Heterogeneity of Loss Aversion and Expectations-Based Reference Points: Replication Pre-Analysis Plan

Lorenz Goette* Thomas Graeber† Alexandre Kellogg‡

University of Bonn University of Bonn UC San Diego

Charles Sprenger§

UC San Diego

This Version: July 3, 2018

JEL classification: D81, D84, D12, D03
Keywords: Reference-Dependent Preferences, Rational Expectations, Personal Equilibrium, Endowment Effect, Expectations-Based Reference Points

*University of Bonn, Institute for Applied Microeconomics, Adenauerallee 24 - 42, 53113 Bonn, Germany; lorenz.goette@uni-bonn.de

†University of Bonn, Bonn Graduate School of Economics, Kaiserstraße 1, 53113 Bonn, Germany; graeber@uni-bonn.de

‡Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093; alexkellogg@ucsd.edu

§Department of Economics and Rady School of Management, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093; csprenger@ucsd.edu

1 Introduction

The goal of this experiment is to replicate the results contained in Goette, Graeber, et al. (2018)[1], which shed light on the extent to which heterogeneity of loss aversion contributes to the mixed experimental evidence of Kőszegi and Rabin (2006)'s (KR) expectations-based reference dependence model. Previous literature has found conflicting evidence (Abeler et al., 2011; Ericson and Fuster, 2011; Gneezy et al., Forthcoming; Goette, Harms, et al., 2016; Heffetz and List, 2014), which we believe is in large part attributable to the lack of emphasis placed on the heterogeneity in loss aversion between agents. In particular, Goette, Graeber, et al. (2018) show that after structurally measuring and controlling for levels of loss aversion, the introduction of a forced exchange treatment a la Goette, Harms, et al. (2016) had directionally opposite impacts on the loss loving and loss averse agents. We propose herein a pre-analysis plan to replicate this experiment with this hypothesis in mind, in order to confirm that our results are indicative of a robust phenomenon.

2 Experiment Design, Hypotheses, and Power

2.1 Design Overview

As this is an exact replication, the design of the experiment will follow Goette, Graeber, et al. (2018), so we will be relatively brief in our description of the experiment. A more thorough description of the design is contained in Goette, Graeber, et al. (2018), and the instructions from that experiment are contained in Appendix A below. The experiment has two stages and there are two bundles of goods, Bundle 1 (USB stick or Pen Set) and Bundle 2 (Picnic Mat or Thermos). In Stage 1 we randomly assign half the subjects to Bundle 1 (USB stick or Pen Set) and Bundle 2 (Picnic Mat or Thermos) at the session level, and within each session (i.e., bundle), randomly assign half of the subjects to be

[1] `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3170670`

endowed with each good[2]. Thus, on aggregate in Stage 1, about 25% of subjects would be endowed with any given item. Drawing the requisite number of subjects from the subject pool of the BonnEconLab at University of Bonn, we begin by showing subjects both items in their Stage 1 bundle on a projector before directing them to a private booth in which they will find their endowed item. Next, subjects will be asked to fill out a survey eliciting their general mood (0-10), as well as their liking (0-8) and wanting (0-8) for each good in the bundle. They will then be informed that half of them will be forced to exchange their endowed item for the other item in the bundle, with the randomization determined by lottery balls numbered 1-20 in a rotating drum. Immediately after the confiscation, subjects will once again be asked to "Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?" by adjusting a slider ranging from 0 (labelled "Unhappy, Angry, Unsatisfied, Sad, Desperate") to 10 (labeled "Happy, Thrilled, Satisfied, Content, Hopeful"). At this point, Stage 1 of the experiment is complete and subjects will be asked to return to the presentation room wherein Stage 2 will commence.

In Stage 2, subjects will be exposed to the other Bundle, mapping those endowed in Stage 1 with USB sticks to Picnic Mats and those endowed in Stage 1 with Pen sets to Thermoses. After seeing both goods in their Stage 2 bundle projected in the presentation room, subjects will re-enter their cubicle, discover their Stage 2 endowment, and read the treatment-specific instructions on the computer. At this point, the subjects receive a different set of instructions depending on their assignment to the Baseline Condition (50%) or Forced Exchange Treatment (50%), with treatment status randomized within-session[3]. For the Baseline, subjects are instructed that they will have the option to choose to exchange their endowed object for the other good in the bundle, or keep their endowment;

[2]There are slight differences between the goods used in the original experiment compared to our replication described herein. For example, the USB sticks are now 16GB instead of 8GB because the company no longer makes the same USB stick with 8GB, but it looks identical. Similarly, the picnic mat has the same functionality but a somewhat different print. These differences should not affect our results because our taxonomies are designed to address preferences via ratings, and failure to categorize subjects into their true types would adversely affect our predictive abilities in stage 2.

[3]In the original experiment, 60% of subjects were randomized into the Forced Exchange Treatment. However, since this reduces our power to detect treatment effects by raising our standard errors, we will

whatever choice they make will be the final decision. In the Forced Exchange Treatment, subjects are instead told that regardless of their choice, subjects will be forced to exchange 50% of the time based on the same drum randomization device as in Stage 1; subjects who choose to exchange voluntarily will have that choice implemented, while those who choose to keep will be forced to exchange half of the time. After verifying that they've understood the instructions, their exchange decision will be elicited, the randomization will be implemented, and exchange takes place with the experimenter clearing the market. Finally, after the decision has been elicited, we will have subjects fill out questionnaires, allowing us to collect sociodemographic information as well as unincentivized measures of cognitive skills. The specific questions are included in the appendix, along with the 5 Raven matrices we will use. Although this was not part of the design of the original experiment, it allows us to include control variables in our analyses to include statistical precision for our hypothesis tests described below, and has no meaningful interaction with our decision of interest.

2.2 Hypotheses and Analyses

Following the theory laid out in Goette, Graeber, et al. (2018), we will provide evidence on whether agents follow the predictions of expectations-based reference dependent as formalized by KR (2006). If subjects behave according to KR (2006), we expect differential responses to the forced exchange treatment introduced above. Specifically, Goette, Graeber, et al. (2018) demonstrate that loss averse agents (characterized by $\lambda > 1$ in the paper) are predicted to respond to the treatment by increasing their willingness to exchange compared to the baseline treatment, and loss loving agents ($\lambda < 1$) are predicted to be less willing to exchange under forced exchange compared to baseline.

To determine whether subjects are classified as loss averse, loss neutral, or loss loving, we follow the structural and reduced form approaches introduced by Goette, Graeber, et al.

instead randomize half of the subjects into treatment for this replication. This is a cosmetic difference in the two designs, and has no impact beyond reducing our standard errors.

(2018). For the structural classification, we use the *personal equilibria* conditions defined by KR (2006) and described in Goette, Graeber, et al. (2018) to estimate the relative utility of goods within a bundle, a loss aversion parameter, and an indifference threshold for all subjects who faced a given bundle in Stage 1. Using the aggregate ratio of relative utility, we solve for bounds on λ for each agent based on their liking preference, with a lower bound estimated for subjects who prefer their endowment, an upper bound for those who prefer the other good, and both an upper and lower bound for those who are indifferent. Subjects whose bounds include one are categorized as loss neutral, whereas those whose bounds are strictly above one are labelled loss averse, and those with bounds strictly below one are defined as loss loving. For robustness, we also use a reduced form taxonomy which categorizes agents who rate their own good higher as loss averse, those who rate the two goods the same as loss neutral, and those who rate the other good higher as loss loving. The results of this structural approach will be presented in the same way as in Goette, Graeber, et al. (2018) by filling in Table 2 below with the relevant findings.

Once these taxonomies are established, we provide evidence of their predictive validity by running a regression of the change in elicited mood rating before and after forced exchange in Stage 1 on whether the subject lost their Stage 1 endowment and a constant. We are particularly interested in the coefficients on the *Lost Stage 1 Endowment*$_i$ dummy variable, which we expect to be negative for the loss averse sub-population and positive for the subjects classified as loss loving. The results from this regression will be presented by filling out Table 3 below.

Next, we run the following Linear Probability Model on the aggregate data as well as on each loss averse category separately to test our hypothesis of the impact of treatment:

$$VoluntaryExchange_i = \beta_0 + \beta_1 \times Treatment_i + \epsilon_i.$$

β_0 tells us how willing to exchange agents in the Baseline condition are, with a willingness to exchange below 0.5 indicating an endowment effect (which we test using a standard

t-test). The effect of the treatment, β_1, provides our test of KR (2006), with loss averse agents expected to trade more ($\beta_{1,LA} > 0$) and loss loving agents expected to trade less ($\beta_{1,LL} < 0$) than baseline. To test whether loss averse agents and loss loving agents respond differentially to treatment, we use SUR to run a χ^2 test of $\beta_{1,LA} = \beta_{1,LL}$. The results of these hypotheses will be presented in the same format as Table 5 in Goette, Graeber, et al. (2018) by entering the new values into the blank Table 4 presented below. Since we are collecting sociodemographic information as well as data on cognitive skills, we will run the same regressions including these controls to increase statistical precision, and present them as in Table 5.

Finally, as this is an exact replication, we will additionally recreate Table 4 pooling both our original sample of 607 and this new sample of about 400 subjects and create Table 6. The Full Sample (Column 1) analysis will contain approximately 1000 subjects, while the analysis on the Loss Averse, Loss Neutral, and Loss Loving subsamples (Column 2-4) will include the subjects from both samples who are classified as such. In particular, Column 2 will contain the 217 (285) subjects who are structurally (reduced form) classified as loss averse in the original sample of 607 as well as any additional subjects who are classified as loss averse in our new sample. The same applies to Columns 3 and 4. We take this approach instead of structurally estimating the parameters on the full sample of 1000 subjects because of the slight differences in the USB stick and picnic mats used in the original experiment compared to this replication.

2.3 Power Calculations: Sample Size

Using the results from the original experiment, we can first get a sense of the sample size required to ensure an adequately powered test of our main hypotheses. Table 5 of Goette et al. (2018) shows a 40 percentage point difference between the treatment coefficient in the loss averse versus loss loving agent subpopulations, which is significant at the 1% level. The test between these two coefficients can equivalently be studied through a differences-

in-differences specification of the form

$$VoluntaryExchange_i = \beta_0 + \beta_1 Treatment_i + \beta_2 LossNeutral_i + \beta_4 LossAverse_i + \beta_5 \times LossNeutral_i \times Treatment_i + \beta_6 LossAverse_i \times Treatment_i + \epsilon_i,$$

where we are interested in the power for testing the hypothesis $\beta_6 = 0$. The results of this regression using the data in Goette, Graeber, et al. (2018) are presented in Table 1, indicating that a $\hat{\beta}_6 = 0.406$ is a reasonable starting point for our power calculations.

Table 1: **Differences in Differences Results**

	(1) Voluntary Exchange (=1)
Constant (Loss Loving (=1))	0.429***
	(0.0665)
Loss Neutral (=1)	-0.0671
	(0.0850)
Loss Averse (=1)	-0.0988
	(0.0824)
Forced Exchange (Loss Loving × Forced Exchange)	-0.248***
	(0.0775)
Loss Neutral × Forced Exchange	0.275***
	(0.102)
Loss Averse × Forced Exchange	0.406***
	(0.102)
Observations	607
R^2	0.0388
t-Test $\beta_6 = 0$	15.78

Notes: OLS regression with robust standard errors in parentheses.
Loss Aversion is calculated by looking at the bounds for $\hat{\lambda}$ given individuals decisions.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We begin by sampling observations with replacement from our initial population of 607 to determine how the Minimum Detectable Effect (MDE) varies with sample size. First, we draw N observations with replacement from the population in Goette, Graeber, et al. (2018), with 50% sampled from the treatment group (as half of subjects in the repli-

cation will be treated). Then, we run the differences-in-differences specification on this bootstrapped sample to obtain $(\hat{\beta}_6, \hat{\sigma}_{\beta_6})$, the estimated coefficient on our parameter of interest as well as its standard error. Using the equation for the minimum detectable effect, $MDE = (t_{\alpha/2} - t_{1-\kappa})\sigma_{\hat{\beta}}$, under 80% power ($\kappa = 0.8$), an $\alpha = 0.05$, with $\sigma_{\hat{\beta}}$ as the standard error of $\hat{\beta}_6$, we compute the MDE as $2.8\hat{\sigma}_{\beta_6}$ for each of these sampling sizes[4].

Figure 1 shows how our MDE relates to our bootstrapped sampling size, with a sample of about 300 subjects yielding 80% power to detect our effect size of 0.406. However, since it's plausible that the estimated effect size in Goette, Graeber, et al. (2018) is on the upper end of the distribution of true effect sizes, we conservatively opt for an ideal sample of around 400 subjects, which would leave us adequately powered to detect effect sizes on the order of 0.34.

To verify that this sample size will provide the requisite power, we additionally use G*Power (Faul et al., 2007), a program created to help researchers compute power analyses for various tests. In particular, we select the option for "linear bivariate regression with two groups, testing the difference between slopes" and choose "A priori: Compute required sample size" in G*Power, which is listed under the menu for t-tests. Following the user manual, this procedure is intended for computing the power to test slope coefficients from regressions using the same outcome and control variables but within different groups.[5] In our case, the appropriate formulation is to test the null $\beta_{1,LA} - \beta_{1,LS} = 0$, where the coefficients come from the regressions

$$VoluntaryExchange_{i,LA} = \beta_{0,LA} + \beta_{1,LA} Treatment_{i,LA} + \epsilon_i,$$

and

$$VoluntaryExchange_{i,LS} = \beta_{0,LS} + \beta_{1,LS} Treatment_{i,LS} + \epsilon_i.$$

[4]Slide 16 of `https://www.povertyactionlab.org/sites/default/files/L2-Power%20Calculation.pdf` contains the formula used here.

[5]`http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf`

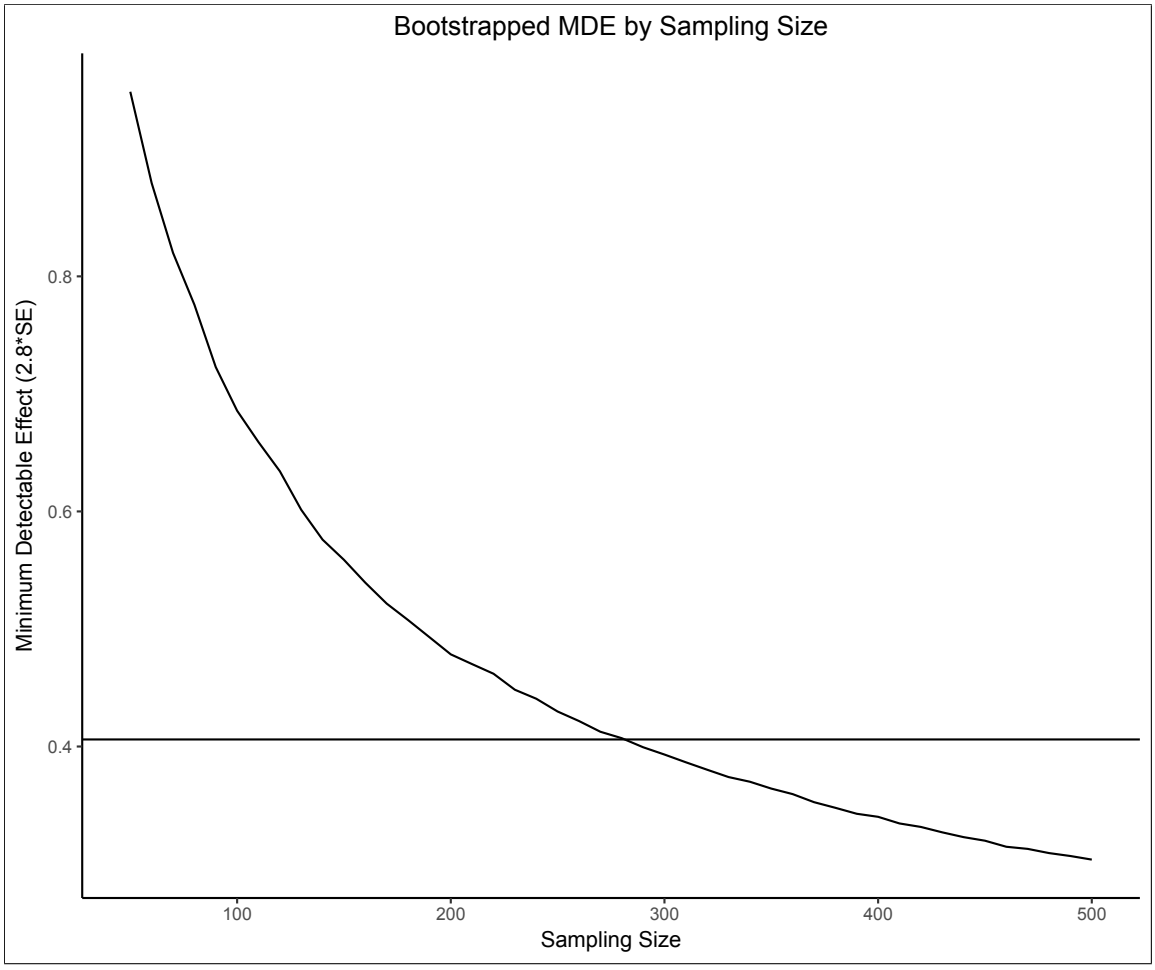


Figure 1: **Minimum Detectable Effect Size and Sample Size.** MDE with 80% power at the 5% level for a two-sided test is calculated as $2.8 \times \hat{\sigma}_{\beta_6}$. Observations are sampled from the population of 607 from Goette, Graeber, et al. (2018) with replacement, with 50% sampled from the treatment group, representing the proportion of treated subjects in the replication sample. Horizontal line at $y = 0.406$ indicates the differences-in-differences estimate from Goette, Graeber, et al. (2018).

We specify a two tailed test with a $|\Delta slope| = 0.406$, an $\alpha = 0.05$, power of 0.8, an allocation Ratio of $N_2/N_1 = 150/217 = 0.69$, and standard deviation of $X_1 = 0.5$, and of $X_2 = 0.49$. Next, to calculate the pooled standard deviation of the residual we follow the formula outlined in the manual,

$$\sigma = \sqrt{\frac{n_1\sigma^2_{r_{LA}} + n_2\sigma^2_{r_{LS}}}{n_1 + n_2}},$$

where $\sigma^2_{r_i}$ is the Mean Square Error of the regression for in the given group, ($\sigma^2_{r_{LA}} = 0.240$, $\sigma^2_{r_{LA}} = 0.187$). This computation yields a $\sigma = 0.467$. Using all these values, we allow G*Power to calculate a required $N_1 = 107$ and an $N2 = 73$. Because we cannot recruit loss averse or loss loving agents directly, we estimate the total sample size we would need to get approximately 107 loss averse and 73 loss loving subjects. We do this by dividing the desired number of loss averse subjects by the fraction of the population in Goette, Graeber, et al. (2018) that is loss averse, $107 \times \frac{1}{\frac{217}{607}} = 300$ and analogously for loss loving agents $73 \times \frac{1}{\frac{150}{607}} = 295$. We take the maximum of these two numbers, 300 subjects, to give us the best chance of reaching the desired number of loss averse and loss loving subjects.

Since we may not get the same magnitude for the treatment effect in the replication, we additionally run the analysis with $|\Delta slope| \in \{0.26, 0.3, 0.34, 0.38, 0.42, 0.46\}$ to see the required sample sizes to reach 80% power under these different effect sizes. Because G*Power tells us how many loss averse and loss loving subjects we require, we compute the total sample sizes implied by these numbers and take the maximum of the two to be conservative. More formally, recognizing the fixed allocation ratio of 0.69, we observe that the total sample from G*Power will consist of 41% loss loving and 59% loss averse subjects. However, these figures do not take into account the fact that we expect only 31% and 35% of the total sample to be loss loving and loss averse, with the remaining fraction classified as loss neutral. Thus, we re-weight these percentages, dividing them by the assumed fraction of the total population that is loss loving or loss averse, respectively, to see how many subjects we would actually need to get the desired number of loss averse and loss loving agents as calculated by G*Power. Taking the maximum of these two re-weighted proportions suggests that we need 1.69 times the number of "Total Subjects" calculated by G*Power, which should leave us with about the right number of loss averse subjects and a surplus of loss loving agents (assuming the allocation ratio is constant). The power for the re-weighted sample sizes and effect sizes is plotted in Figure 2, which indicates that the power of the test is about 80% with a sample size of 400 subjects and an effect size of 0.34, supporting our conclusion from the resampling approach.

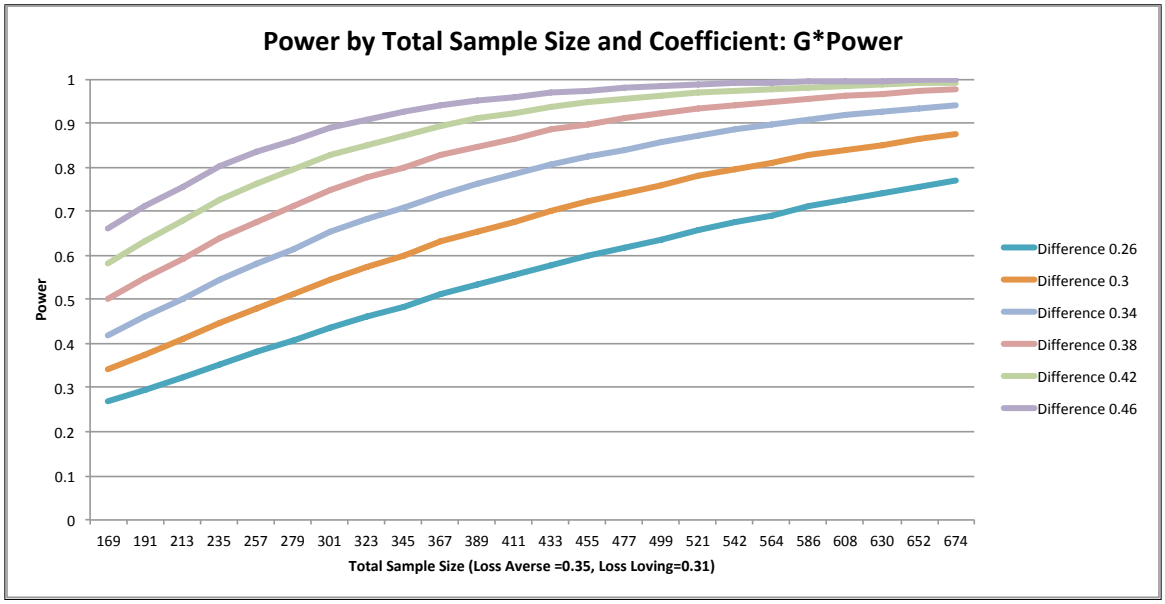


Figure 2: **Power by Sample Size and Effect Size.** For each effect size, $|\Delta slope|$, G*Power computes the number of loss averse and loss loving subjects required to reach 80% power, fixing the level of the test as well as the allocation ratio, and the standard deviations. Total sample size is calculated as the number of observations that yields at least the required number of loss averse and loss loving subjects, where we assume that 35% of the subjects are loss averse, 31% are loss loving, and the rest are loss neutral as per Goette, Graeber, et al. (2018). In practice, this takes the form $1.69 \times N_{gpower}$ as described in the text.

As an additional check, we use the G*Power option "Post hoc: compute acheived power" to determine the power to reject $\beta_6 = 0$ in Goette, Graeber, et al. (2018), where the full sample size was 607, with 217 loss averse and 150 loss loving agents. Plugging in all the same parameters as before as well as the number of loss averse and loss loving agents, we see that the two-tailed test of 5% level had 98% power to reject the null. Thus, the fact that the required sample size for 80% power under the same level is smaller than the sample in Goette, Graeber, et al. (2018) is in line with our expectation.

All of this suggests that 300-400 subjects would leave us adequately powered to test our hypothesis of interest assuming the true effect size close to 0.406 and the allocation ratio is fixed at 0.69. Because uncertainty relating to both of these factors could leave us underpowered (if the ratio changes in a way that decreases either the loss averse or the loss

loving fraction, or if the effect size is smaller), we opt for caution and aim for a sample of 400 subjects, which we estimate would yield 143 loss averse and 99 loss loving subjects.

3 Blank Tables

Table 2: **Aggregate Parameter Estimates**

	(1)	(2)	(3)	(4)
	Estimate	(Std. Error)	Estimate	(Std. Error)
	Bundle 1		*Bundle 2*	
Loss Aversion:				
$\hat{\lambda}$		()		()
Utility Values:				
$\hat{X}_1$ *(Pen Set)*		()		
$\hat{Y}_1$ *(USB Stick)*	1	-		
$\hat{X}_2$ *(Picnic Mat)*				()
$\hat{Y}_2$ *(Thermos)*			1	-
Discernibility:				
$\hat{\delta}$		()		()

Notes: Maximum likelihood estimates. Robust standard errors in parentheses.

Table 3: **Preference Types and Subjective Experience**

	(1)	(2)	(3)	(4)
	Dependent Variable: Δ *Stage 1 Happiness*			
	Full Sample	Loss Averse	Loss Neutral	Loss Loving
Panel A: Structural Bounds Taxonomy				
Lost Stage 1 Endowment				
	()	()	()	()
Constant				
	()	()	()	()
R-Squared				
# Observations				
Panel B: Reduced Form Taxonomy				
Lost Stage 1 Endowment				
	()	()	()	()
Constant				
	()	()	()	()
R-Squared				
# Observations				

Notes: Ordinary least squares. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Panel A: taxonomy of types based on structural bounds. Panel B: taxonomy of types based on reduced form rating statements.

Table 4: **Exchange Behavior and Probabilistic Forced Exchange**

	(1)	(2)	(3)	(4)
	Dependent Variable: Exchange (=1)			
	Full Sample	Loss Averse	Loss Neutral	Loss Loving
Panel A: Structural Bounds Taxonomy				
Forced Exchange				
	()	()	()	()
Baseline Exchange (Constant)				
	()	()	()	()
R-Squared				
# Observations				
H_0: No Baseline Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: No Forced Ex. Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: Baseline (col. 2) = Baseline (col. 4)				$\chi^2(1)=$
				$(p=)$
H_0: Forced Ex. (col. 2) =Forced Ex(col. 4)				$\chi^2(1)=$
				$(p=)$
Panel B: Reduced Form Taxonomy				
Forced Exchange				
	()	()	()	()
Baseline Exchange (Constant)				
	()	()	()	()
R-Squared				
# Observations				
H_0: No Baseline Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: No Forced Ex. Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: Baseline (col. 2) = Baseline (col. 4)				$\chi^2(1)=$
				$(p=)$
H_0: Forced Ex. (col. 2) =Forced Ex(col. 4)				$\chi^2(1)=$
				$(p=)$

Notes: Ordinary least square regression. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Null hypotheses tested for 1) zero baseline endowment effect, regression (Constant = 0.5); 2) zero forced exchange endowment effect (Constant + Forced Exchange = 0.5); 3) Identical baseline behavior across loss averse and loss loving agents (Constant (col. 2) = Constant (col. 4)); 4) Identical treatment effects of forced exchange across loss averse and loss loving agents (Forced Exchange (col. 2) = Forced Exchange (col. 4)). Hypotheses 3 and 4 tested via seemingly unrelated regression. Panel A: taxonomy of types based on structural bounds. Panel B: taxonomy of types based on reduced form rating statements.

Table 5: **Exchange Behavior and Probabilistic Forced Exchange with Controls**

	(1)	(2)	(3)	(4)
	Dependent Variable: Exchange (=1)			
	Full Sample	Loss Averse	Loss Neutral	Loss Loving
Panel A: Structural Bounds Taxonomy				
Forced Exchange				
	()	()	()	()
Baseline Exchange (Constant)				
	()	()	()	()
R-Squared				
# Observations				
Controls	Yes	Yes	Yes	Yes
H_0: No Baseline Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: No Forced Ex. Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: Baseline (col. 2) = Baseline (col. 4)				$\chi^2(1)=$
				$(p=)$
H_0: Forced Ex. (col. 2) =Forced Ex(col. 4)				$\chi^2(1)=$
				$(p=)$
Panel B: Reduced Form Taxonomy				
Forced Exchange				
	()	()	()	()
Baseline Exchange (Constant)				
	()	()	()	()
R-Squared				
# Observations				
Controls	Yes	Yes	Yes	Yes
H_0: No Baseline Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: No Forced Ex. Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: Baseline (col. 2) = Baseline (col. 4)				$\chi^2(1)=$
				$(p=)$
H_0: Forced Ex. (col. 2) =Forced Ex(col. 4)				$\chi^2(1)=$
				$(p=)$

Notes: Ordinary least square regression. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Null hypotheses tested for 1) zero baseline endowment effect, regression (Constant = 0.5); 2) zero forced exchange endowment effect (Constant + Forced Exchange = 0.5); 3) Identical baseline behavior across loss averse and loss loving agents (Constant (col. 2) = Constant (col. 4)); 4) Identical treatment effects of forced exchange across loss averse and loss loving agents (Forced Exchange (col. 2) = Forced Exchange (col. 4)). Hypotheses 3 and 4 tested via seemingly unrelated regression. Panel A: taxonomy of types based on structural bounds. Panel B: taxonomy of types based on reduced form rating statements.

Table 6: **Combined Studies Exchange Behavior and Probabilistic Forced Exchange**

	(1)	(2)	(3)	(4)
	Dependent Variable: Exchange (=1)			
	Full Sample	Loss Averse	Loss Neutral	Loss Loving
Panel A: Structural Bounds Taxonomy				
Forced Exchange				
	()	()	()	()
Baseline Exchange (Constant)				
	()	()	()	()
R-Squared				
# Observations				
Study Fixed Effects	Yes	Yes	Yes	Yes
H_0: No Baseline Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: No Forced Ex. Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: Baseline (col. 2) = Baseline (col. 4)				$\chi^2(1)=$
				$(p=)$
H_0: Forced Ex. (col. 2) =Forced Ex(col. 4)				$\chi^2(1)=$
				$(p=)$
Panel B: Reduced Form Taxonomy				
Forced Exchange				
	()	()	()	()
Baseline Exchange (Constant)				
	()	()	()	()
R-Squared				
# Observations				
Study Fixed Effects	Yes	Yes	Yes	Yes
H_0: No Baseline Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: No Forced Ex. Endowment Effect	$F=$	$F=$	$F=$	$F=$
	$(p=)$	$(p=)$	$(p=)$	$(p=)$
H_0: Baseline (col. 2) = Baseline (col. 4)				$\chi^2(1)=$
				$(p=)$
H_0: Forced Ex. (col. 2) =Forced Ex(col. 4)				$\chi^2(1)=$
				$(p=)$

Notes: Ordinary least square regression. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Null hypotheses tested for 1) zero baseline endowment effect, regression (Constant = 0.5); 2) zero forced exchange endowment effect (Constant + Forced Exchange = 0.5); 3) Identical baseline behavior across loss averse and loss loving agents (Constant (col. 2) = Constant (col. 4)); 4) Identical treatment effects of forced exchange across loss averse and loss loving agents (Forced Exchange (col. 2) = Forced Exchange (col. 4)). Hypotheses 3 and 4 tested via seemingly unrelated regression. Panel A: taxonomy of types based on structural bounds. Panel B: taxonomy of types based on reduced form rating statements.

4 Instructions and Material Presented to Participants

All instructions and information presented to participants have been translated from German to English.

4.1 Images of Objects Presented to Participants

The following images were projected to the wall of the lecture room at the beginning of the respective stage. For the displayed example, the Stage 1 bundle consisted of the USB stick and erasable pens, but this was counter-balanced at the session level.

Figure 3: **Image 1 Projected on the Wall to Present Objects.** For Stage 1 with goods bundle consisting of USB stick and erasable pens.

Figure 4: **Image 2 Projected on the Wall to Present Objects.** For Stage 2 with goods bundle consisting of thermos and picnic mat.

4.2 Instructions (computer-based)

Welcome to part 1 of 2 in this experiment!

Please close the curtain of you cabin and read the following information. All computer entries that you make in this experiment are fully anonymous and cannot be traced back to you. Speed is not important at any point in this experiment. Please always take sufficient time to read and understand the instructions.

The [USB stick / erasable pens / thermos / picnic mat] now belongs to you. You can touch and inspect it at any time. However, please do not yet open the packaging and do not use the object yet. The two objects presented to you ([USB stick and erasable pens / thermos and picnic mat]) have been randomly allocated to the cabins in equal

quantities. Your cabin number was also randomly determined based on your choice of seat in the presentation room.

Please click on OK when you have read these information. If you have questions, please call an experimenter.

Please answer the questions.

[USB stick / thermos]

How much do you like this product?

How much would you want to have this product?

[Erasable pens / picnic mat]

How much do you like this product?

How much would you want to have this product?

Please read the following information carefully.

The experimenter will soon draw a random number between 1 and 20 using a lotto drum. The drawn number will then be announced loudly. If the drawn number is a number [from 11 to 20 / from 1 to 10], your [USB stick / erasable pens / thermos / picnic mat] will be taken away from you and you instead receive [USB stick / erasable pens / thermos / picnic mat]. If the drawn number is a number [from 1 to 10 / from 11 to 20], you will keep your [USB stick / erasable pens / thermos / picnic mat] and nothing happens. After the number has been drawn and the exchange of objects has taken place (if applicable), nothing else happens in this part of the experiment. You can then keep your object for good.

Please only confirm below once you have understood everything. If you have questions, please call the experimenter and wait until he comes to your cabin.

[Mood elicitation 1]

Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?

"Unhappy, Angry, Unsatisfied, Sad, Desperate" – "Happy, Thrilled, Satisfied, Content, Hopeful"

The time has come. Please wait until the number has been drawn.

Remember: If the drawn number is a number [from 11 to 20 / from 1 to 10], your [USB stick / erasable pens / thermos / picnic mat] will be taken away from you and you instead receive [USB stick / erasable pens / thermos / picnic mat]. If the drawn number is a number [from 1 to 10 / from 11 to 20], you will keep your [USB stick / erasable pens / thermos / picnic mat].

The drawn number is [1 / 2 / ... / 20].

This number is a number [from 1 to 10 / from 11 to 20]. Therefore [you can keep your [USB stick / erasable pens / thermos / picnic mat] / your [USB stick / erasable pens / thermos / picnic mat] will be taken away from you and you instead receive [USB stick / erasable pens / thermos / picnic mat]]. Please wait while the experimenter carries out the exchange in all cabins.

[Mood elicitation 2 and control question.]

Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?

"Unhappy, Angry, Unsatisfied, Sad, Desperate" – "Happy, Thrilled, Satisfied, Content, Hopeful"

Regarding the lottery draw, that has just taken place: What was the probability (in per-

cent) that you would lose your initial object? Please enter a number between 0 and 100.

Part 1 of the experiment is over!

Please follow the instructions.

- Memorize your cabin number.
- You can no go back to the presentation room.
- Please leave your [USB stick / erasable pens / thermos / picnic mat] in the cabin. You will be back in the same cabin in a few minutes.
- Remember: The object now belongs to you for good and you will take it away from this experiment.

Welcome to part 2 in this experiment!

Please close the curtain of you cabin and read the following information. The [USB stick / erasable pens / thermos / picnic mat] now also belongs to you. You can touch and inspect it at any time. However, please do not yet open the packaging and do not use the object yet. The two objects presented to you for part 2 ([USB stick and erasable pens / thermos and picnic mat]) have again been randomly allocated to the cabins in equal quantities.

Please click on OK when you have read these information. If you have questions, please call an experimenter.

[Instructions Stage 2 – ONLY BASELINE (p=0.0)]

Please read the following information carefully. The [USB stick / erasable pens / thermos / picnic mat] from part 2 of the experiment now belongs to you and you can keep it for

good. If you like, you can exchange your [USB stick / erasable pens / thermos / picnic mat] voluntarily for [USB stick / erasable pens / thermos / picnic mat]. Whichever way you decide, your choice is final and you will take your selected object with you from this experiment.

Please only confirm below once you have understood everything. If you have questions, please call the experimenter and wait until he comes to your cabin.

[Instructions Stage 2 – ONLY FORCED EXCHANGE (p=0.5)]

Please read the following information carefully. You have received a new object in part 2 of the experiment ([USB stick / erasable pens / thermos / picnic mat]). You will soon get the opportunity to exchange your [USB stick / erasable pens / thermos / picnic mat] voluntarily for [USB stick / erasable pens / thermos / picnic mat].

If you decide to exchange, you will receive [USB stick / erasable pens / thermos / picnic mat] as requested for your [USB stick / erasable pens / thermos / picnic mat] and you can then keep your [USB stick / erasable pens / thermos / picnic mat] for good. The experiment is then finished.

If you decide against an exchange, there will be a probability of 50% that the exchange will be forced anyways and you have to exchange nevertheless.

Concretely, the following happens in the case that you decide against a voluntary exchange: The experimenter will draw a random number between 1 and 20 using a lotto drum (as in part 1 of the experiment). The drawn number will then be announced loudly. If the drawn number is a number [from 11 to 20 / from 1 to 10], your [USB stick / erasable pens / thermos / picnic mat] will be taken away from you and you instead receive [USB stick / erasable pens / thermos / picnic mat]. If the drawn number is a number [from 1 to 10 / from 11 to 20], you will keep your [USB stick / erasable pens / thermos / picnic mat] and nothing happens. After the number has been drawn and the exchange of objects has taken place (if applicable), nothing else happens in this part of the experiment. You can then keep your object for good.

Please only confirm below once you have understood everything. If you have questions, please call the experimenter and wait until he comes to your cabin.

[Mood elicitation 3]

Before you get the opportunity to exchange your object, please answer the following questions about how you currently feel. Which expressions better apply to you at the moment? "Unhappy, Angry, Unsatisfied, Sad, Desperate" – "Happy, Thrilled, Satisfied, Content, Hopeful"

Do you want to exchange your [USB stick / erasable pens / thermos / picnic mat] for a [USB stick / erasable pens / thermos / picnic mat]?

Yes, I want to exchange.

No, I do not want to exchange.

[ONLY BASELINE (p=0.0)]

You have decided [for / against] a voluntary exchange. Please wait while the experimenter carries out the exchange in all cabins.

[ONLY FORCED EXCHANGE (p=0.5)]

You have decided [for / against] a voluntary exchange. Please wait while the experimenter carries out the exchange in all cabins.

[ONLY NON-TRADERS] After this, it will be determined whether you have to exchange anyways.

[ONLY TRADERS] Please wait until the experiment continues. A random number will now be drawn for those who decided against a voluntary exchange. After that the experiment continues for you.

[ONLY NON-TRADERS] Remember: If the drawn number is a number [from 11 to 20 / from 1 to 10], your [USB stick / erasable pens / thermos / picnic mat] will be taken away from you and you instead receive [USB stick / erasable pens / thermos / picnic mat]. If the drawn number is a number [from 1 to 10 / from 11 to 20], you will keep your [USB stick / erasable pens / thermos / picnic mat].

[ONLY NON-TRADERS]

The drawn number is [1 / 2 / ... / 20]

This number is a number [from 1 to 10 / from 11 to 20]. Therefore [you can keep you [USB stick / erasable pens / thermos / picnic mat] / your [USB stick / erasable pens / thermos / picnic mat] will be taken away from you and you instead receive [USB stick / erasable pens / thermos / picnic mat]. Please wait while the experimenter carries out the exchange in all cabins.

[Mood elicitation 4]

Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?

"Unhappy, Angry, Unsatisfied, Sad, Desperate" – "Happy, Thrilled, Satisfied, Content, Hopeful"

[Questionnaire]

What is your gender?

What is your age?

How long have you been in education? Please indicate years including school years.

What is your highest school degree?

What is your highest educational degree?

Are you currently enrolled as a student?

(Only if student) Which subject do you study?

(Only if student) Generally speaking, how is your knowledge in economics?

(Only if student) On average, how many hours per week do you spend on your studies and, if applicable, on part-time jobs?

(Only if non-student) What is your current field of work? Please describe as short and precise as possible.

(Only if non-student) Generally speaking, how is your knowledge in economics?

(Only if non-student) On average, how many hours per week do you spend on professional activities?

How high is your monthly income after taxes? Please also include transfer payments, funds from the federal educational grant, capital income etc.

On a monthly level, du you usually have some amount of money left to save or put aside such as for bigger investments, emergencies, or wealth formation?

How do you expect your financial situation to evolve in the next 12 months?

[CRT4[6]]

If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?

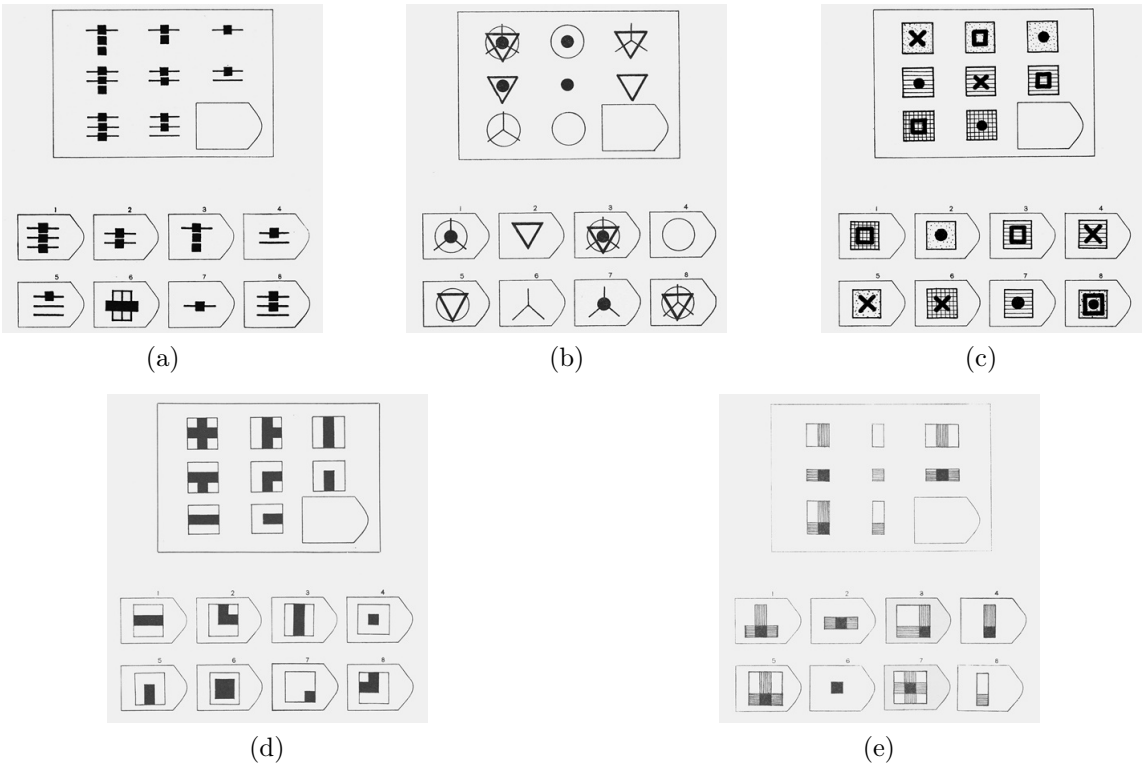
Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?

A man buys a pig for \$60, sells it for \$70, buys it back for \$80, and sells it finally for \$90. How much has he made?

Simon decided to invest \$8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money.

[6]From Toplak and Stanovich (2014).

Raven Matrices



(a) (b) (c)

(d) (e)

Figure 5: Raven Matrices

The experiment is over!

You can keep both your objects. You will also receive a show-up fee of 4 euros. Please wait shortly in you cabin until the experimenter calls you out. Thank you for your participation!

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman (2011). "Reference points and effort provision". In: *The American Economic Review*, pp. 470–492.

Ericson, Keith M. Marzilli and Andreas Fuster (2011). "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments". In: *The Quarterly Journal of Economics* 126.4, pp. 1879–1907.

Faul, Franz, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner (2007). "G*Power 3:A flexible statistical power analysis program for the social, behavioral, and biomedical sciences". In: *Behavior Research Methods* 39.2, pp. 175–191.

Gneezy, Uri, Lorenz Goette, Charles Sprenger, and Florian Zimmermann (Forthcoming). "The Limits of Expectations-Based Reference Dependence". In: *Journal of the European Economic Association*.

Goette, Lorenz, Thomas Graeber, Alexandre Kellogg, and Charles Sprenger (2018). "Heterogeneity of Loss Aversion and Expectations-Based Reference Points". Unpublished Working Paper.

Goette, Lorenz, Anette Harms, and Charles Sprenger (2016). "Randomizing Endowments: An Experimental Study of Rational Expectations and Reference-Dependent Preferences". Unpublished Working Paper.

Heffetz, Ori and John A. List (2014). "Is the Endowment Effect an Expectations Effect?" In: *Journal of the European Economic Association* 12.5, pp. 1396–1422. ISSN: 1542-4774.

Kőszegi, Botond and Matthew Rabin (2006). "A model of reference-dependent preferences". In: *The Quarterly Journal of Economics*, pp. 1133–1165.

Toplak, Maggie and Keith Stanovich (2014). "Assessing Rational Thinking Using an Expansion of the Cognitive Reflection Test". In: *Canadian Journal of Experimental Psychology* 68.4.