FUNGIBILITY, LABELS, AND CONSUMPTION

Johannes Abeler
University of Oxford

Felix Marklein
German Ministry of Finance

Abstract
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 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. We first provide suggestive evidence from the field and then use a laboratory experiment to show this result in a tightly controlled environment. In the lab, 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: C91, C93, D01, H31, I38)

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 and Kahneman 1981; Barberis, Huang, and Thaler 2006; Rabin and Weizsäcker 2009)

The editor in charge of this paper was Stefano DellaVigna.

Acknowledgments: 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 Abi Adams, Ian Crawford, Vince Crawford, Stefano DellaVigna, Gabrielle Fack, Simon Gächter, Lorenz Goette, Lars Garn Hansen, Paul Heidhues, Steffen Huck, David Huffman, Alexander Koch, Peter Kooreman, Anne Laferrère, Ulrike Malmendier, Julia Nafziger, Jörg Oechssler, Susanne Goldlücke, Robert Oxoby, Collin Raymond, Matthew Rabin, Klaus Schmidt, Lael Schooler, Uwe Sunde, Georg Weizsäcker, Matthias Wibral, James Wisson, 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.

E-mail: johannes.abeler@economics.ox.ac.uk (Abeler)

Journal of the European Economic Association January 2017 15(1):99–127DOI: 10.1093/jeea/jvw007
© The Authors 2016. Published by Oxford University Press on behalf of the European Economic Association. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

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, for example, 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.[1] Consumers might violate fungibility also for other psychological reasons, for example, because of feelings of reciprocity toward 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 and 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 *nondistortionary* in-kind grant, that is, an in-kind grant with an amount lower than what the consumer would have spent on the targeted good anyway. For example, we 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. A similar phenomenon has been called a "flypaper effect" as the money "sticks where it hits" (e.g., Hines and Thaler 1995).

Field evidence on how consumers spend in-kind benefits might therefore help to answer whether consumers treat money as fungible. Unfortunately, these field data are 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, even after controlling for income (e.g., Susin 2002; 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, among many things, the low elasticity of housing supply. More direct

1. In the original formulation, mental accounting includes also other heuristics and biases, for example, a preference for a good deal and loss aversion. Recently, the term "mental accounting" has often been used in a more narrow sense of "narrow bracketing of consumption choices". We will use narrow bracketing and mental accounting in the latter sense interchangeably.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

evidence on fungibility comes from labeled cash grants. Beatty et al. (2014) use a regression discontinuity approach and show that the UK winter fuel payment, a cash grant, is disproportionately spent on heating. 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 intrahousehold bargaining or the characteristics of the benefit payment (e.g., periodicity) influence results (see, e.g., Blow, Walker, and Zhu 2012). Hastings and Shapiro (2013) also investigate violations of fungibility but they rely on price shocks instead of in-kind benefits as source of variation. They find that consumers substitute toward lower octane gasoline when gasoline prices rise. The magnitude of the effect is too big to be driven by an income effect and can only be explained by consumers treating their "gas budget" as nonfungible to the rest of their budget.

In this paper, we use a laboratory experiment as an alternative and complementary way to investigate whether consumers treat money as fungible. We first provide suggestive evidence from a quasi-experimental intervention in the field to investigate the effect of nondistortionary in-kind grants on behavior in a natural environment. We then conduct a laboratory experiment to analyze behavior and its potential reasons in a tightly controlled environment. Both studies are based on the same general design. Subjects make consumption decisions and either receive a nondistortionary in-kind grant or a cash grant of equal size in addition to their cash budget. 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 the in-kind grant disproportionately on the targeted good and thus consume too much of this good.

We chose a restaurant as setting for our field study 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 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 since vouchers came as a surprise to participants after they had entered the restaurant. We find that an €8 beverage voucher increases beverage consumption on average by €3.84 compared to a bill voucher. The difference between treatments is very large, almost half of the treatment manipulation. This indicates that guests are influenced by the label attached to the voucher and thus spend too much on the targeted good, violating fungibility.

Although the field study examines a natural setting with exogenously assigned treatments, the causal interpretation of the treatment dummies relies on a number of assumptions. The study was run over only a relatively small number of days and there could be omitted variables that influence consumption on a daily level and that we cannot control for. Additionally, some guests would have initially liked to spend less than the amount of the voucher. These guests were distorted and should increase spending even if fully rational. We need to make additional assumptions on

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

the distribution of beverage preferences to rule out that this rational distortion can generate the observed treatment effect.

In the laboratory, assignment to treatments is fully randomized and we do not need to rely on additional identifying assumptions. The grant is also nondistortionary for all subjects, so we do not need assumptions on the distribution of preferences. We supplement the two main laboratory treatments with five additional treatments and collect further information in a questionnaire to investigate the mechanisms underlying behavior. In our lab experiment, subjects again decide on consumption and had at their disposal a cash budget and either an in-kind grant or a cash grant. We induce a payoff function by specifying monetary payoffs for all possible consumption levels; we thus know the optimal decision. We find that, as in the field study, 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. About 20% of subjects who receive the in-kind grant increase their consumption of the targeted good by the full amount of the grant. We show in several additional treatments that subjects' tendency to violate fungibility is not influenced by their inclination to honor the stated preferences of a potential benefactor and that the treatment effect is robust to learning and to increases in stake size. We also find that the treatment difference is driven by subjects with lower cognitive skills.

There exist potentially several psychological mechanisms for a violation of fungibility. In Section 5, we discuss a formal model of mental accounting or narrow bracketing that can explain most of the results from both studies. In Online Appendix A, we discuss to what extent several other models can explain our data. Loss aversion (Kőszegi and Rabin 2006) and salience (Bordalo, Gennaioli, and Shleifer 2013) can align many results but only with a relatively strong additional assumption similar to narrow bracketing. Reciprocity (Rabin 1993) can explain some of our results in the field study but cannot explain the laboratory results.

Our paper provides several novel insights: First, whereas 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, Laibson, and Madrian (2009) study a 401(k) plan for which the firm allocated half of the assets and the employees the remaining part. They show that employees do not consider the assets allocated by the firm when deciding how to invest their contribution (see also Odean 1998). Second, previous laboratory experimental studies on narrow bracketing of consumption decisions were not incentivized (e.g., Heath and 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. Third, the few existing studies investigating consumption decisions in incentivized environments analyze how people spend a gift or a windfall gain (e.g., Bodkin 1959; Milkman and 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. These studies find,

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

however, that people spend more after receiving a gift or windfall gain (but see Kuhn et al. 2011 for more mixed evidence for a large in-kind windfall, namely a car). 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[2] or for other reasons), it is not evident whether a change in spending can be clearly linked to a violation of fungibility. This is much less likely to 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 study 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. The specific design of our study also suggests implications for public policy. Our results suggest that in-kind benefits will distort consumers' decisions more than previously assumed. We discuss the policy implications in more detail in Section 6.

The paper is organized as follows: The general design of both studies is described in Section 2. Section 3 reports the detailed setup and results of the field study. Section 4 presents design and results of the laboratory experiment. We present a model that is in line with our results in Section 5. Section 6 concludes.

2. Conceptual Framework and Hypotheses

We examine whether consumers treat money as fungible by investigating how they spend different kinds of lump-sum grants. In our experiments, subjects can decide over consumption of up to three goods: consumption of the targeted good c_t, of the nontargeted good c_{nt} and of a composite other good c_o. Overall utility is thus $U = V(c_t, c_{nt}, c_o)$, where $V()$ is the consumption utility. Prices are p_t, p_{nt}, and 1, respectively. Subjects have a cash budget B at their disposal. The three treatments differ in whether subjects additionally receive a grant of size G and in whether the grant is restricted to the use for $p_t c_t$ (*Label treatment*) or for $p_t c_t + p_{nt} c_{nt}$ (*Cash treatment*). In the *Baseline treatment*, subjects do not receive a grant. The budget constraints are thus

$$\begin{aligned} \text{Baseline treatment:} \quad & p_t c_t + p_{nt} c_{nt} + c_o \leq B \\ \text{Cash treatment:} \quad & p_t c_t + p_{nt} c_{nt} + c_o \leq B + G \ \wedge \ c_o \leq B \\ \text{Label treatment:} \quad & p_t c_t + p_{nt} c_{nt} + c_o \leq B + G \ \wedge \ p_{nt} c_{nt} + c_o \leq B \end{aligned}$$

Denote by c_t^*, c_{nt}^*, and c_o^* the payoff-maximizing consumption level of the three goods for decision out of B, that is, without any grant.

2. For some examples of the extensive research in psychology and marketing on the influence of mood on (consumption) choices, see Lewinsohn and Mano (1993) or Groenland and Schoormans (1994).

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

The crucial feature of our design is that the amount of the grant G is lower than the amount $p_t c_t^*$ spent optimally on the targeted good without grant. Even if G is given in kind for the targeted good, the consumer can reallocate parts of his cash budget B to still reach the first-best choice, if they treat money as fungible. The type of grant will thus not influence consumption of rational consumers.

In the restaurant study, B is the guests' own wealth. G is an 8-euro voucher, therefore $G \ll B$. Receiving a grant at all will thus also not matter as the income effect of G will be negligibly small. Therefore, in the limit as $G/B \to 0$ and if guests are rational, $c_t = c_t^*$ in all three treatments. The additional budget G will be spent entirely on nonrestaurant consumption. Figure 1a depicts this situation graphically. To make the correspondence to the lab experiment clearer, we depict consumption choices conditional on c_o^*, the optimal level of nonrestaurant consumption without grant. Consumption of the targeted good (c_t) is on the horizontal axis and consumption of the nontargeted good (c_{nt}) is on the vertical axis. The dot-and-dash line is thus the resulting budget constraint if the consumer has only B at his disposal and is consuming c_o^* of nonrestaurant consumption. The optimal consumption bundle in *all* three treatments is at point A. Now consider a guest who does not treat money as fungible. In the Baseline treatment, we would still expect such a consumer to choose the first-best bundle A. The difference to the standard model occurs in the two other treatments. The grant will be spent disproportionately on the targeted good in the Label treatment or on the targeted and nontargeted good in the Cash treatment as the consumer does not take advantage of the possibility to reallocate parts of his cash budget. Subjects who treat money as completely nonfungible will choose bundle C in the Label treatment where $c_t = c_t^* + G/p_t$. In the Cash treatment, these subjects will choose a bundle on the line between C and D.

In the lab experiment, there is no third good and B and G can only be spent on the targeted and nontargeted good. There is also no Baseline treatment; instead, subjects make two decisions, once on consumption out of B (*baseline stage*) and once out of $B + G$ (*grant stage*). We induce the payoff function and thus know what a rational subject will consume in the two treatments. Figure 1b depicts the situation graphically. In the baseline stage the budget constraint is the dot-dash-line and the optimal consumption bundle (c_t^*, c_{nt}^*) is at point E. The payoff function guarantees that $p_t c_t^* > G$ for all subjects. In the grant stage of both Label and Cash treatment, the optimal bundle is at H. The dashed line is the budget constraint in the Cash treatment; the solid line is the budget constraint in the Label treatment. As in the field study, a rational subject's consumption will not differ by treatment.[3] A subject who does not treat money as fungible will still choose the first-best bundle H in the Cash treatment, as both income component are as good as cash. In the Label treatment, the grant will, however, be spent disproportionately on the targeted good. In the extreme case of full nonfungibility, subjects will choose bundle F.

3. In fact, the choice sets are exactly identical in both treatments, as we cap the maximum consumption for each good, thus effectively cutting off the triangle with $c_{nt} > B/p_{nt}$ also in the Cash treatment.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

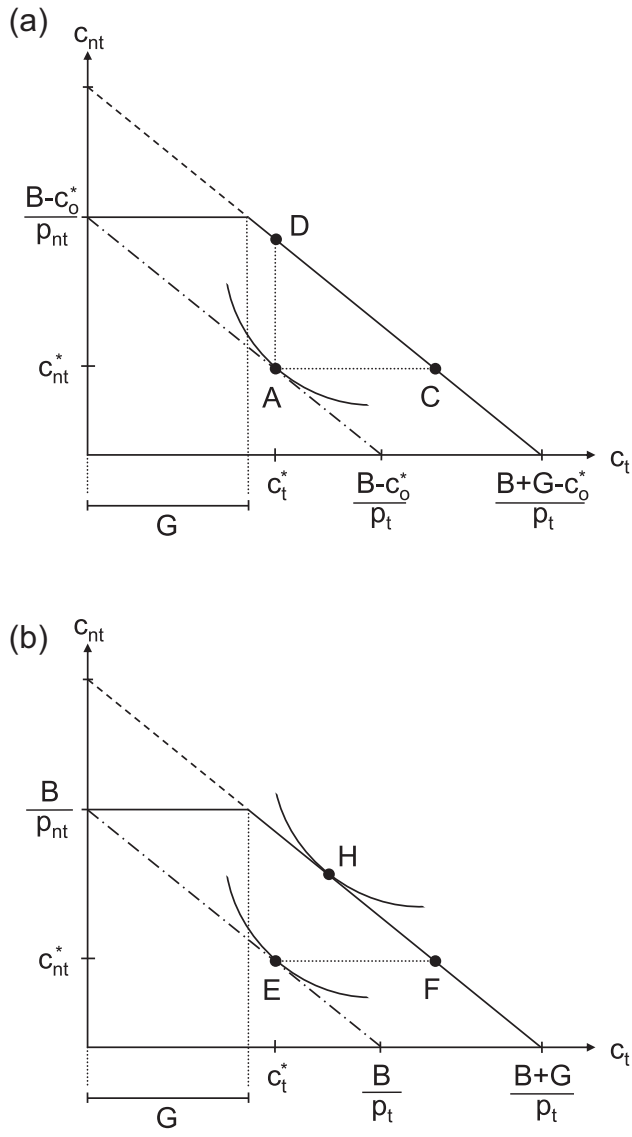

FIGURE 1. (a) Consumption decision in the field study. (b) Consumption decision in the laboratory experiment. Consumption of the targeted good is on the horizontal axis, consumption of the nontargeted good on the vertical axis.

Overall, across both studies, if at least some subjects do not treat money as fungible, we thus expect average consumption of the targeted good in the Label treatment to be higher than in the Cash treatment. In the next section, we present the detailed design and results of the field study. The lab experiment is discussed in Section 4.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

3. Field Study

We chose a restaurant as setting to conduct the field study. 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 study took place in a wine restaurant situated in the Palatinate, a wine-growing region in southern Germany. The restaurant itself is located in a winery. Usual per-person spending in this restaurant is about €40 (∼$54 at the time of the study); about 40% of the total is spent on beverages (the majority of beverage spending is used for wine and 99% of tables consume at least some wine). This setting thus matches the two-goods case presented in Figure 1a well.

Guests were not aware of participating in a scientific study. 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 voucher per person (∼$11). The type of voucher differed by day: during 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. During the *Label treatment*, vouchers were given as a "gourmet beverage voucher". These vouchers were restricted to be spent on beverages. Both types of vouchers had to be redeemed the same evening. 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 (it is very unusual to not consume beverages in German restaurants; water must also be purchased). Therefore, the beverage voucher should be nondistortionary for most guests. We will first present data on the full sample and then address potentially distorted guests. In a third treatment, the *Baseline treatment*, guests did not receive any voucher.

We have data on 552 guests spread over 22 days and 187 tables: four days (37 tables) in the Label treatment, three days (34 tables) in the Cash treatment, and 15 days (116 tables) in the Baseline treatment (nine days before and six after the main treatments). Overall, 107 vouchers were distributed in the Label treatment and 89 vouchers in the Cash treatment, one per person. 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.[4] 356 persons participated in the Baseline treatment. We 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. During the observed period, the menu did not change and the same

4. The Label treatment took place from 15 August to 22 August 2007. The Cash treatment took place from 23 August to 29 August 2007. We have data on the Baseline treatment from 29 July to 10 August 2007, and from 5 September to 29 September 2007. The restaurant is closed on Mondays and Tuesdays and we only conducted the main treatments on weekdays to reduce variance (we control for days of the week in the regressions below). Unfortunately, the restaurant did not collect data for the four days directly after the two main treatments. The restaurant was closed for a couple of days in the second half of September.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

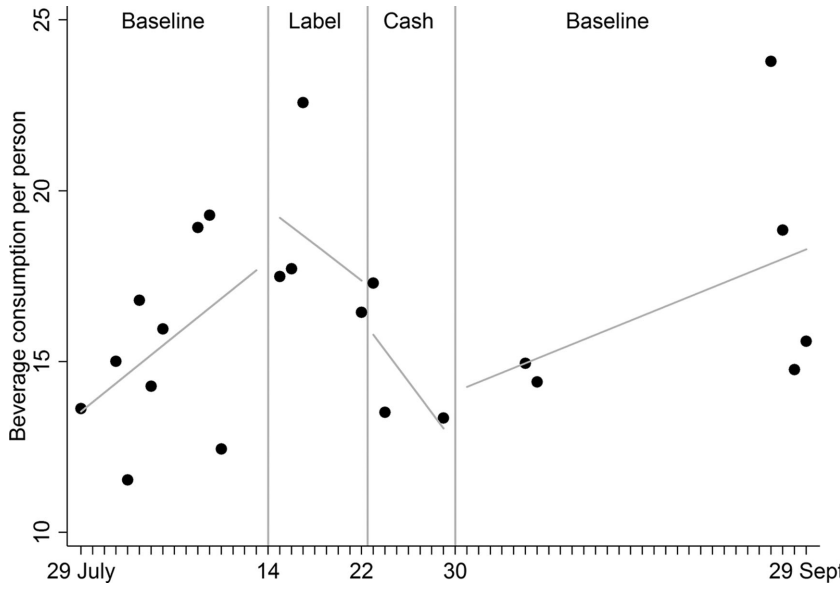

FIGURE 2. Scatter plot of per-person beverage consumption (averaged to daily level).

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 and whether the table was inside the restaurant or on the terrace. Additionally, we collected data about the local weather (average outside temperature, duration of sunshine on this day, whether it rained, average atmospheric pressure). Table C.1 in the Online Appendix documents averages of consumption choices and control variables for the three treatments. Figures C.1 and C.2 show histograms of beverage and meal consumption in the three treatments.

Figure 2 shows a scatter plot of per-person beverage consumption averaged on the daily level. Our main focus will be on the comparison of Label to Cash treatment, since this comparison allows for the cleanest assessment of whether fungibility is violated. Participants in both treatments receive a voucher; changes in consumption patterns because of receiving a voucher per se therefore cannot influence the treatment comparison. The days in the Label treatment show higher consumption on average than the days in the Cash treatment, even though a clear break is difficult to discern with so few days. Participants in the Label treatment—who receive a beverage voucher—spend on average €18.55 per person on alcoholic and nonalcoholic beverages, €3.84 more than participants in the Cash treatment and also more than participants who did not receive either voucher. This treatment difference is very large compared the value of the grant (€8) and to the average beverage consumption in the Baseline treatment (€16.01).

Table 1 summarizes the regression evidence for the field study. We average all data by day and allow for interdependence across days by using Newey-West standard

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

TABLE 1. Treatment effects in the field study (data aggregated to day).

Dependent variables:	Beverage consumption (1)	(2)	log(beverage consumption) (3)	Volume of consumed beverage (4)	Price of consumed beverage (5)	Total consumption (6)
1 If either voucher	−1.295	0.545	0.046	−0.017	0.561	7.648*
	(1.296)	(2.257)	(0.165)	(0.070)	(2.247)	(3.518)
1 If Label treatment	3.838**	5.114*	0.319*	0.064	4.457*	
	(1.691)	(2.758)	(0.170)	(0.052)	(2.398)	
Additional controls	No	Yes	Yes	Yes	Yes	Yes
Constant	16.012***	−40.261	−0.898	−0.319	75.271	234.839
	(0.892)	(157.913)	(10.313)	(3.503)	(169.288)	(222.592)
No. obs.	22	22	22	22	22	22

Notes: Data are averaged on day. Newey-West standard errors (2 lags) are in parentheses. The dependent variable is beverage consumption (in euros per person) in columns (1)–(2), the logarithm of beverage consumption in column (3), volume of consumed beverages (in liter per person) in column (4), price of consumed beverages (in euros per liter) in column (5), and total consumption (in euros per person) in column (6). The additional controls are: outside temperature, duration of sunshine on this day, whether it rained, atmospheric pressure, dummies for the day of the week, share of tables served on the terrace, and the average number of guests at a table on this day. *Significant at 10%; **significant at 5%; ***significant at 1%.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

errors. The dependent variable is the gross per-person beverage consumption, that is, any spending on beverages regardless of whether it is covered by a voucher or out-of-pocket expense. We regress this variable on a dummy for receiving any voucher at all and a dummy for being in the Label treatment. 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 by €3.84. This means that merely attaching a label to the grant increases consumption of the targeted good. In column (2) we control for several variables that might influence consumption: the local weather (average outside temperature, duration of sunshine on this day, whether it rained, average atmospheric pressure), dummies for the day of the week, the share of tables that were served on the terrace instead of inside, and the average number of guests at a table on that day. None of the control variables is significantly different at the 5% level between treatments. There were also no major holidays during the two main treatments. The treatment effect is less significant but the point estimate is even larger when we control for these variables. In column (3) we regress the logarithm of beverage consumption on the treatment dummies and the control variables.[5]

For these regressions to identify the causal effect of the treatments we need to assume that there are no omitted variables that we cannot control for but that affect consumption on a daily level. Reverse causality is not an issue since treatments were exogenously imposed by us. Even though we control for several variables that might influence outcomes in column (2), this is a strong assumption and we would thus treat the field study only as suggestive evidence.[6] In the laboratory experiment, described in the next section, assignment to treatments is truly random, allowing for a direct causal interpretation of the treatment dummies.

In the analysis above we ignored the fact that the grant was distortionary for some guests. In the Cash and Baseline Treatment, 16% of tables consume beverages for less than the value of the voucher, spending €5.81 on average. For these guests, the voucher was distortionary. They should increase their beverage consumption even if they treat money as fungible. It could be, however, that more than 16% of individual guests were distorted. A table with an average per-person consumption of more than €8 could be composed of low-consuming and thus distorted guests and very high-consuming guests. Distorted guests could thus be at tables with any average beverage consumption, given that the other guests consume enough. Furthermore, it

5. We can also derive a nonparametric estimate of the treatment effect by repeatedly randomizing days to treatments, then running our regressions on these placebo treatments and finally comparing the distribution of the placebo effects to the actual treatment effect. This bootstrap procedure yields a p-value for the Label treatment dummy of 0.048 for the regression in column (1) and of 0.126 for column (2).

6. If one is additionally willing to assume that observations across different days are independent, we can use each table as an observation and cluster standard errors on each day to allow for within-day spill overs or other dependencies within day. Table C.2 in the Online Appendix reports regressions like in Table 1 for this data set. Coefficients become generally more significant.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

could be that distorted guests rationally increase beverage consumption beyond €8, as beverage bundles above €8 become relatively cheaper compared to beverage bundles below €8. Without further assumptions on the share of distorted guests and on how distorted guests change their behavior, these rational distortions could account for our treatment effect. In Online Appendix B we discuss this in detail and document sufficient assumptions that rule out rational distortion. For example, if we are willing to assume that guests do not increase their out-of-pocket expenditure, when they receive a (distortionary) beverage voucher, then up to 48% of guests could be distorted without rational distortion being able to generate the observed treatment effect. This is another reason why we treat the field study only as a starting point for the laboratory experiment. There, we can ensure that the grant is nondistortionary for all subjects by choosing an appropriate payoff function and by endowing every subject with the same budget.

The additional spending on beverages that we have documented above is not uniformly allocated to all kinds of beverages. Guests in the Label treatment do not consume more beverages in terms of volume but they consume more expensive beverages. In column (4) of Table 1, we regress the volume of consumed beverages on treatment dummies and controls. The effect of the voucher and of the type of voucher on consumed volume is relatively small (e.g., Cash vs. Label: 64 ml, or 2 fl oz). In contrast, the type of voucher influences the average price of the consumed beverages quite strongly (column (5)). Receiving any voucher at all does not have a significant impact. But if the voucher is targeted to beverages, the average price rises by €4.46 per liter.

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 increase in lifetime income can surely be neglected. Rational guests will not be affected by the voucher but guests who violate fungibility should increase overall 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 and Mano 1993; Groenland and Schoormans 1994). 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.

We find that overall spending is higher in both voucher treatments compared to the Baseline treatment. In Table 1, column (6), we regress total consumption on a dummy for receiving a voucher on the controls described above. Participants in the two voucher treatments spend on average €7.65 more than in the Baseline treatment.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

4. Laboratory Experiment

4.1. Design of the Laboratory Experiment

In the two main treatments of 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, which we framed as "housing" and "clothing". 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. Payoffs were converted into euros and paid to subjects at the end of the experiment. In the first decision stage, which we will call *baseline stage*, subjects received a cash budget of 50 money units that 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, that is, the money had to be spent entirely on the targeted good, namely housing. Parameters were chosen such that the in-kind grant was by design nondistortionary 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. We will refer to the two main treatments as "BG Cash" and "BG Label" (for baseline stage—grant stage).

The exact specification of the payoff functions is presented in Table 2. For each good, payoff increases in consumption and marginal payoff weakly decreases. Prices per unit were $p_t = 3$ for the targeted good (housing) and $p_o = 2$ for the other good (clothing). 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

TABLE 2. Payoff functions in the laboratory experiment.

Consumed units	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Payoff													
Housing (targeted good)	0	36	70	102	132	160	186	210	232	252	270	286	299
Clothing (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													
Housing (targeted good)	310	316	322	328	333	338	343	347	351	355	358	361	364
Clothing (other good)	200	208	216	223	230	237	244	251	256	261	266	271	276

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

consumption bundle in the baseline stage is $E = (12, 7)$; the optimal bundle in the grant stage is $H = (13, 20)$; the bundle F is $(22, 7)$.[7]

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.[8] 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, Frykblom, and Shogren 2002).

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 Online Appendix D for an English translation of the instructions of the two main treatments and the five additional treatments described below). Detailed instructions for the second stage were given only later on the computer screen. This allowed us to have subjects of the two main treatments in the same session and thus to align the delivery of the two treatments as much as possible (the five control treatments were conducted at a later point). Payoff points (cf. Table 2) were paid out after the experiment at a rate of 100 points = €1. Subjects received a show-up fee of €2.50 besides their earnings from the consumption decisions.

We additionally conducted five variants of the BG Label treatment described above. Table 3 gives an overview of all seven treatments.

In the treatment "Repetition Label", the baseline stage was followed by five grant stages in which subjects could spend 50 money units cash plus an additional in-kind grant of 24, 27, 30 (as in the main treatments), 33, or 36 money units. The order of the five grant stages was randomized for each subject. A subject thus faced the 50+30 decision as the second, third, . . . , or sixth decision, that is, with different degrees of experience. By comparing subjects who decided early on over the budget of 50+30 with subjects who made this decision later we can investigate the effect of learning. We chose to investigate learning effects by presenting subjects with slightly different repeated decisions because this forces subjects to think anew for each decision, facilitating their learning process. After each stage, subjects were informed about their earnings in that stage, further assisting learning. The exchange rate was set to 300 points = 1 euro to

7. 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 additional effort to reach a higher monetary payoff, see Section 4.3 for details. There are several ways to find the optimal bundle. Probably the fastest is to realize that both payoff functions are increasing and that marginal payoff weakly decreases. We can thus equalize (as much as possible) the marginal payoff while exhausting the budget and then check neighboring bundles whether the discreteness of the consumption space makes adjustments profitable.

8. 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 min. If subjects did not complete the task, they got an endowment of ten money units only. This was the case for 14 out of 442 subjects who will be excluded from the analysis.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

TABLE 3. Overview of lab experimental treatments.

Treatment name	Targeted good	Order of stages	Note
BG Cash	–	Baseline, Grant	
BG Label	Housing	Baseline, Grant	
Repetition Label	Housing	Baseline, 5 × Grant	
GB Label	Housing	Grant, Baseline	
High Stakes Label	Housing	Baseline, Grant	Stakes doubled
Transparent Randomization (Housing)	Housing	Baseline, Grant	Open randomization into treatments
Transparent Randomization (Clothing)	Clothing	Baseline, Grant	Open randomization into treatments

keep the overall incentive level comparable to the main treatments. Everything else, including the payoff function, was held constant.

In the treatment "GB Label" we switched the order of baseline and grant stage: subjects first had a 50 money unit cash budget plus a 30 money unit in-kind grant at their disposal (as in the grant stage of BG Label). In the second stage they received only a cash budget of 50 money units (as in the baseline stage of BG Label). The design of GB Label is even closer to the setup in the field, as restaurant guests also decided directly on consumption with a grant at their disposal without a preceding baseline stage.

The design of the treatment "High Stakes Label" was identical to BG Label except that we altered the exchange rate to 100 points = €2, thus doubling the monetary incentives. If subjects do not bother to optimize in BG Label because of the perceived low stakes, subjects in High Stakes Label should choose consumption bundles closer to the optimum.

A violation of fungibility might be driven by a cognitive bias like narrow bracketing or by a perceived obligation to reciprocate the receipt of the grant by complying with the stated preferences of the giver. The latter effect should be greatly reduced in the lab experiment as 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.[9] Still, we could imagine that there is a whiff of stated preferences in the two main treatments: some subjects might wonder whether the experimenter chose the targeted good on purpose and wanted thus to signal a preference for this good. We completely eliminate this line of reasoning by conducting two treatments with open and Transparent Randomization into treatments. In the grant stage of these

9. Any suggestion (or informed-principal) effect should also be reduced in the lab experiment. Such an effect assumes that subjects believe that the grant provider is benevolent and 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 did have all necessary information available and could easily check whether the suggestion of the label lead to higher payoffs, which was indeed not the case.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

treatments, subjects either received an in-kind grant for the good called "housing" (as in BG Label) or an in-kind grant for the other good ("clothing"). Importantly, subjects knew about the other treatment as both treatments were conducted in the same laboratory session. And since the random assignment to treatments was done openly and transparently, they were well aware that they were randomly allocated to their respective treatment. They could therefore not attach any intentionality to the choice of their targeted good. For the exact procedure of the randomization see the instructions of these treatments in Online Appendix D. We will refer to the treatment in which subjects ended up receiving an in-kind grant for housing (like in BG Label) as "Transparent Randomization (Housing)" and to the other treatment as "Transparent Randomization (Clothing)". The remaining procedure was exactly as in BG Label. If the stated-preferences effect drives most of a potential difference between the two main treatments, both Transparent Randomization treatments should lead to similar consumption choices. If, however, a violation of fungibility is mostly driven by cognitive reasons, Transparent Randomization (Housing) will lead to behavior similar to BG Label whereas Transparent Randomization (Clothing) will yield choices close to the optimal bundle, as also mental accounting and narrow bracketing push subjects towards the optimal solution in this treatment.[10]

At the end of the experiment, subjects answered a questionnaire. The experiment was computerized using z-Tree and ORSEE (Greiner 2004; Fischbacher 2007). Subjects in all treatments were students from the University of Bonn studying various majors. Treatments were assigned randomly and no subject participated in more than one treatment. 427 subjects participated in the seven treatments and completed the real-effort task successfully.[11] On average, subjects earned €13.20. Sessions lasted between 60 and 70 min.

4.2. Results of the Laboratory 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 main (BG) treatments. Accordingly, we find that behavior in this stage is not different across treatments. In Table 4, column (1), we regress consumption of housing (the later to be targeted good) in the baseline stage on a dummy for the Label treatment.[12] We use tobit estimates to account for the fact that subjects could only buy between 0 and 25 units of each good. In column (2), we also control for gender, age,

10. As mentioned above, the optimal bundles are $A' = (12, 7)$ in the baseline stage and $A = (13, 20)$ in the grant stage, regardless of the framing of the grant. To maximize payoffs, we should thus spend the additional income of the in-kind grant almost exclusively on the second good, that is, clothing.

11. One subject chose a consumption bundle close to zero by mistake, thereby foregoing almost all earnings. The subject wanted to change the decision but this was not possible during the experiment. We thus exclude this subject from the analysis. Results are robust to including the subject.

12. For ease of exposition, we report only the consumption of housing. Consumption of clothing, the other good, can then be readily calculated as only few subjects choose a consumption bundle that is not on the budget frontier (13 out of 300 decisions in the two main treatments).

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

TABLE 4. Consumption in the two main treatments of the laboratory experiment.

Dependent variables:	Consumption of the targeted good ("housing")				Absolute distance to optimal consumption			
	Baseline stage		Grant stage		Baseline stage		Grant stage	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 If Label treatment	−0.133	−0.124	2.205***	2.145***	−0.000	0.031	2.634***	2.274***
	(0.350)	(0.354)	(0.636)	(0.582)	(0.455)	(0.456)	(0.767)	(0.675)
Consumption in baseline stage				0.607***				
				(0.135)				
Distance to optimum in baseline stage								0.811***
								(0.205)
Controls for gender, age, major	No	Yes	No	Yes	No	Yes	No	Yes
Constant	11.440***	12.150***	14.442***	7.679***	0.741**	2.496	1.020*	−1.350
	(0.248)	(1.417)	(0.449)	(2.848)	(0.333)	(1.855)	(0.558)	(2.692)
No. obs.	150	150	150	150	150	150	150	150

Notes: Tobit estimates. The dependent variable is consumption of the targeted good ("housing") in the baseline stage (columns (1)–(2)) or in the grant stage (columns (3)–(4)) and absolute distance to optimal consumption of the targeted good in the baseline stage (columns (5)–(6)) or in the grant stage (columns (7)–(8)). Data from BG Label and BG Cash are included in the analysis. 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). Standard errors are in parentheses. *Significant at 10%; **significant at 5%; ***significant at 1%.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

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. More importantly, the regressions 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.[13]

Next, we analyze outcomes in the grant stage.

RESULT 1. *Consumption of the targeted good is significantly higher in BG Label than in BG Cash.*

In the grant stage, subjects in BG Label buy too much of the targeted good. They buy 16.6 units on average, compared to 14.4 units in BG Cash and an optimal consumption level of 13. The estimates in Table 4, 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.[14] The treatment effect is also significant when we take the distance to the optimal consumption as dependent variable (column (7)). In column (8), 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 BG Label choose consumption bundles significantly further away from the optimal bundle. By consuming too much of the targeted good, subjects in BG Label leave money on the table as their choices translate into significantly lower payoffs. With payoff in the grant stage as dependent variable, the p-values of the treatment dummy are 0.014 and 0.004 in tobit regressions equivalent to the specifications in Table 4, columns (3) and (4) (limiting payoff at the maximal reachable level). The treatment effect is about -18 points. This relatively small effect size is due to the concave payoff function: any deviation from the optimal choice will lead to second-order losses in payoff.

This result confirms the main finding of our field study: even in this stylized and tightly controlled environment subjects do not treat money as fungible. Our next result documents a considerable heterogeneity within treatments.

RESULT 2. *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 a within-person measure of behavior by comparing decisions in the grant stage to decisions in the baseline stage. A histogram of the within-person change in consumption for the two

13. 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 (see columns (5) and (6) in Table 4). Most subjects choose consumption levels close to the optimum; the average distance is 1.5 units.

14. Results for this and subsequent regressions do not change if we interact the baseline stage outcome with the treatment dummy, allowing for heterogenous inertia across treatments. See Table C.4 in Online Appendix C for details.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

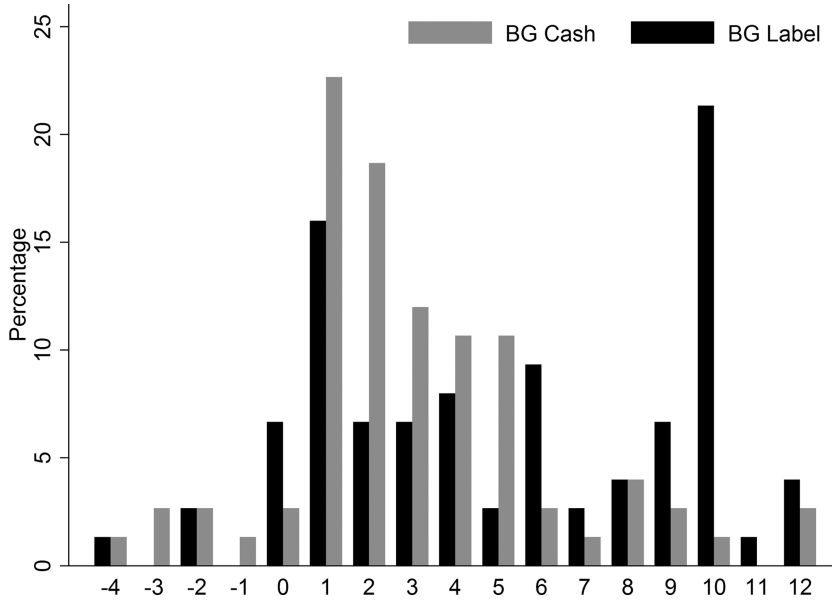

FIGURE 3. Consumption increase of housing, the targeted good, from baseline stage to grant stage in the two main treatments. The grant is worth ten units of the targeted good.

main treatments is shown in Figure 3. The grant was worth ten units of the targeted good. In line with the results reported above, the consumption increase is higher in BG Label than in BG Cash (t-test, $p < 0.001$). Moreover, decisions are highly heterogeneous in BG Label. The most frequent consumption increase in BG Cash is by either one or two units, often leading to a choice of the optimal consumption bundle in the grant stage. By contrast, the modal choice in BG Label is a consumption increase by ten units, that is, 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 nonfungible will do exactly this (cf. bundle F in Figure 1b). In BG Label, 21% of subjects spend the whole grant on the targeted good, whereas this is true for only 1% of subjects in BG Cash. 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 persists, especially in the distance to optimal consumption, although it is smaller and not always significant (see Table C.5 in Online Appendix C). Interestingly, subjects who spend the entire grant on the targeted good also decide much faster than the remaining subjects. They need on average 116 s for their decision, whereas the other subjects need 267 s, 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, like mental accounting, rather than extensive deliberations.

Our next two results show that the treatment effect also obtains in the additional treatments.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

RESULT 3. *The treatment effect does not change if we give subjects the possibility to gain more experience, switch grant and baseline stage or double the stake size.*

In the Repetition Label treatment, subjects faced five grant stages with in-kind grants of 24–36 money units in random order. The data in Table C.3 (in Online Appendix C) include only the consumption decision for the grant of 30 money units (the same size as in all other treatments). Thus, the table includes subjects with different degrees of experience. The treatment effect remains unchanged: consumption of the targeted good and absolute distance to optimal consumption are significantly higher in Repetition Label than in BG Cash (Table C.3). At the same time, behavior is not different from BG Label (all p-values are between 0.100 and 0.958 in regressions akin to columns (3)–(4) and (7)–(8) in Table C.3 with BG Label as omitted category).[15]

The treatment effect is also unchanged if we switch baseline and grant stage in treatment GB Label or if we double the stake size in treatment High Stakes Label (Table C.3). Moreover, behavior in these two treatments is not different from behavior in BG Label (all p-values are between 0.217 and 0.958). The fact that GB Label yields the same treatment effect as BG label implies that the violation of fungibility works similarly for gains and losses, that is, the increase in consumption when receiving an additional in-kind grant is not different from the decrease in consumption when losing a previously held in-kind grant.

RESULT 4. *Eliminating any stated-preferences effect by making the random treatment allocation transparent does not eliminate the treatment difference.*

In the two Transparent Randomization treatments, subjects knew about the other treatment and knew that they were randomly allocated to receive a grant for the good "housing" (as in all other Label treatments) or for the other good, called "clothing". We find that consumption and absolute distance to optimal consumption are significantly higher in Transparent Randomization (Housing) than in BG Cash (see Table C.3 in the Online Appendix) and also higher than in Transparent Randomization (Clothing).[16] A stated-preference effect therefore does not drive our main treatment effect.[17]

15. We can also investigate the effect of experience or learning directly by comparing subjects who decided early on over the budget of 50+30 with subjects who made this decision later. In a regression of consumption on the position of the 50+30 decision (one to five), this trend is not significant ($p = 0.338$). The same holds true with absolute distance to optimal consumption as dependent variable ($p = 0.167$) or when we compare periods individually against each other (all $p > 0.187$).

16. In regressions like in Table C.3 but restricted to the two Transparent Randomization treatments all treatment dummies have $p < 0.008$.

17. The results on the heterogeneity in consumption increase (Figure 3) also hold for the sample including the relevant additional treatments, that is, all except Transparent Randomization (Clothing), see Figure C.3 in Online Appendix C.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

4.3. Determinants of Behavior in the Laboratory

The Transparent Randomization treatments show that a stated-preferences effect does not cause the behavior we observe. Rather, a cognitive bias like narrow bracketing seems to drive the treatment effect. But can we find more direct evidence that narrow bracketing indeed underlies behavior 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, Loewenstein, and Rabin (1999) also conjecture that: "Cognitive limitations—in perception, attention, memory, and analytical processing, and so forth—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. Our next result supports this hypothesis.

RESULT 5. *The treatment difference in consumption is driven by subjects with lower cognitive skills.*

We use subjects' math grade in their final high school exam "Abitur" (which is a weighted average of their exam grades in the last two years of school) as a proxy for their cognitive, in particular 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; and the grade covers performance over a long period, reducing measurement error. Most importantly, however, math grades are highly correlated with other measures of intelligence and cognitive ability.[18] The grades were elicited in the postexperimental questionnaire. Grades range from 1 (best grade) to 6 (fail), a higher grade thus indicates a poorer performance.

Figure 4 shows the results graphically. We split subjects at the median of the math grade distribution, that is, grades 1 and 2 are classified as "good grade" and grades 3 and worse are classified as "bad grade". Although the grade has no influence in BG Cash, subjects with bad grades choose higher consumption and are further away from the optimum than subjects with good grades. In Table 5, columns (1)–(4), we regress consumption of the targeted good on a Label dummy, the math grade of

18. 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. Standard cognitive-ability tests (or "IQ"-tests) aim to measure *g*. 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, Brown, and Shapiro 2013). The correlation between SAT-score and *g* has been estimated as 0.70 (Brodnick and Ree 1995), very similar to the correlation between math grade and *g*.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

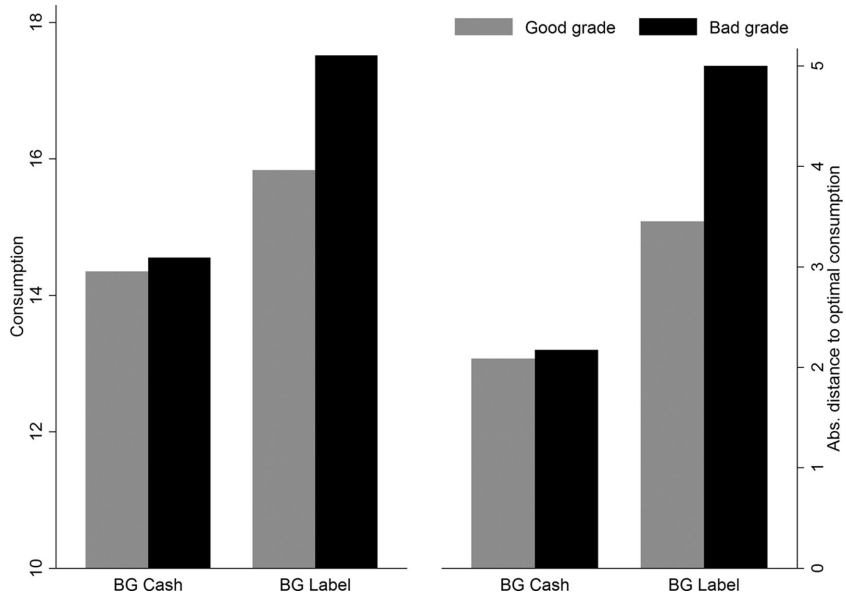

FIGURE 4. Consumption and absolute distance to optimal consumption in the grant stage for subjects with above and below median math grades.

subjects and an interaction of grade and Label. This table includes only data from the two main treatments, but the results also hold if we include data from all relevant variants of the main treatment (see Table C.6 in Online Appendix C). We include the same control variables as before. By including dummies for subjects' major we control for any additional effects of university (math) education that might influence decision making. The specification in column (1) of Table 5 shows that the 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 BG Cash (the coefficient of grade is very small and not significant). Only the effect on behavior in BG Label is pronounced and significant. A one grade higher (i.e., worse) math score results in a consumption increase of 1.01 units on average. The result that the math grade has only an effect on behavior in BG Label but not in BG Cash is confirmed when we estimate separate regressions for each treatment (columns (3) and (4)). Although the effect is not significant in BG Cash, it is again sizable and significant in BG Label. The results also obtain if we take the absolute distance to the optimal consumption as dependent variable (see Table 5, columns (5)–(8)).[19] After the five additional treatments, we also collected data on verbal cognitive

19. In the German high school system, there are two types of math course: intensive and basic course (*Leistungskurs* and *Grundkurs*). If we include a dummy for the type of subjects' high school course and an interaction term of course and Label treatment, results of Table 5 do not change and the math grade

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

TABLE 5. Impact of cognitive ability in the two main treatments.

	Consumption of the targeted good				Absolute distance to optimal consumption			
	Baseline	Grant	Grant stage		Baseline	Grant	Grant stage	
Dependent variables:	Full sample		BG Cash	BG Label	Full sample		BG Cash	BG Label
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 If BG Label	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 baseline 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)
No. obs.	150	150	75	75	150	150	75	75

Notes: Tobit estimates. The dependent variable is consumption of the targeted good (i.e., housing) in the baseline stage in column (1) or in the grant stage (columns (2)–(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 two main treatments; columns (3) and (7) report results for BG Cash and columns (4) and (8) for BG Label only. The lower limit for the tobit estimation is 0 in all specifications; the upper limit is 25 for columns (1)–(4) and 12 for columns (5)–(8). Standard errors are in parentheses. *Significant at 10%; **significant at 5%; ***significant at 1%.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

ability. This allows us to further pin down which type of cognitive ability exactly influences fungibility. We know subjects' grade in the German course of their final high school exam; subjects additionally took a standard test assessing their vocabulary, a measure of verbal skill (Lehrl et al. 1991). These variables do not correlate with behavior in the experiment, suggesting that the observed differences in behavior are indeed driven by nonverbal cognitive ability. Note that the additional treatments are all Label treatments, we thus cannot correlate verbal scores with the Cash-Label treatment effect.

As further suggestive evidence, we measured subjects' feelings of obligation to comply with the stated preference of a grant giver directly. After the experiment, we described in a short vignette that a couple spent less money on their kids than what they received as child benefits. Subjects then had to state how "justified" or "appropriate" they found this behavior. 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 correlate with behavior in the Label treatments. This corroborates our findings from the Transparent Randomization treatments that behavior in the lab is not driven by a stated-preferences effect.

5. Model of Mental Accounting

In this section, we present a model of mental accounting or narrow bracketing of consumption decisions that can align most of our results.[20] The literature suggests two reasons why people bracket narrowly, namely, limited self-control and decision costs. If people suffer from limited self-control (Thaler and Shefrin 1981), it can be optimal for them to adopt a set of personal rules that are painful to violate and that restrict the consumption of future selves (Bénabou and Tirole 2004). In particular, Thaler (1980, 1985) has suggested that many people adopt a system of mental accounts that split the overall budget into sub-budgets reserved for the consumption of specific goods. Such a system counteracts overspending on tempting goods by limiting the maximally available amount for consumption of these goods.[21] Another reason to bracket narrowly is people facing direct costs of making decisions (e.g., Conlisk 1996). Making globally optimal decisions is much more difficult than making a sequence of locally optimal decisions, that is, for each bracket separately. If people know that

stays significant. Neither the course dummy nor the interaction term are significantly different from zero. Apparently, the type of math course does not add much information beyond math grade and university major.

20. For related models of narrow bracketing see Barberis et al. (2006), Rabin and Weizsäcker (2009), Hastings and Shapiro (2013), and Koch and Nafziger (2016).

21. We can also interpret this as a dual-system approach (Kahneman 2003): the intuitive system would like to spend more on tempting goods and the reasoning system sets rules and limits the budget for these kinds of goods. The ability of the reasoning system to set and enforce these rules could depend on the type of grant.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

they face decision costs, it can be optimal for them to bracket narrowly to save on decision costs.[22]

We refrain from modeling self-control problems or decision costs beyond them being a reason for bracketing narrowly. We rather treat narrow bracketing as a heuristic that people use even in situations where the underlying problem is not present. For example, in the lab experiment, temptation is not an issue, so even subjects who have problems with self-control in other situations should act rationally in this setting. We assume that subjects still bracket narrowly.[23]

Consumers decide on consumption out of a set of goods. The set of goods is partitioned into a hierarchy of mental accounts or groups of goods. This means that on the first (or top) level all goods are in one account.[24] On the second level, the set of goods is partitioned into a number of accounts $g_1, g_2, g_3, \ldots$ In lower levels, these accounts are again split into subaccounts, that is, account g_1 contains subaccounts $g_{1,1}, g_{1,2}, \ldots$, and so on. For example, goods could be split into food and nonfood goods on the second level, food is then split into food-away-from-home and food-at-home in the third level, food-away-from-home is split into meals-away-from-home and beverages-away-from-home in the fourth level, and so forth. All consumers have the same hierarchy of accounts.

The consumer reaches a consumption decision in two phases, a posting phase and a spending phase. In the posting phase, the consumer considers each budget coming from a different income source (e.g., wage earnings, capital gains, in-kind grants) and posts it to one account on a particular level of the account hierarchy. More precisely, we assume that a budget is posted to the lowest level at which it cannot be split over several accounts. This means that no decision is yet required on how to spend the budget on particular goods, budgets are just matched to accounts in the hierarchy. Cash budgets are thus posted to the top level, as the consumer would need to decide how to split it across different accounts on the next-lower, that is, second level. In-kind grants and labeled cash grants (like child benefits) are posted to the lowest level of the hierarchy on which there is still an account containing all goods for which the in-kind grant can be used. On this level, it is then posted to the account that contains these goods. In our field study, the cash budget B is posted to the top level. The bill voucher in the restaurant is posted to the food-away-from-home account (on level 3 in the example above) whereas the beverage voucher is posted to the beverages-away-from-home account (level 4). At the end of the posting phase, all budgets have been

22. This reasoning predicts that narrow bracketing is only used in situations that feature at least a minimal amount of complexity. If the decision situation is simple enough for the optimal solution to be directly apparent, we would not expect people to bracket narrowly.

23. Decision costs might still apply in the lab experiment.

24. If consumers bracket narrowly for self-control reasons, the top level could already be split into several accounts for different groups of goods, such that even cash budgets cannot be spent freely (see below). The amounts of these accounts would be set to reign in overconsumption on tempting goods. For the remainder of this section, we assume that there is only one account on the top level.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

posted to specific accounts that could be situated on different levels of the account hierarchy.

In the spending phase, the consumer considers the account hierarchy sequentially level by level, starting at the top level. For each account, he spends all budgets posted to this account rationally on all goods available in this account. However, he is naive in the sense that he takes consumption allocated on higher levels as given and only takes a share $\theta \in [0, 1]$ of any budgets further down the hierarchy into account. If $\theta = 1$, the consumer is fully rational and can reach the overall first-best consumption bundle. If $\theta = 0$, he ignores subsequent budgets. Intermediate values of θ correspond to partial narrow bracketing. Once all budgets are allocated, consumption is implemented as the sum of all allocations to each good.[25]

Consider first a rational consumer with $\theta = 1$ in the field study. The cash budget B is allocated to the top level and all goods are available in this account. In the Baseline treatment, she thus chooses the first-best bundle (c_t^*, c_{nt}^*, c_o^*) (point A in Figure 1a; c^* denotes again the rationally optimal consumption of a good if only B were available). In the Label and Cash treatments, she foresees that she will subsequently have a grant G at her disposal. In the Label treatment, she thus only allocates $p_t c_t^* - G$ to the targeted good. This preliminary allocation allows her to later spend G entirely on the targeted good while still reaching the first-best consumption bundle. In the Cash treatment, she similarly allocates less to the targeted and nontargeted goods at the top level such that she can ultimately reach the first-best bundle. She then goes down the hierarchy and on each level spends the posted budgets on the goods in the respective account. The beverage voucher is posted to an account that only contains the targeted good whereas the bill voucher account encompasses targeted and nontargeted good. Given her allocation at the top level, she can comply with the condition of the vouchers, spend G on the targeted good (or on targeted and nontargeted good in the Cash treatment) and overall reach the first-best consumption bundle (c_t^*, c_{nt}^*, c_o^*) (remember that $G \ll B$).

In the Label treatment of the lab experiment, a rational subject only allocates $p_t c_t^{**} - G$ on the top level to the targeted good (where c_t^{**} is the first-best consumption of the targeted good out of $B + G$, see point H in Figure 1b). She can thus subsequently spend G on the targeted good and end up with the first-best consumption bundle at point H. In the Cash treatment of the lab experiment, both B and G are as good as cash and thus posted to the top level. The consumer can thus directly reach the optimal bundle at point H.

Now consider a consumer with $\theta = 0$. In the field study, he also spends B rationally on the top level but ignores a potential subsequent G. In the Label treatment, he subsequently spends the beverage voucher on the targeted good in addition to c_t^* and ends up consuming $c_t^* + G/p_t$ (point C in Figure 1a). In the cash treatment, he spends the bill voucher on targeted and nontargeted good in addition to c_t^* and c_{nt}^* and ends up

25. Although the account hierarchy looks similar to a multistage budgeting structure or "utility tree" as pioneered by Strotz (1957) and Gorman (1959), multistage budgeting is quite different from our model as it does not have budgets on different levels of the hierarchy and no naivete.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

on the line between points C and D. In the Label treatment of the lab experiment, the consumer again allocates B in a way that is optimal only in the naive sense, and reaches (c_t^*, c_{nt}^*) (point E in Figure 1b). He then spends the in-kind grant on the targeted good, ending up at bundle $(c_t^* + G/p_t, c_{nt}^*)$ (point F). In the Cash treatment, B and G are as good as cash and thus posted to the top level account. The consumer can thus reach the first-best bundle at point H.

For intermediate values of θ, the consumer is able to move some of the in-kind grant to the nontargeted or to the other good (or rather to hold back spending out of B on the targeted good) but, for θ low enough, will fail to reach the first-best consumption bundles.

These predictions are in line with the main treatment differences in our two studies. In the lab experiment, 21% of subjects increase spending on the targeted good by the exact amount of the grant and the remaining subjects also show a (smaller) treatment effect. In the field study, the overall treatment effect is positive but smaller than the amount of the grant. These facts point to heterogeneity in θ, with a sizable share of subjects with $\theta = 0$ and a share of subjects with intermediate values of θ. If subjects bracket narrowly for cognitive reasons, we would expect subjects with lower cognitive skills to be more likely to violate fungibility as they will benefit most from reducing decision costs. This is indeed what we find in the lab experiment.

It does not follow directly from narrow bracketing that restaurant guests with a beverage voucher will use their increased consumption to buy more expensive wines rather than more wine. This finding is, however, easily rationalizable if we assume that cheap wine is an inferior good (at least for the level of consumed beverage volume in our field study, i.e., around 0.83 l or 29 fl oz). For this we have to assume that there are four instead of three goods: low-quality wine (which is cheap), high-quality wine (which is expensive), meals and nonrestaurant consumption and that cheap wine is an inferior good. If a guest decides to spend more on beverage consumption (perhaps due to narrow bracketing), the guest will thus shift more of their beverage consumption to higher quality, expensive wine, thereby increasing the price-per-volume ratio.

6. Conclusion

In this paper we pursue a dual research strategy by combining a field quasi-experiment and an incentivized laboratory experiment to test whether consumers treat money as fungible. Both studies 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 skills. This points to cognitive biases like narrow bracketing or mental accounting as mechanism underlying the 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 can lead to pronounced rent increases (see, e.g., Susin 2002; Fack 2006). Our results suggest that this effect is partly due to a violation of fungibility and could thus be mitigated by linking housing

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

benefits less saliently to rent payments to make it easier for tenants to treat this income source as fungible. There are, however, other situations in which it might be desirable if recipients violate fungibility. If the government 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 a grant or by replacing a cash grant with a nondistortionary in-kind grant, consumers who violate fungibility could be induced to buy more of the targeted good.

References

Barberis, Nicholas, Ming Huang, and Richard H. Thaler (2006). "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing." *American Economic Review*, 96(4), 1069–1090.

Beatty, Timothy, Laura Blow, Thomas Crossley, and Cormac O'Dea (2014). "Cash by Any Other Name? Evidence on Labelling From the UK Winter Fuel Payment." *Journal of Public Economics*, 118, 86–96.

Bénabou, Roland and Jean Tirole (2004). "Willpower and Personal Rules." *Journal of Political Economy*, 112, 848–886.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro (2013). "Who is 'Behavioral'?" *Journal of the European Economic Association*, 11, 1231–1255.

Blow, Laura, Ian Walker, and Yu Zhu (2012). "Who Benefits from Child Benefit?" *Economic Inquiry*, 50, 153–170.

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

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2013). "Salience and Consumer Choice." *Journal of Political Economy*, 121, 803–843.

Brodnick, R. J. and Malcolm James Ree (1995). "A Structural Model of Academic Performance, Socioeconomic Status, and Spearman's *g*." *Educational and Psychological Measurement*, 55, 583–594.

Cherry, Todd L., Peter Frykblom, and Jason F. Shogren (2002). "Hardnose the Dictator." *American Economic Review*, 92(4), 1218–1221.

Choi, James J., David Laibson, and Brigitte C. Madrian (2009). "Mental Accounting in Portfolio Choice." *American Economic Review*, 99(5), 2085–2095.

Conlisk, John (1996). "Why Bounded Rationality?" *Journal of Economic Literature*, 34, 669–700.

Deary, Ian, Steve Strand, Pauline Smith, and Cres Fernandes (2007). "Intelligence and Educational Achievement." *Intelligence*, 35, 13–21.

Fack, Gabrielle (2006). "Are Housing Benefit an Effective Way to Redistribute Income? Evidence from a Natural Experiment in France." *Labour Economics*, 13, 747–771.

Falk, Armin and Urs Fischbacher (2006). "A Theory of Reciprocity." *Games and Economic Behavior*, 54, 293–315.

Fischbacher, Urs (2007). "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10, 171–178.

Frank, Björn (1998). "Good News for Experimenters: Subjects do not Care about Your Welfare." *Economics Letters*, 61, 171–174.

Gorman, William M (1959). "Separable Utility and Aggregation." *Econometrica*, 27, 469–481.

Greiner, Ben (2004). "An Online Recruitment System for Economic Experiments." *Forschung und wissenschaftliches Rechnen*, 63, 79–93.

Groenland, Edward and Jan Schoormans (1994). "Comparing Mood-Induction and Affective Conditioning as Mechanisms Influencing Product Evaluation and Product Choice." *Psychology and Marketing*, 11, 183–197.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022

Hastings, Justine and Jesse Shapiro (2013). "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks." *Quarterly Journal of Economics*, 128, 1449–1498.
Heath, Chip and Jack Soll (1996). "Mental Budgeting and Consumer Decisions." *Journal of Consumer Research*, 23, 40–52.
Hines, James R. and Richard H. Thaler (1995). "The Flypaper Effect." *Journal of Economic Perspectives*, 9(4), 217–26.
Kahneman, Daniel (2003). "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93(5), 1449–1475.
Koch, Alexander K. and Julia Nafziger (2016). "Goals and Bracketing under Mental Accounting." *Journal of Economic Theory*, 162, 305–351.
Kooreman, Peter (2000). "The Labeling Effect of a Child Benefit System." *American Economic Review*, 90(3), 571–583.
Kőszegi, Botond and Matthew Rabin (2006). "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121, 1133–1165.
Kőszegi, Botond and Adam Szeidl (2013). "A Model of Focusing in Economic Choice*." *The Quarterly Journal of Economics*, 128, 53–104.
Kuhn, Peter, Peter Kooreman, Adriaan Soetevent, and Arie Kapteyn (2011). "The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery." *American Economic Review*, 101(5), 2226–2247.
Lehrl, Siegfried, Jürgen Merz, Georg Burkhard, and Bernd Fischer (1991). *Mehrfach-Wortschatz-Intelligenztest: MWT-A*. perimed Fachbuch-Verlagsgesellschaft, Erlangen.
Lewinsohn, Shai and Haim Mano (1993). "Multi-Attribute Choice and Affect: The Influence of Naturally Occurring and Manipulated Moods on Choice Processes." *Journal of Behavioral Decision Making*, 6, 33–51.
Milkman, Katherine L. and John Beshears (2009). "Mental Accounting and Small Windfall Gains." *Journal of Economic Behavior and Organization*, 71, 384–394, Mimeo.
O'Curry, Suzanne (1997). "Income Source Effects." *DePaul University Discussion Paper*.
Odean, Terrance (1998). "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53, 1775–1798.
Rabin, Matthew (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5), 1281–1302.
Rabin, Matthew and Georg Weizsäcker (2009). "Narrow Bracketing and Dominated Choices." *American Economic Review*, 99(4), 1508–1543.
Read, Daniel, George Loewenstein, and Matthew Rabin (1999). "Choice Bracketing." *Journal of Risk and Uncertainty*, 19, 171–197.
Strotz, Robert H. (1957). "The Empirical Implications of a Utility Tree." *Econometrica*, 25, 269–280.
Susin, Scott (2002). "Rent vouchers and the Price of Low-Income Housing." *Journal of Public Economics*, 83, 109–152.
Thaler, Richard H. (1980). "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior & Organization*, 1, 39–60.
Thaler, Richard H. (1985). "Mental Accounting and Consumer Choice." *Marketing Science*, 4, 199–214.
Thaler, Richard H. (1999). "Mental Accounting Matters." *Journal of Behavioral Decision Making*, 12, 183–206.
Thaler, Richard H. and Hersh M. Shefrin (1981). "An Economic Theory of Self-Control." *Journal of Political Economy*, 89, 392–410.
Tversky, Amos and Daniel Kahneman (1981). "The Framing of Decisions and the Psychology of Choice." *Science*, 211, 453–458.

Supplementary Data

Supplementary data are available at *JEEA* online.

Downloaded from https://academic.oup.com/jeea/article/15/1/99/2732085 by Vienna University Library user on 23 February 2022