Title
Essays on Expectations-Based Reference Dependence

Author
Kellogg, Alexandre

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on Expectations-Based Reference Dependence

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Alexandre Kellogg

Committee in charge:

Professor Charles Sprenger, Co-Chair
Professor Isabel Trevino, Co-Chair
Professor Jim Andreoni
Professor Renee Bowen
Professor Sally Sadoff
Professor Joel Watson

2021

The dissertation of Alexandre Kellogg is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Thanks to my advisor, Charles Sprenger, for guiding me throughout my journey into the reference-dependent landscape. Thanks to my family for the support provided; this would not have been possible without you.

Chapter 1 is currently being prepared for submission for publication. Goette, Lorenz; Graeber, Thomas; Sprenger, Charles. The dissertation author was a principal researcher on this paper.

Chapter 3 is coauthored with Campos-Mercade, Pol; Goette, Lorenz; Sprenger, Charles. The dissertation author was a principal researcher on this paper. The work was supported by the National Science Foundation Doctoral Dissertation Award 1949517.

VITA

2015	B. A. in Economics and Mathematics *cum laude*, Pomona College
2016-2021	Graduate Teaching Assistant, University of California San Diego
2021	Ph. D. in Economics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Essays on Expectations-Based Reference Dependence

by

Alexandre Kellogg

Doctor of Philosophy in Economics

University of California San Diego, 2021

Professor Charles Sprenger, Co-Chair
Professor Isabel Trevino, Co-Chair

This dissertation explores the interplay between heterogeneity in gain-loss attitudes and a leading model of expectations-based reference dependence, referred to throughout as KR (Kőszegi and Rabin, 2006, 2007). Reference dependence posits that individuals consider outcomes relative to some reference point (e.g., the status quo or an expectation) rather than evaluating outcomes in isolation; these models have helped rationalize behavior inconsistent with the neoclassical model of expected utility.

Since the development of this KR expectations-based mechanism, a number of studies have sought to experimentally test the comparative static predictions in the exchange (Ericson

and Fuster, 2011; Goette, Harms, et al., 2016; Heffetz and List, 2014) and labor supply (Abeler et al., 2011; Gneezy et al., 2017) contexts. The mixed experimental results initially cast doubt on KR's ability to predict behavior. Importantly, these tests were all conducted under an implicit assumption of universal loss aversion – that individuals weight losses below the reference point more than gains above. Recent work, however, documents a substantial fraction of gain-lovers (Chapman, Snowberg, et al., 2018), meaning these experiments were incidentally testing a joint hypothesis of the KR model and loss aversion.

Throughout these chapters, experiments are specifically designed to overcome this potential confound by measuring gain-loss attitudes in a first stage and relating these measures to theoretical predictions in a second stage. The results provide evidence in support of the KR predictions after accounting for this underlying heterogeneity. Moreover, a Bayesian re-analysis of the exchange experiments suggests that a highly heterogeneous distribution of gain-loss attitudes best rationalizes the mixed results, with an estimated posterior indicating 35% to 55% of participants as gain-loving.

Chapter 1

Heterogeneity of Gain-Loss Attitudes and Expectations-Based Reference Points

1.1 Introduction

Models of reference-dependent preferences are regarded as a major advance in behavioral economics, rationalizing a range of observations at odds with the canonical model of expected utility over final wealth (C. Camerer et al., 1997; Kahneman, J. L. Knetsch, et al., 1990; Odean, 1998; Rabin, 2000). Critical to such applications is the formulation of the reference point around which gains and losses are encoded. A recent literature has examined characterizations of the reference point based on rational expectations of potential outcomes (Kőszegi and Rabin, 2006, 2007) (henceforth KR).[1] These expectations-based formulations have the promise to be readily and broadly applicable, closing the model with a foundation to which economic tools are already well adapted.

Despite the promise of the KR formulation of the reference point, tests of the theory

[1] Our analysis will focus on the formulations of KR. An earlier literature also provided formulations of reference dependence grounded in rational expectations, but without the equilibrium concepts we analyze (Bell, 1985; Loomes and Sugden, 1986).

have yielded mixed results (see, e.g., Abeler et al., 2011; Ericson and Fuster, 2011; Gneezy et al., 2017; Goette, Harms, et al., 2016; Heffetz and List, 2014; Smith, 2019). While early experimental applications in exchange behavior and effort provision showed treatment effects in line with KR comparative statics, subsequent replications and extensions have shown more limited or contradictory effects. A plausible interpretation from this literature is that the KR model of expectations-based reference points lacks a strong empirical foundation.

While explicitly posed as tests of the hypothesis that reference points are derived from expectations, prior experimental tests of KR have actually been tests of an inadvertent joint hypothesis, implicitly assuming that individuals are universally loss-averse — weighting losses more than commensurately-sized gains. Ignored to date is the alternative possibility that individuals, even a minority, could be 'gain-loving' — weighting gains more than commensurately-sized losses. Tests conducted thus far have investigated the joint hypothesis that reference points are based in expectations *and* that all subjects are loss-averse. The mixed empirical evidence noted above can be read as rejecting expectations-based reference points or, possibly, as rejecting as rejecting the notion that gain-loss attitudes are universal. We explore the latter interpretation based on theoretical results for the confounding effects of heterogeneous gain-loss attitudes. We provide the first evidence of expectations-based reference dependence after splitting the joint hypothesis into its component parts.

The behavioral economics literature since Kahneman and Tversky (1979) has generally advocated for loss aversion based on a psychological generality that "the aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount (p.279)." The psychology for gain-loving behavior is the opposite balance between disappointment and elation: the aggravation that one experiences in losing a sum of money being less than the pleasure associated with gaining the same amount. Such a psychology would lead subjects to accept, rather than reject, a small-stakes positive-expected-value lottery over a gain and a loss or exhibit an 'anti-endowment effect' in exchange settings, rather than the standard behavioral patterns.

While not widely appreciated, previous work has documented precisely the sort of heterogeneity in gain-loss attitudes ignored in prior tests of the KR model. Chapman, Snowberg, et al. (2018) note seven prior papers documenting the distribution of gain-loss attitudes with lottery choice, and one doing so with exchange behavior (a prior version of this paper).[2] These distributional assessments identify sizable minorities of gain-loving subjects. That is, sizable minorities of subjects do indeed accept small-stakes positive (and even some negative) expected value lotteries over a gain and a loss. Summarizing these prior data, Chapman, Snowberg, et al. (2018) report a weighted average of 22 percent gain-loving subjects. One may be tempted to eschew the minority of gain-loving subjects as reflecting decision errors or natural sampling variability. Such a view is challenged by prior work indicating that individual differences in measured gain-loss attitudes from lottery choices are predictive of anomalies in labor supply and exchange decisions (see, e.g., Abeler et al., 2011; Dean and Ortoleva, 2015; Fehr and Goette, 2007; Gachter et al., 2007). Such correlations across domains should not occur if the variation in measured gain-loss attitudes was mere sampling error.

The previously documented heterogeneity in gain-loss attitudes has a critical influence on what can be inferred from prior empirical tests of the KR model. We show that KR comparative statics used as the basis of prior experiments change sign when individuals are gain-loving rather than loss-averse. If loss aversion is not a universal characteristic, prior empirical tests have unintentionally aggregated these different signed effects. As we show in detail in section 1.2.2, the treatment effects are non-linear in gain-loss attitudes. Hence, the average treatment effect need not coincide with the treatment effect of the average preference. Indeed, null and mis-signed average treatment effects (relative to the loss averse prediction) can easily occur with an average preference of loss aversion.

Heterogeneity and aggregation issues also affect the power of tests for the average treat-

[2]The manuscripts are Burks et al. (2009), Erev et al. (2008), Harinck et al. (2007), Nicolau (2012), Sokol-Hessner et al. (2009), and Sprenger (2015); and Chapman, Dean, et al. (2017).

ment effect. Our distributional findings indicate that well powered experiments for detecting average KR treatment effects require sample sizes around an order of magnitude larger than current practice. If one recognizes the prior findings of heterogeneity in gain-loss attitudes and the challenge they raise for the empirical study of the KR model, one is thus effectively forced to examine the model's predicted differential treatment effects over gain-loss attitudes.

We implement an experimental study of gain-loss attitudes and exchange behavior in an initial sample of 607 subjects and a pre-registered replication sample with a further 417 subjects. Our design has two stages. Stage 1 is dedicated to measuring gain-loss attitudes, and Stage 2 focuses on testing the heterogeneous treatment effects over gain-loss attitudes predicted by the KR model.

In Stage 1, subjects are randomly endowed with one of two objects and are asked for their hypothetical choice between the two objects, how much they 'want' each object on a nine-point scale, and how much they 'like' each object on a nine-point scale. The Stage 1 preference statements allow us to form a taxonomy of gain-loss types, constructed from a structural model of the preference statements for the endowed and alternative object.[3]

In Stage 2, subjects are endowed with one of two new objects, both completely different from those used in Stage 1. The objects used in the study are balanced across Stage 1 and Stage 2, with roughly equal numbers of subjects exposed to each pair in each stage. The design's balanced use of different objects in Stage 2 relative to those used for the elicitation of gain-loss attitudes in Stage 1 is critical. Any behavioral connection across stages *cannot* be attributable to the valuation of the objects in question. Once endowed with their new object, subjects are randomly assigned to one of two conditions. One group of subjects is assigned to Condition B,

[3]An alternative design would attempt to measure gain-loss attitudes either through small-stakes risk aversion or some other choice. Such tests would face a number of challenges, requiring both additional assumptions (e.g., about the correlation between intrinsic utilities and gain-loss parameters across contexts) and additional experimental choices. Recognizing both the polluting potential of such choices and the challenge of modeling the full body of experimental behavior through the lens of the KR model (Sprenger, 2015), we believe there is substantial value in our method. In section 1.4.2, we show predictive power for our measure of gain-loss attitudes and exchange behavior in a standard exchange paradigm. Of course, failure to correctly categorize gain-loss types should lead to a lack of predictive validity in Stage 2 of the experiment, working against these and other identified results.

a baseline endowment effect condition, where they decide whether they would like to exchange their object for the alternative. Another group of subjects is assigned to Condition F, where they decide whether to exchange their object under a probabilistic forced exchange mechanism. With probability 0.5, regardless of their decision, exchange will be forced. Similar probabilistic mechanisms are the standard experimental approach for investigating the KR model's predictions in exchange behavior (see, e.g. Ericson and Fuster, 2011; Goette, Harms, et al., 2016; Heffetz and List, 2014).

Under the KR model, loss-averse subjects should be more willing to exchange in Condition F than in Condition B. Intuitively, exchanging in Condition F eliminates the potential loss associated with attempting to retain the endowed object but being forced to exchange. Thus, a loss-averse subject who is unwilling to exchange in Condition B may be willing to exchange in Condition F. In contrast, gain-loving subjects should exhibit the opposite pattern, growing less willing to exchange in Condition F relative to Condition B. Not exchanging in Condition F creates potential gains that outweigh the potential losses associated with probabilistic forced exchange.

We document two key findings. First, we reproduce findings of heterogeneity in gain-loss attitudes. On average, subjects state a preference for their randomly endowed object in Stage 1, indicating an endowment effect. However, preference statements also exhibit marked heterogeneity.[4] Roughly 25 percent of subjects' Stage 1 statements indicate an anti-endowment effect. Our primary structural model interprets the distribution of choices as driven by uniform object-specific valuations and heterogeneous gain-loss attitudes. This model identifies a distribution of the gain-loss parameter, λ, with loss aversion ($\lambda > 1$) on average but substantial heterogeneity. Indeed, the fitted distribution estimates 38 percent (95% C.I. $= [13, 49]$ percent) of individuals as gain-loving ($\lambda < 1$). This finding reinforces prior work on the heterogeneity of gain-loss at-

[4] Fifty-seven percent of subjects state they would choose their randomly endowed object, and two-thirds provide weakly higher wanting and liking ratings for their endowed object. These preference statements are highly correlated with each other: all pairwise correlations exceed 0.7.

titudes. Previous estimates of the fraction of gain-loving individuals comfortably fit within our confidence intervals. This indicates that homogeneous loss aversion in our sample would be the wrong assumption with which to proceed.

Second, Stage 2 behavior differs substantially by the gain-loss attitudes measured in Stage 1 with different objects. Loss-averse subjects are markedly less willing to exchange than gain-loving subjects in Condition B. This intuitive relationship between λ and behavior in a standard exchange paradigm demonstrates the predictive validity of the gain-loss attitudes measured in Stage 1. In Condition F, the relationship between gain loss attitudes and exchange behavior reverses, leading to precisely the heterogeneous treatment effects proposed by the KR model. Loss-averse subjects are more willing to exchange in Condition F than in Condition B, while gain-loving subjects are less so. Both the sign and magnitude of our heterogeneous treatment effects are closely in line with the predictions of the KR model. These represent the first experimental findings on the KR model accounting for heterogeneous gain-loss attitudes.

Our results also hold under alternate methodologies for identifying gain-loss attitudes. In particular, we provide a reduced-form exercise which infers gain-loss attitudes based upon residualized Stage 1 behavior. There again, individual differences in gain-loss attitudes closely relate to differences in Stage 2 treatment effects. Additionally, our findings replicate in both an initial study and an exact replication. Appendix A.1 details our pre-registered analysis, which featured centrally in prior drafts of this manuscript,[5] and Table 1.5 in the main text shows the reproducibility.

We interpret our findings as documenting the importance of heterogeneity in gain-loss attitudes for expectations-based reference dependence. A number of alternative interpretations may be proposed for our effects. First, one may misinterpret our findings as being indicative of heterogenous valuations for the objects in the experiment. As noted above, our design uses

[5]We are indebted to an anonymous referee for suggesting the current path of structural analysis. This analysis differs from our prior presentation of the results, as we have adopted a mixed logit methodology for identifying gain-loss attitudes rather than our prior standard logit methods.

different sets of objects for the measurement of gain-loss attitudes in Stage 1 and for our core test of KR comparative statics in Stage 2. Given random assignment, heterogeneous valuations for the objects in Stage 1 would lead to no average preference for the endowed object; and, more importantly, there would be no correlation between the preferences elicited in Stage 1 and behavior with different objects in Stage 2. Because this is true for both Condition B and Condition F, there should also be no heterogeneity in treatment effects. We reinforce this point with a companion structural exercise to our main analysis, allowing heterogeneity in object valuations, that shows, unsurprisingly, zero predictive power. Second, one may attribute preferring the non-endowed object in Stage 1 to a 'grass-is-greener' effect — subjects wanting that with which they are not endowed — rather than gain-loss attitudes. In Stage 2, both grass-is-greener and grass-is-not-greener subjects should employ their prior logic in Condition B and in Condition F, delivering, again, zero heterogeneity in treatment effects. Third, perhaps subjects are confused by the elements of our two-stage design, and so provide noisy responses that are challenging to interpret in any regard. Random noise should do only one thing in our design: work against any obtained findings of predictive power across stages. For noise in response to accurately characterize our data it must have several systematic features: it must systematically lead to an average preference for the randomly endowed object in Stage 1; it must generate a systematic positive correlation between Stage 1 behavior and behavior in Stage 2 Condition B; and it must generate a systematic negative correlation between Stage 1 behavior and the treatment difference between Condition F and Condition B in Stage 2. Random noise, such as what one might normally attribute to confusion, should not exhibit such systematic features. Fourth, perhaps one should interpret findings of heterogeneous gain-loss attitudes as reflective of sampling variation around a homogeneous degree of loss aversion, spuriously revealing some subjects as gain-loving. Such a possibility should be reflected in both our confidence interval for the probability of $\lambda < 1$ allowing for extremely low values, and in a lack of predictive power in Stage 2. Both are rejected by the data.

Our identified heterogeneity of gain-loss attitudes, reflected in heterogeneous treatment effects over gain-loss types, carries important implications for experimental and theoretical work on expectation-based reference dependence. First, given our theoretical development and results, heterogeneity in gain-loss attitudes appears to be an issue of first-order importance. Prior work showing inconsistent average treatment effects in experiments on the KR model should not be interpreted as a rejection of the theory. Indeed, we show that our own average treatment effect is a null effect, quite close to theoretical average treatment effect under our distribution of gain-loss attitudes, 5.9 percentage points. Mixed evidence on the KR model is likely not driven by a failure of the expectations-based formulation of reference points, but rather by a failure of the second component of the joint hypothesis inherent to this prior work: that gain-loss attitudes are universal. In a simple and reproducible way, we show that the predictions of KR are reliably recovered once one accounts for heterogeneity in gain-loss attitudes.

Second, we show that even with over 1000 collective observations, we are dramatically underpowered to identify the theoretical average treatment effect in our experiment. The theoretical average treatment effect of 5.9 percentage points would require a sample size of around 2250 observations to estimate with 80 percent power. Prior experimental tests focused on average treatment effects may be similarly underpowered given the findings of heterogeneous gain-loss attitudes both here and in the prior literature.

Third, while the data are markedly supportive of the KR formulation, the average *level* of exchange behavior in our design falls below the *level* predictions of KR under strict formulations of refined equilibrium behavior (and nothing else influencing choice). We view it as unlikely that *only* the KR expectations-based mechanism is driving behavior. This highlights the importance of investigating *treatment effects* common to all KR equilibrium formulations rather than *levels*, as also pointed out in Ericson and Fuster (2011) (see Section 1.2.1 and Appendix A.2 for detail).

Lastly, we add a key observation on the heterogeneity of gain-loss attitudes to a growing literature on the topic. Chapman, Snowberg, et al. (2018) indicate eight prior studies with a documented distribution of gain-loss attitudes, only one of which is measured outside of lottery choice (a prior version of this paper). Ours are the first findings to document the distribution of gain-loss attitudes in exchange settings, and the predictive validity of resulting individual measures. In our exchange setting, we document an average attitude of loss aversion, but a sizable proportion of the distribution, 38 percent, exhibits gain lovingness. This proportion of gain-loving subjects somewhat exceeds estimates from risk experiments in the lab, but falls below the field estimates of Chapman, Snowberg, et al. (2018). Future work providing further documentation and evaluation of the heterogeneity in gain-loss attitudes across domains is equipped with an initial observation from exchange behavior.

The paper proceeds as follows. In Section 1.2, we set the theoretical background and derive behavioral predictions. Section 1.3 and 1.4 present the experimental design and results, respectively. Section 1.5 concludes.

1.2 Theoretical Considerations and Design Guidance

We examine the predictions of the KR model in simple exchange settings with two objects, recognizing heterogeneity of gain-loss attitudes. The theoretical development hues closely to our experimental design, providing motivation for our analysis. We contrast two conditions: a standard exchange paradigm, termed Condition B, where subjects are endowed with an object and decide whether to exchange or not; and a probabilistic forced exchange paradigm, termed Condition F, identical to Condition B except that with probability 0.5, regardless of choice, exchange will be forced. We show that loss-averse subjects should grow more willing to exchange in Condition F relative to Condition B. In contrast, gain-loving subjects should grow less willing to exchange in Condition F relative to Condition B.

There is a central intuition for the heterogeneous response to probabilistic forced exchange. When attempting not to exchange their endowment in Condition F, an individual faces the potential of having this object taken from them and exchanged regardless of their desire. A loss-averse individual disproportionately dislikes the sensation of potential loss and so may choose to exchange to avoid the possible loss. In contrast, a gain-loving individual disproportionately likes the sensation of potential gain and so may choose not to exchange to maintain the possible gain.

Consider a two-dimensional utility function over two objects of interest, object X and object Y. Let $\mathbf{c} = (m_X, m_Y)$ and $\mathbf{r} = (r_X, r_Y)$ represent vectors of intrinsic utility and reference utility, respectively. The KR model specifies a utility function with two components, intrinsic utility, $m(\mathbf{c}) \equiv m_X + m_Y$, and gain-loss utility, $n(\mathbf{c}|\mathbf{r}) \equiv n_X(m_X|r_X) + n_Y(m_Y|r_Y) \equiv \mu(m_X - r_X) + \mu(m_Y - r_Y)$, with separability across consumption dimensions. Let $m_X \in \{0, X\}$ and $m_Y \in \{0, Y\}$ stand for both the outcome and the corresponding intrinsic utility of owning zero or one unit of object X, and zero or one unit of object Y, respectively. Overall utility is described by

$$\begin{aligned} u(\mathbf{c}|\mathbf{r}) = u(m_X, m_Y|r_X, r_Y) &= m_X + n_X(m_X|r_X) + m_Y + n_Y(m_Y|r_Y) \\ &= m_X + \mu(m_X - r_X) + m_Y + \mu(m_Y - r_Y), \end{aligned}$$

where

$$\mu(z) = \begin{cases} \eta z & \text{if } z \geq 0 \\ \eta\lambda z & \text{if } z < 0. \end{cases}$$

In this piece-wise linear gain-loss function, the parameter η captures the magnitude of changes relative to the reference point, and λ captures gain-loss attitudes. If $\lambda > 1$, the individual is loss-averse, experiencing losses more than commensurately-sized gains. If $\lambda < 1$, the individual is gain-loving, experiencing gains more than commensurately-sized losses.

1.2.1 Determination of the Reference Point in Exchange Behavior

In the KR model, unless exogenously determined, the vector **r** is established as part of a consistent forward-looking plan for behavior. The KR model posits a reference-dependent expected utility function $U(F|G)$, taking as input a distribution F over consumption outcomes, **c**, which are valued relative to a distribution G of reference points, **r**. That is:

$$U(F|G) = \int \int u(\mathbf{c}|\mathbf{r}) dF(\mathbf{c}) dG(\mathbf{r}).$$

A *Personal Equilibrium* is a situation where, given that the decision-maker expects as a referent some distribution, F, they indeed prefer F as a consumption distribution over all alternative consumption distributions, F'. Ex-ante optimal behavior has to accord with expectations of that behavior. Formally, given a choice set, $\mathcal{D}$, of lotteries, F, over consumption outcomes $\mathbf{c} = (m_X, m_Y)$, KR's *Personal Equilibrium* states the following:

Personal Equilibrium (PE): A choice $F \in \mathcal{D}$, is a personal equilibrium if

$$U(F|F) \geq U(F'|F) \; \forall \; F' \in \mathcal{D}.$$

Regardless of endowment, if object X is to be chosen in a PE, then $\mathbf{r} = (X, 0)$, and if object Y is to be chosen in a PE then $\mathbf{r} = (0, Y)$.

Given the potential for the multiplicity of PE selections, the KR model is constructed with a notion of equilibrium refinement, *Preferred Personal Equilibrium* (PPE), and an alternate non-PE criterion, *Choice-Acclimating Personal Equilibrium* (CPE). In both of these constructs, ex-ante utility is used as a basis for selection and, hence, for making more narrow predictions. For ease of explication, we focus our analysis on the CPE criterion. In Appendix A.2 we provide theoretical analyses under the PE and PPE approaches. Importantly, all three formulations share common comparative statics, and therefore make qualitatively similar predictions, for our KR test.

Given a choice set, $\mathcal{D}$, of lotteries, F, over consumption outcomes $\mathbf{c} = (m_X, m_Y)$, *Choice-Acclimating Personal Equilibrium* states the following:

Choice-Acclimating Personal Equilibrium (CPE): A choice $F \in \mathcal{D}$, is a choice-acclimating personal equilibrium if

$$U(F|F) \geq U(F'|F') \; \forall \, F' \in \mathcal{D}.$$

Under CPE, an individual selects between options like $[\mathbf{c},\mathbf{r}] = [(X,0),(X,0)]$ and $[\mathbf{c},\mathbf{r}] = [(0,Y),(0,Y)]$.[6]

Manipulating r: Probabilistic Forced Exchange

The CPE concept noted above requires consistency between the distributions of $\mathbf{c}$ and $\mathbf{r}$. We consider a baseline simple exchange condition, Condition B, for an individual endowed with object X. We focus on the choice set consisting of pure strategy choices $\mathcal{D} = \{(X,0),(0,Y)\}$, with the first element reflecting choosing not to exchange and the second choosing to exchange.[7]

In this setting, there are two potential CPE selections, $[\mathbf{c},\mathbf{r}] = [(X,0),(X,0)]$ and $[\mathbf{c},\mathbf{r}] = [(0,Y),(0,Y)]$. The individual can support not exchanging, $[\mathbf{c},\mathbf{r}] = [(X,0),(X,0)]$, in a CPE if

$$U(X,0|X,0) \geq U(0,Y|0,Y),$$

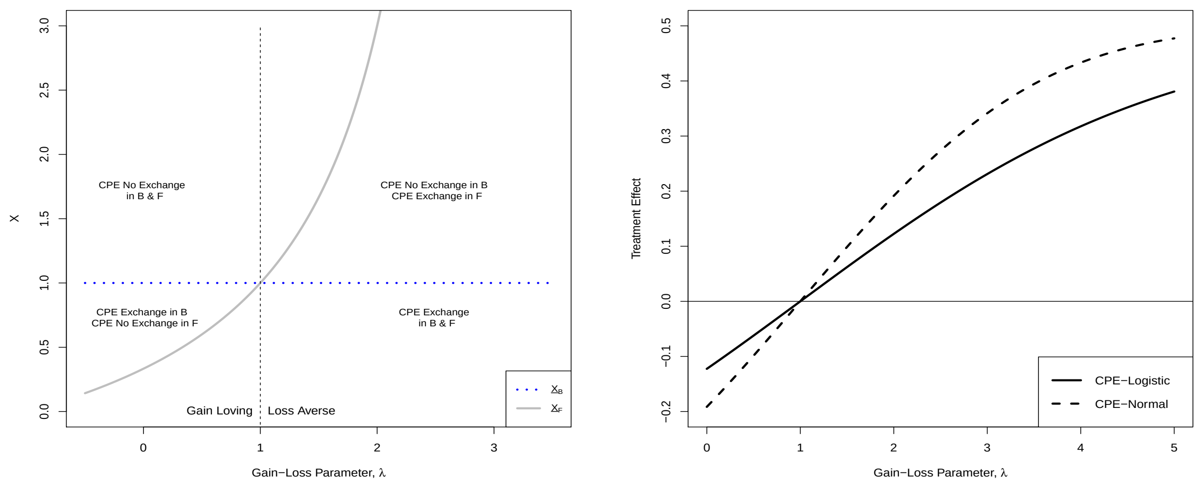
which, under our functional form assumptions, becomes

$$X \geq Y. \tag{1.1}$$

Figure 1.1, Panel A graphs the Condition B CPE cutoff, $\underline{X}_B = Y$, the smallest value of X at

[6]Note that a selection need not be PE in order to be CPE. The alternate concept, PPE, requires F and F' to be PE, rather than simply elements of $\mathcal{D}$.

[7]In Appendix A.2, we conduct the analysis with $\mathcal{D}$ including all mixtures of exchanging and not exchanging and reach quite similar results.



(a) Theoretical CPE Thresholds

(b) Predicted Treatment Effects

Figure 1.1: **Gain-Loss Attitudes, Theoretical Thresholds, and Treatment Effects**

Notes: Panel A: CPE cutoff values for agent endowed with X, $Y = 1$ and $\eta = 1$. For $X \geq \underline{X}_B = Y$, individuals can support not exchanging as a CPE in a baseline exchange environment (Condition B). For $X \geq \underline{X}_F = \frac{1+0.5\eta(\lambda-1)}{1+0.5\eta(1-\lambda)}Y$, individuals can support not exchanging as a CPE in a forced exchange environment (Condition F). Panel B: Simulated treatment effects for the probability of exchange plotted by λ with $Y/X = 1$, $\eta = 1$ under logistic or normal probability distributions.

which the individual can support not exchanging, which is constant for all values of the gain-loss parameter, λ.

The value $\underline{X}_B = Y$ implies that choice in Condition B is governed only by intrinsic utility. This represents the inability of KR-CPE to rationalize the standard endowment effect. This prediction is not shared by the PE formulation, wherein the value of gain-loss attitudes tunes the set of permissible PE choices and can lead to an endowment effect (see Appendix A.2). Nonetheless, the critical comparative static shared by both formulations is delivered by comparing exchange behavior in this baseline Condition B with probabilistic forced exchange.

Now, consider an environment of probabilistic forced exchange, Condition F. With probability 0.5, the agent, assumed endowed with X, will be forced to exchange X for Y regardless of their choice. If the individual wishes to retain their object, they are subject to a stochastic reference point, as with probability 0.5 their object will be exchanged. Now, the potential selections for someone endowed with X are $\mathcal{D} = \{0.5(X,0) + 0.5(0,Y), (0,Y)\}$, with the first element reflecting attempting not to exchange and the second reflecting exchange, as before. They can support attempting not to exchange as a CPE if

$$U(0.5(X,0)+0.5(0,Y)|0.5(X,0)+0.5(0,Y)) \geq U(0,Y|0,Y),$$

which, under our functional form assumptions, becomes

$$\begin{aligned} 0.5X + 0.5Y + 0.25\eta(1-\lambda)(X+Y) &\geq Y \\ X &\geq \frac{1+0.5\eta(\lambda-1)}{1+0.5\eta(1-\lambda)}Y. \end{aligned}$$

The manipulation of probabilistic forced exchange changes the CPE threshold from $\underline{X}_B = Y$ to $\underline{X}_F = \frac{1+0.5\eta(\lambda-1)}{1+0.5\eta(1-\lambda)}Y$. Figure 1.1, Panel A illustrates the changing CPE cutoff values associated with not exchanging. In Condition F, the individual can support attempting to retain X in CPE on the basis of both intrinsic utility and gain-loss attitudes.

1.2.2 Heterogeneity in Gain-Loss Attitudes and Aggregation Issues

The gain-loss parameter, λ, tunes precisely how behavior should change between Conditions B and F. Figure 1.1, Panel A is partitioned into four regions. Two critical regions of changing CPE choice are identified. For $X > Y$ and $\lambda > 1$, it is CPE to not exchange in Condition B, and CPE to exchange in Condition F. This region has been the basis of prior experimental tests under the assumption of universal loss aversion; such individuals become more willing to exchange when probabilistically forced. Ignored to date is the region where $X < Y$ and $\lambda < 1$. In this region, it is CPE to exchange in Condition B, and CPE to not exchange in Condition F. In contrast to the loss-averse prediction, such gain-loving individuals become less willing to exchange when probabilistically forced. The KR comparative static for the difference between

Condition B and Condition F changes sign at $\lambda = 1$. Prior aggregate empirical tests assuming homogeneous gain-loss attitudes have thus aggregated these different signed effects for loss-averse and gain-loving subjects.

The CPE threshold determining behavior in Condition F, $\underline{X}_F = \frac{1+0.5\eta(\lambda-1)}{1+0.5\eta(1-\lambda)}Y$, is non-linear in λ. This non-linearity means that in a sample with heterogeneous gain-loss attitudes, the average threshold value of $\underline{X}_F$ will not coincide with the threshold value of the average preference. Correspondingly, treatment effects based the changing thresholds between conditions will inherit the same non-linearity. Hence, the average treatment effect will also not coincide with the treatment effect of the average preference.

In order to make numerical predictions for behavior and the confounding effects of heterogeneity, we map from the CPE thresholds to the probability of making a specific selection. We simulate behavior assuming that X and Y have equal intrinsic utility, $Y/X = 1$, $\eta = 1$, and the CPE utilities are followed probabilistically subject to a specific logit choice model.[8] That is, an individual chooses to exchange in Condition B with probability

$$\begin{aligned} Prob(Exchange)_B &= Prob(Y > X) = Prob(Y/X - 1 > 0) \\ &= logistic(0) = 0.5. \end{aligned}$$

Similarly, the individual chooses to exchange in Condition F with probability

$$\begin{aligned} Prob(Exchange)_F &= Prob(Y > 0.5X + 0.5Y + 0.25\eta(1-\lambda)(X+Y)) \\ &= Prob(0.5(Y/X - 1) + 0.25\eta(\lambda - 1)(1 + Y/X) > 0) \\ &= logistic(0.5(\lambda - 1)). \end{aligned}$$

[8] In our actual empirical results, we estimate the value of Y/X rather than fix it by assumption. We maintain $\eta = 1$ throughout.

And the treatment effect is simulated as

$$\begin{aligned} TE &= Prob(Exchange)_F - Prob(Exchange)_B \\ &= logistic(0.5(\lambda - 1)) - logistic(0) \end{aligned}$$

Figure 1.1, Panel B graphs this treatment effect against the value of λ under the assumptions $\eta = 1$ and $Y/X = 1$. Appendix A.2 provides the same analysis under PE and PPE, reaching similar conclusions for the effects of heterogeneity. The theoretical simulated treatment effect is negative for $\lambda < 1$, positive for $\lambda > 1$, and is generally concave in λ. Figure 1.1, Panel B also provides a theoretical benchmark under a normal probability distribution rather than the logistic, highlighting the robustness of the non-linearity prediction.

The apparent concavity of simulated treatment effects in λ implies a substantial challenge for the aggregation of treatment effects. Not only do treatment effects change sign at $\lambda = 1$, but gain-loving subjects can have an outsized impact on identified average effects. Given the non-linearity of treatment effects over gain-loss attitudes, the average treatment effect doesn't coincide with the treatment effect of the average preference. Furthermore loss aversion on average does not guarantee positive average treatment effects. This makes heterogeneity a confound of first-order importance for experiments in this vein. Null and mis-signed (relative to the average preference) average treatment effects can be consistent with average loss aversion. Any test of KR must account for heterogeneity in gain-loss attitudes to credibly test the underlying expectation-based reference-dependent mechanism. Motivated by this point, our study combines the experimental manipulation of probabilistic forced exchange with a prior measurement of gain-loss attitudes.

1.3 Experimental Design and Procedures

Our design is comprised of two stages. In Stage 1, a taxonomy of gain-loss types is created. In Stage 2, subjects are assigned to either a standard exchange study or one with probabilistic forced exchange. Stage 1 measures of gain-loss attitudes can then be connected to Stage 2 behavior. Figure 1.2 illustrates the experimental order of events.

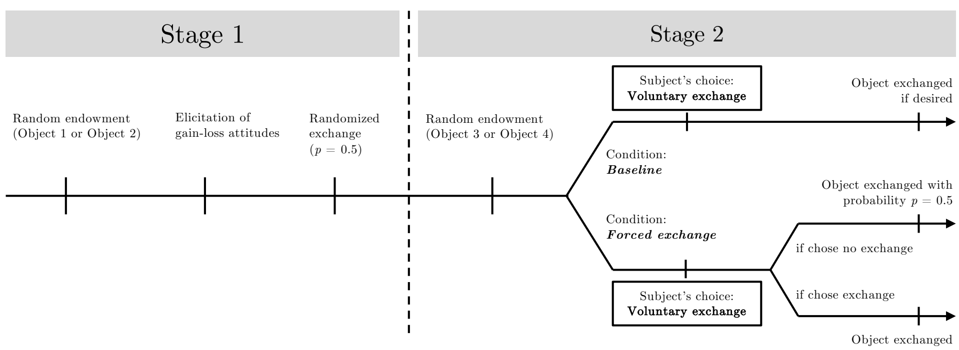


Figure 1.2: **Timeline of Laboratory Experiment**
Notes: The figure displays the course of events in both treatment conditions, Condition B(aseline) and Condition F(orced exchange).

1.3.1 Stage 1: Measuring Gain-Loss Attitudes

Procedures. The experimenter welcomed the participants in a small presentation room and informed them that the study would consist of two stages. At each seat was a card with a number (placed face down). Then, without further explanation, the experimenter projected on the wall two equally-sized pictures of the respective Stage 1 objects for that session, along with the description and two short bullet points on the characteristics of each product. The exact information presented to subjects is reproduced in Appendix A.5 in German and translated to English.

After allowing sufficient time (three minutes) to study the projected information, the experimenter asked subjects to turn the card in front of them over and move to the cubicle with the corresponding number in the adjacent computer laboratory. In their private cubicle, which was separated and not visible from the outside, subjects would find one of the two presented objects. Computer instructions then informed the subject that the object in front of them was in their possession, and that they were free to inspect it more closely.

After three minutes allotted for inspection of the object, we asked subjects three questions. First, for each object subjects were asked "How much do you like this product?" with response scales ranging from 0="Not at all" to 8="Very much". Second, for each object they were asked "How much would you want to have this product?" on the same response scales. Third, they were asked "If you had to choose one of the objects, which one would you prefer to keep?", and were asked to provide a hypothetical choice between the two objects. These three preference statements are the raw data upon which our structural estimates of gain-loss attitudes are constructed.

Our Stage 1 design exogenously endows subjects with objects and elicits preference statements under this fixed endowment prior to any discussion of exchange. In the case of an exogenous endowment that cannot be expected to change, the KR model coincides exactly with the standard model of reference-dependent preferences with a fixed reference point. This allows us to elicit gain-loss attitudes under our exogenous endowment. See Kőszegi and Rabin (2006) for additional discussion of this point in the particular context of the endowment effect. Had we conducted an alternate design without such exogenous endowments or with salient discussion of exchange, the reference point would plausibly not be fixed, challenging our assumptions for measurement of gain-loss attitudes (see section 1.4.1 for further discussion).

Stage 1 of our experiment also featured one additional element of random variation: an experience with probabilistic exchange. After subjects provided their preference statements, the computer instructions announced that the experimenter would randomly draw a number between 1 and 20 using a rotating lottery drum placed on a table in the middle of the room. Half of the subjects were informed that the object in front of them would be replaced by the alternative object if a number between 1 and 10 was drawn. Instructions for the other half read that this replacement would only take place if a number between 11 and 20 was drawn.[9] The experimenter drew the number in a way that both the lotto device containing the 20 balls and the drawn number was visible from every cubicle. Immediately following the draw, and without further comment, the experimenter replaced objects as dictated by the drawn number. As noted above, it is critical that introduction of random replacement procedure was done *after* all preference statements were elicited under the exogenously endowed object.

This random experience serves two purposes in our design. First, regardless of Stage 2 treatment assignment, individuals will have had some prior experience with probabilistic exchange (albeit without choice). Second, it removes a potential challenge to our interpretation associated with complementarities between objects across rounds. If there existed some unmodeled, unintentional complementarity between the objects endowed in Stage 1 and Stage 2, a subject might state a preference for or against both of their endowed objects in order to consume both endowed objects or both alternatives together. Random replacement within Stage 1 breaks these potential complementarities as a driver of Stage 2 choice, and we can explore the relationship between Stage 1 experience and Stage 2 behavior.[10] After completing Stage 1, the instructions asked subjects to return to the main lecture room for Stage 2.

[9]This *loss condition* was counterbalanced within each subsample endowed with the same object, such that irrespective of the draw, exchange would take place for exactly half of the subjects initially endowed with either object.

[10]Immediately before and immediately after the random replacement was conducted, we elicited subjects' mood using standard psychological scales (Bradley and Lang, 1994). Subjects answered the question "Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?" by positioning a slider on an 11-point response scale. The lower end (0) was labeled using the words "Unhappy, Angry, Unsatisfied, Sad, Desperate" and the upper end (10) was labeled "Happy, Thrilled, Satisfied, Content, Hopeful'". The changes in these values were used as an initial validation of gain-loss types in prior versions of this manuscript. For space considerations we do not conduct this intermediate analysis here, but the results can be found in https://papers.ssrn.com/sol3/papers.cfm?abstractid=3170670 and https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3589906.

1.3.2 Stage 2: Baseline Exchange and Probabilistic Forced Exchange

Procedures. The basic procedures in Stage 2 were deliberately kept exactly identical to those in the Stage 1. Upon their return to the lecture room, the experimenter projected another page onto the wall, this time presenting the objects for Stage 2 of that session. In the meantime, a second experimenter allocated objects to the cubicles in the computer laboratory next door in a pre-specified order. Subjects were ushered back to their cubicle where they found their second endowed object and were allowed sufficient time for inspection. In Stage 2, subjects were randomized into one of two conditions: a baseline exchange condition, Condition B, and a treatment condition with probabilistic forced exchange, Condition F. The randomization was conducted at the session level.[11] Across our two studies, 59 percent (603 of 1024 subjects) were randomly assigned to Condition F sessions.[12]

Condition B: Baseline Exchange. In Condition B, subjects had an opportunity to voluntarily exchange their endowed object for the alternative. Their decision was final — whatever they chose they would receive. The baseline condition is a standard exchange setting common to endowment effect experiments.

Condition F: Probabilistic Forced Exchange. Condition F implemented an exchange study with probabilistic forced exchange. The instructions specified that regardless of their choice, exchange would take place with probability 0.5 based on a draw from the lotto drum, as in Stage 1. This means that for a subject who decided to exchange, the treatment had no effect. However, for a subject who attempted to keep their object, exchange would be forced probabilistically with a 50 percent chance.

Several noted issues with experimental investigations of market exchange motivated our

[11] We present our analysis with robust standard errors in the main text and Appendix Tables A.4 through A.6 reproduce our results with standard errors clustered at the session level. Statistical significance is enhanced with clustering, and so we opt to provide the more conservative values in the main text.

[12] In our initial study 62 percent (374 of 607) were assigned to Condition F under an assignment probability of 60 percent, and in our replication study 55 percent (229 of 417) were assigned to Condition F under an assignment probability of 50 percent.

purposefully simple design (Plott and Zeiler, 2005, 2007). First, subjects take a simple binary choice, alleviating potential concerns related to the use of 'multiple price lists' in exchange experiments. Specifically, we do not need to elicit a willingness to pay or willingness to accept in monetary terms, but simply ask whether the subject is willing to trade the endowed object for the alternative. As such, mistaken perceptions of market power do not play a role, nor do income effects. Second, unlike previous market exchange experiments, we create a private environment that limits confounds from social interaction. In particular, subjects make their decisions anonymously in a private cubicle; they find their endowment placed in front of them when entering the cubicle instead of receiving it personally through the hands of the experimenter, which has been criticized for triggering the misperception of the endowment as a gift (see, e.g., Plott and Zeiler, 2005, 2007); and subjects do not interact with other subjects at any stage during the experiment.

1.3.3 Sample Details

An initial sample of 607 students and a replication sample of a further 417 students from the University of Bonn participated in the experiment which was conducted using the software z-Tree (Fischbacher, 2007) in June and July 2015 and July 2018 at the BonnEconLab.[13] We conducted 53 sessions with 16 to 20 participants each. Table 1.1 provides an overview of the subject pool by treatment conditions.

The objects used for the exchange experiment included a USB stick, a set of three erasable pens, a picnic mat, and a thermos.[14] We selected these four objects on the basis of

[13]Several minor differences between the original sessions and those in the replication deserve note. We opted to split the treatment assignments between Condition B and Condition F at 50 percent-50 percent rather than the original 40 percent-60 percent to maximize power. Since storage technology rapidly advanced, the 8GB USB stick had to be replaced by a 16GB USB stick, as that was the new minimum. In addition, we were unable to repurchase the identical pattern for the picnic mat, so we opted for a visually similar one. Further, while only one experimenter ran the sessions for the original study, a total of 4 experimenters ran sessions during the course of the replication. In the Appendix, we repeat the analysis with experimenter fixed effects and find quantitatively similar results. Lastly, there was a small error in the implementation of sessions run by one specific experimenter who reversed the coded, randomly selected, endowments. Although this has no effect on the experiment, it did require us to recode the endowments for these sessions. The results excluding this experimenter's sessions also reproduce the findings here.

[14]Pictures and information presented to subjects are reproduced in Appendix A.5.

Table 1.1: **Summary Statistics and Treatment Assignment**

Stage 1

	Pair 1		Pair 2	
	USB stick	Pen set	Picnic mat	Thermos
A) Initial Endowment	274	264	242	244
– in % of subject pool	26.76%	25.78%	23.63%	23.83%

Stage 2

	Pair 1		Pair 2	
	USB stick	Pen set	Picnic mat	Thermos
B) Initial Endowment	242	244	274	264
– in % of subject pool	23.63%	23.83%	26.76%	25.78%
C) Condition B	113	117	97	94
– in % of B)	46.70%	47.95%	35.40%	35.60%
D) Condition F	129	127	177	170
– in % of B)	53.30%	52.05%	64.60%	64.40%
Total number of observations	1024			

Notes: Stage 2 condition (Condition B or Condition F) is randomized within each session. The use of each pair as the Stage 1 pair was counterbalanced at the session level.

a pre-experimental survey evaluation of 12 candidate objects. We put particular emphasis on ruling out complementarities between items across rounds. The former two (USB stick and pens) and the latter two objects (picnic mat and thermos) each constituted a pair. Every subject faced exactly one stage with each pair of objects. The use of each pair as the Stage 1 pair was counterbalanced at the session level, with the respective other pair used in Stage 2. Within each session, the endowments of one of the two objects within the pair was counterbalanced in both stages.[15]

[15]That is, if for a given session the USB stick and pens pair constituted the Stage 1 pair, the picnic mat and thermos pair would be the Stage 2 pair. Half of the subjects were initially endowed with the USB stick in Stage 1. Among this half of the session participants, again half would initially receive the picnic mat and the other half the thermos at the beginning of Stage 2.

1.4 Experimental results

We present the results in three subsections. First, we examine the Stage 1 preference statements leading to our taxonomy of gain-loss attitudes. Second, we examine behavior in Stage 2, linking heterogeneity in gain-loss attitudes to the behavioral response to probabilistic forced exchange. Third, we provide robustness tests and separate analyses for our initial and replication studies.

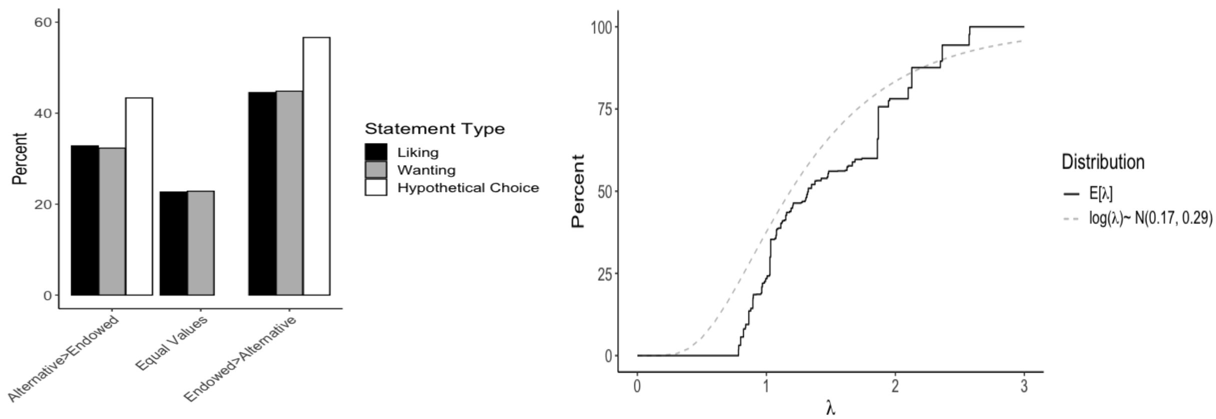
1.4.1 Stage 1: Gain-Loss Attitudes

In Stage 1, we collect three critical preference statements for the endowed and alternative object. These statements are used to infer the gain-loss attitude for each individual. Figure 1.3, Panel A provides histograms for our three preference statements: hypothetical choice, and wanting and liking ratings for the two objects. We summarize the direction of preference in ratings statements using the ordinal information of rating the endowed object higher than the alternative, giving them equal rating, or rating the alternative higher. These values are aggregated across the four potential endowments of Stage 1. Given random assignment of endowed objects and the counterbalanced design, the distributions of preference statements should be identical between endowed and alternative objects. Instead, all three distributions show a clear preference for the subject's endowed object relative to the alternative.[16] For each measure we reject the null hypothesis that stated preferences are equal over the endowed and alternative objects.[17]

The collected preference statements show a clear endowment effect. Thirty-eight percent of subjects (385 of 1024) state that they would hypothetically choose, strictly like, and strictly

[16]Fifty-seven percent of subjects state that they would choose their endowed object, 45 percent provide a higher liking rating for their endowed object compared to 33 percent for the alternative, and 45 percent provide a higher wanting rating for their endowed object compared to 32 percent for the alternative. The different preference statements are remarkably correlated within individual. The pairwise correlations between hypothetical choice, relative liking, and relative wanting statements all exceed 0.7.

[17]Two sided t-tests comparing "Endowed > Alternative" to "Alternative>Endowed" are significant for all statements (Liking: $t = 5.48$, Wanting: $t = 5.86$, Hypothetical Choice: $t = 6.06$, $(p < 0.01)$ for all).



(a) Stated Preferences and Endowments

(b) Distributions of Gain-Loss Attitudes

Figure 1.3: **Stage 1 Statements and Implied Gain-Loss Attitudes**

Notes: Panel A: relative preference statements for endowed and alternative objects. Wanting and liking ratings mapped from a nine-point scale onto a categorical variable representing whether the rating was higher for the endowed or alternative object, or equal for the two. Two sided t-tests comparing "Endowed¿Alternative" to "Alternative¿Endowed" are significant for all statements (Liking: $t = 5.48$, Wanting: $t = 5.86$, Hypothetical Choice: $t = 6.06$, $p < 0.01$ for all). Panel B: The dashed line represents estimated distribution $log(\lambda) \sim N(0.17, 0.29)$, resulting in a mean $\lambda = 1.37$, a median $\lambda = 1.18$, and $P(\lambda < 1) = 0.38$. Solid line represents the expected value of λ conditional on the Stage 1 statements, $E[\lambda]$, as described in the main text.

want their endowed object. Importantly, however, twenty-six percent of subjects (262 of 1024) exhibit the opposite pattern of hypothetically choosing, strictly liking, and strictly wanting the alternative object.

Estimation of Gain-Loss Attitudes

The preference statements summarized in Figure 1.3, Panel A provide a basis for estimating the distribution of gain-loss attitudes or utility values using standard mixed logit methods. Consider an individual endowed with object X who is asked to provide preference statements between X and Y. Under the KR model, the individual will state a preference in the form of a higher liking value for X, higher wanting value for X, or hypothetical choice of X if

$$u(X,0|X,0) - u(0,Y|X,0) > \delta,$$

where δ captures the possibility of equal rating levels (note $\delta = 0$ for our hypothetical choice data as there was no possibility of stating indifference). Under our functional form assumptions for KR utility — piecewise linear gain-loss utility with $\eta = 1$ — such a preference statement occurs with probability

$$Prob_{X|X} = Prob((1+\lambda) - 2\frac{Y}{X} - \delta_X > 0),$$

where $\delta_X \equiv \frac{\delta}{X}$. Similarly, an individual endowed with X would state a preference for Y with probability

$$Prob_{Y|X} = Prob(2\frac{Y}{X} - (1+\lambda) - \delta_X > 0),$$

and, where appropriate, would provide equal ratings for the two objects with probability

$$Prob_{E|X} = 1 - Prob_{X|X} - Prob_{Y|X}.$$

Symmetric formulations exist for individuals endowed with object Y.[18] These probabilities summarize the connection between the relative preference statements illustrated in Figure 1.3, Panel A and our structural model of gain-loss attitudes in Stage 1.

We make five assumptions to estimate the distribution of gain-loss attitudes from Stage 1 preference statements. First, following the structure of our design, our exercise recognizes the exogenously endowed object, X, as the subject's reference point. This formulation of a fixed

[18] That is

$$\begin{aligned} Prob_{X|Y} &= Prob(2 - (1+\lambda)\frac{Y}{X} - \delta_X > 0) \\ Prob_{Y|Y} &= Prob((1+\lambda)\frac{Y}{X} - 2 - \delta_X > 0) \\ Prob_{E|Y} &= 1 - Prob_{X|Y} - Prob_{Y|Y}. \end{aligned}$$

reference point is inherent to our design, which elicits preference statements for both objects under this fixed endowment (and prior to any discussion of exchange).[19] Second, we assume that $Prob(\cdot)$ is the logistic function leading to logit choice. Third, we assume that the value λ is drawn from a log-normal distribution with $log(\lambda) \sim N(\mu_\lambda, \sigma_\lambda^2)$, leading to a mixed logit formulation. Fourth, we assume the relative utility value $\frac{Y}{X}$ is homogeneous in the population and a parameter to be estimated. Fifth, we assume δ_X to be fixed at value 0.55, a value that our prior research indicated to be an appropriate aggregate value.[20] We make a similar set of assumptions to estimate the distribution of utility values, $\frac{Y}{X}$: logit choice, homogeneous gain-loss attitudes, $\frac{Y}{X}$ drawn log-normal with $log(\frac{Y}{X}) \sim N(\mu_{\frac{Y}{X}}, \sigma^2_{\frac{Y}{X}})$, and $\delta_X = 0.55$. Appendix A.3 provides the complete simulated likelihood formulation for both the estimation of heterogeneous gain-loss attitudes and heterogeneous utilities.

Table 1.2 provides estimates of these two models using the Method of Simulated Likelihood with 1000 Halton draws for each observation for relevant heterogeneous parameters. Each subject provides three observations to this estimation exercise: their hypothetical choice, their relative liking statement, and their relative wanting statement (assumed independent). We provide separate utility estimates for our initial and replication sample to account for the evolution of tastes over the three years between our studies.

In the first two columns of Table 1.2 we provide estimates and standard errors assuming heterogeneous gain-loss attitudes with homogeneous utility values. In both our initial and replication studies we find the pen set has relatively lower utility than the USB stick. The picnic mat

[19]Though implausible under our design, potential alternatives to this formulation might be the CPE construction or to assume the subject considers retaining their endowed object, X, and gaining the alternative, Y, evaluating $U(X,Y|X,0) = X + (1+\eta)Y$. Importantly, neither alternative would deliver any information on the key gain-loss parameter, λ, and so both would yield null predictions for heterogeneous treatment effects in Stage 2. As such, in addition to the structure of our design, the results we document further invalidate these formulations.

[20]See Appendix A.1 or `https://papers.ssrn.com/sol3/papers.cfm?abstractid=3170670` for these prior estimates. We found some substantial sensitivities of the value σ_λ^2 to attempting to estimate δ_X alongside the other parameters. The challenge is intuitive: a larger value of δ_X implies individuals should more frequently give the two objects equal ratings. All else equal, a higher variance of gain-loss attitudes is required to justify the relative infrequency of such observations. Appendix Table A.3 provides analysis setting δ_X at several different values and demonstrating corresponding sensitivity for the variance of gain-loss attitudes.

Table 1.2: **Method of Simulated Likelihood Estimates**

	(1)	(2)	(3)	(4)
	Estimate	(Std. Error)	Estimate	(Std. Error)
	Heterogeneous λ		*Heterogeneous* $\frac{Y}{X}$	
Gain-Loss Attitudes:				
$\hat{\lambda}$	1.37	(0.08)	1.31	(0.05)
$\hat{\mu_{\lambda}}$	0.17	(0.07)	-	-
$\hat{\sigma^2_{\lambda}}$	0.29	(0.21)	-	-
Pair 1 Utilities (USB Stick (X) - Pen Set (Y)) *:*				
$\hat{\frac{Y}{X}}$ (Initial)	0.62	(0.04)	0.62	(0.03)
$\hat{\frac{Y}{X}}$ (Replication)	0.61	(0.04)	-	-
$\hat{\mu_{\frac{Y}{X}}}$	-	-	-0.55	(0.09)
$\hat{\sigma^2_{\frac{Y}{X}}}$	-	-	0.16	(0.13)
Pair 2 Utilities (Picnic Mat (X) - Thermos (Y))*:*				
$\hat{\frac{Y}{X}}$ (Initial)	1.11	(0.03)	1.03	(0.03)
$\hat{\frac{Y}{X}}$ (Replication)	0.88	(0.04)	-	-
$\hat{\mu_{\frac{Y}{X}}}$	-	-	-0.03	(0.04)
$\hat{\sigma^2_{\frac{Y}{X}}}$	-	-	0.12	(0.08)
Discernibility:				
δ_X	0.55	-	0.55	-
# Observations	3,072		3,072	
Log-Likelihood	-2743.13		-2751.72	
Akaike's Information Criterion (AIC)	5498.26		5513.44	

Notes: Method of simulated likelihood estimates. Standard errors in parentheses.

and thermos carry more similar utility, with aggregate tastes evolving over the years of our study. For gain-loss utility we estimate the parameters of the log-normal distribution to be $\hat{\mu}_\lambda = 0.17$ (s.e. = 0.07). and $\hat{\sigma}^2_\lambda = 0.29$ (0.21). This log-normal distribution has estimated mean equal to $exp(\hat{\mu}_\lambda + 1/2\hat{\sigma}^2_\lambda) = 1.37$ (0.08). Figure 1.3, Panel B provides the estimated distribution of gain-loss attitudes, $log(\lambda) \sim N(0.17, 0.29)$, as a dashed gray line. Under this distribution of gain-loss attitudes, there is a 38 percent chance of an individual being gain-loving, $\lambda < 1$. In order to provide a confidence interval for the probability of $\lambda < 1$, we simulate 100,000 values for μ_λ and σ^2_λ from the joint normal distribution implied by the coefficient vector and estimated variance-covariance matrix underlying Table 1.2, columns (1) and (2). We calculate the corresponding probability of $\lambda < 1$ for each simulated pair to construct the relevant distribution, delivering a bootstrapped 95% confidence interval of $[0.13, 0.49]$.[21]

In the second two columns of Table 1.2, we also provide estimates assuming homogeneous gain-loss attitudes and heterogeneous utility values. There, we find an aggregate $\hat{\lambda} = 1.31$ (0.05), close to the previously reported mean value, and substantial variation in intrinsic utility values with estimated means also close to the previously reported values. Table 1.2 provides the simulated likelihood and Akaike's Information Criterion (AIC) values for our two estimated models. Heterogeneous gain-loss attitudes has substantially better likelihood values, which, by the AIC, justify the additional degree of freedom it uses relative to the model with heterogeneous utilities (six versus five degrees of freedom). Of course, our project is not predicated on in-sample fit in Stage 1, but rather using Stage 1 measures out-of-sample to predict behavior in Stage 2.

[21]Standard errors for $\hat{\sigma}^2_\lambda$ and $exp(\hat{\mu}_\lambda + 1/2\hat{\sigma}^2_\lambda)$ calculated using the delta method from estimated values of $\hat{\mu}_\lambda = 0.17$ (0.07) and $log(\hat{\sigma}_\lambda) = -0.62$ (0.37). The parameters $\hat{\mu}_\lambda$ and $log(\hat{\sigma}_\lambda)$ have estimated covariance -0.021, which is used for generating simulated confidence interval on the probability of $\lambda < 1$.

Individual Gain-Loss Attitudes

Moving from the distribution of gain-loss attitudes to an expected value of λ for each individual is a straightforward step. As proposed in Train (2009), we simulate the distribution of λ, and calculate the expected λ for each possible Stage 1 statement profile. For example, under the estimated log-normal density, $g(\lambda)$, one simulates $Prob_{X|X}(\lambda)$, and the expected value of λ given a preference for X when endowed with X as

$$E[\lambda_{X|X}] = \int \lambda \frac{Prob_{X|X}(\lambda)g(\lambda)}{\int Prob_{X|X}(\lambda)g(\lambda)d\lambda} d\lambda.$$

For each endowment, subjects could provide one of two hypothetical choice statements, one of three relative liking statements, and one of three relative wanting statements, yielding 18 potential statement profiles. With four endowments, there are 72 potential profiles, each with an implication for the expected value of λ.[22] We extend the above example to construct the probability of each such profile assuming independence between the simulated probabilities for hypothetical choice, liking, and wanting statements.

Figure 1.3, Panel B provides the distribution of $E[\lambda]$ implied by Stage 1 preference statements as the solid black line. This distribution has mean 1.49, median 1.32, with 23 percent of subjects exhibiting $E[\lambda] < 1$. The distribution of $E[\lambda]$ is similar in shape and key statistics to the underlying log-normal estimates. However, the distribution of $E[\lambda]$ does exhibit fewer extreme gain-loving and loss-averse observations than its underlying distribution. Individual heterogeneity in $E[\lambda]$ in hand, we are equipped to analyze heterogeneous treatment effects.

[22]Note that because we allow for different utilities in our initial study and replication, there are 72 such values for each.

1.4.2 Stage 2: Probabilistic Forced Exchange and Heterogeneous Treatment Effects

Table 1.3 presents results from Stage 2 of our study, with linear probability models for the indicator $Exchange(=1)$. Column (1) demonstrates the average treatment effect without accounting for heterogeneity in gain-loss attitudes. In Condition B, 38 percent of subjects choose to exchange. Comparing this value to the neoclassical benchmark of 50 percent indicates a significant endowment effect in Condition B, $F_{1,1022} = 25.66,\ (p < 0.01)$. The substantial endowment effect observed in the Condition B is unaffected by probabilistic forced exchange. In contrast to the prediction of the KR model with universal loss aversion (which would predict a positive treatment effect), we find that Condition F decreases the probability of exchange by -0.4 percentage points. We fail to reject that this treatment effect is zero, $F_{1,1022} = 0.01,\ (p = 0.91)$.

Figure 1.4 and Table 1.3, column (2) interact Stage 2 condition assignment with gain-loss attitudes measured in Stage 1. In Panel A of Figure 1.4, we plot binned values of $E[\lambda]$ from Stage 1 against the probability of exchange in Condition B of Stage 2.[23] Our Stage 1 value of gain-loss attitudes closely correlates with prevalence of endowment effects in Stage 2. Subjects with low values of $E[\lambda] < 1$ in Stage 1 exchange more frequently than subjects with $E[\lambda] > 1$. Table 1.3, column (2) indicates a substantial correlation between $E[\lambda]$ and Condition B behavior, with a statistically significant slope coefficient of -0.136, (robust s.e. = 0.041), $F_{1,1020} = 11.23,\ (p < 0.01)$.

Gain-loss attitudes measured with one set of objects in Stage 1 are predictive of baseline exchange levels for a different set of objects in Stage 2. This intuitive connection between Stage 1 gain-loss attitudes and behavior in Condition B of Stage 2 is not strictly within the KR model's CPE predictions, which predicts no standard endowment effect Condition B.[24] In Appendix

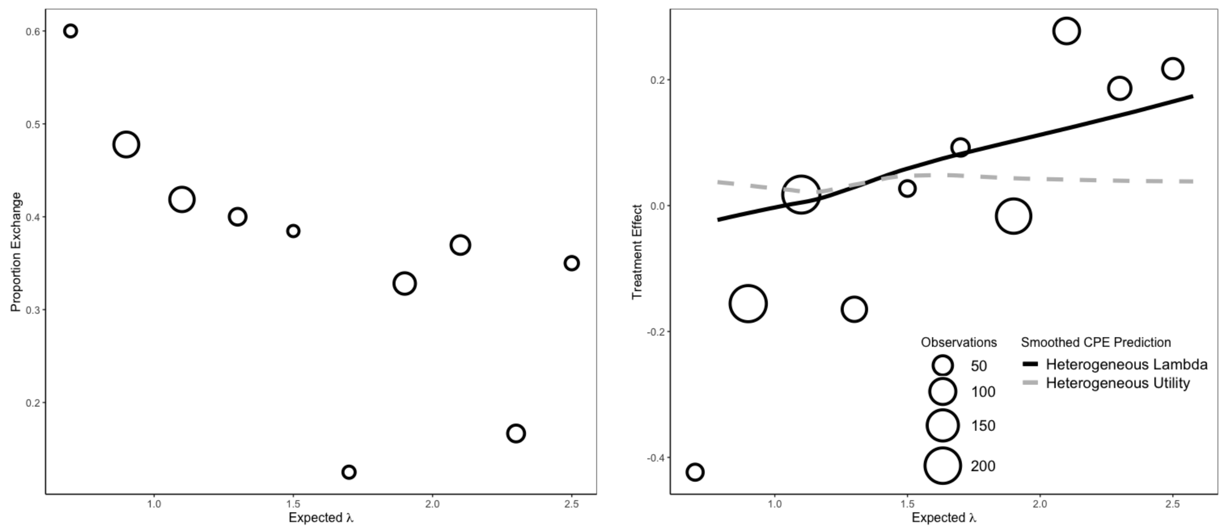
[23]The size of each plotted point corresponds to the number of subjects with $E[\lambda]$ in a bin of size 0.2 around the reported value.

[24]Note that our Stage 1 design exogenously endows subjects with objects and elicits preference statements prior to any discussion of potential replacement. In this case the KR model coincides exactly with the standard model of reference-dependent preferences with a fixed reference point, which allows us to elicit gain-loss attitudes in Stage

Table 1.3: **Exchange Behavior and Probabilistic Forced Exchange**

	Dependent Variable: Exchange (=1)		
	(1)	(2)	(3)
Condition F	-0.004	-0.340	-0.004
	(0.031)	(0.087)	(0.031)
$E[\lambda]$		-0.136	
		(0.041)	
Condition F * $E[\lambda]$		0.225	
		(0.054)	
Reduced Form Measure			-0.050
			(0.015)
Condition F * Reduced Form			0.077
			(0.020)
Constant (Condition B)	0.380	0.584	0.380
	(0.024)	(0.067)	(0.023)
R-Squared	0.000	0.017	0.014
# Observations	1024	1024	1024
H_0 : Zero Endowment Effect in B	$F_{1,1022} = 25.66$	$F_{1,1020} = 1.57$	$F_{1,1020} = 26.07$
	$(p < 0.01)$	$(p = 0.21)$	$(p < 0.01)$
H_0 : Zero Treatment Effect (F-B)	$F_{1,1022} = 0.01$	$F_{1,1020} = 15.12$	$F_{1,1020} = 0.02$
	$(p = 0.91)$	$(p < 0.01)$	$(p = 0.90)$
H_0 : Gain-Loss Attitudes $\perp$ Exchange in B		$F_{1,1020} = 11.23$	$F_{1,1020} = 10.69$
		$(p < 0.01)$	$(p < 0.01)$
H_0 : Gain-Loss Attitudes $\perp$ Treatment Effect		$F_{1,1020} = 17.25$	$F_{1,1020} = 14.65$
		$(p < 0.01)$	$(p < 0.01)$

Notes: Ordinary least square regression. Robust standard errors in parentheses. Null hypotheses tested for 1) zero baseline endowment effect regression (Constant coefficient = 0.5); 2) zero treatment effect (Condition F coefficient= 0); 3) no relationship between gain-loss attitudes and behavior in Condition B behavior ($E[\lambda]$ or Reduced Form Measure coefficient = 0); 4) constant treatment effect over gain-loss attitudes (Condition F * $E[\lambda]$ or Condition F * Reduced Form coefficient = 0). F-statistics and two-sided p-values reported.



(a) Condition B Behavior (b) Heterogeneous Treatment Effect

Figure 1.4: **Stage 1 Gain-Loss Attitudes and Stage 2 Behavior**
Notes: Panel A presents the observed proportion of exchange in Condition B by the expected value of λ, binned from 0.6 to 2.8 in increments of 0.2 and assigned the midpoint of the relevant interval. Panel B presents the observed treatment effect (exchange in Condition F- Condition B) as well as two smoothed KR predictions when we attribute heterogeneity to either gain-loss attitudes or utility values.

A.2 we provide analysis for PE, which permits both an endowment effect in Condition B and predicts the observed relationship between Stage 1 gain-loss attitudes and Condition B behavior. Critically, even though CPE and PE (as well as the PPE refinement) differ in their predictions for the *level* of Condition B behavior, they make the same qualitative prediction for treatment effects between Conditions B and F. Moreover, all three formulations make the same qualitative prediction of heterogeneous treatment effects over gain-loss attitudes (see Appendix A.2 for detail). This highlights the importance of investigating *treatment effects* rather than *levels* in our empirical design.

Figure 1.4, Panel B plots $E[\lambda]$ from Stage 1 against the treatment effect in Stage 2, Condition F-Condition B. Table 1.3, column (2) provides corresponding regression statistics. For subjects with low values of $E[\lambda] < 1$, probabilistic forced exchange decreases exchange, while for those with values $E[\lambda] > 1$, Condition F generally increases trading probabilities. The inter-

1. See Kőszegi and Rabin (2006) for additional discussion of this point in the particular context of the endowment effect.

action effect between $E[\lambda]$ in Stage 1 and the Stage 2 treatment effect reported is 0.225 (0.054), and statistically significant at all conventional levels, $F_{1,1020} = 17.25$, $(p < 0.01)$. Appendix Table A.4 provides the same analysis with standard errors clustered at the session level and reaches the same statistical conclusions. Loss-averse and gain-loving subjects respond differently to probabilistic forced exchange, delivering the heterogeneous treatment effects predicted by KR.

Also graphed in Figure 1.4, Panel B are two smoothed predictions for theoretical treatment effects. The solid black line corresponds to the KR-CPE prediction under heterogeneous gain-loss attitudes. To construct this prediction, we use the values of $E[\lambda]$ obtained in our sample combined with the object assignments in Stage 2, and calculate the probability of exchange in CPE under logit choice. Under our estimated mixed logit model of heterogeneous gain-loss attitudes in Table 1.2, columns (1) and (2), this is the predicted relationship between $E[\lambda]$ and the treatment effect. The observed heterogeneous treatment effects are in line with the sign and magnitude predicted by the KR model.

The theoretically predicted and observed heterogeneity in treatment effects provides an alternative interpretation to the null aggregate result presented in Table 1.3, column (1). Rather than indicating a failure of the KR model, it indicates a failure to account for the model's heterogeneous predictions. Indeed, the predicted average treatment effect under our mixed logit model is quite close to the zero average treatment effect observed in column (1). Even with a minority of gain-loving subjects, the average treatment effect is predicted to be only 5.9 percentage points under KR-CPE.[25] In our concluding discussion, we reflect on this relatively low average treatment effect for conducting appropriately powered aggregate experiments on the KR model.

The dashed gray line in Figure 1.4, Panel B corresponds to the predictions from an alternative model of Stage 1 behavior: that differences in preference statements are driven by heterogeneous utilities rather than heterogeneous gain-loss attitudes. By design, such an interpretation of Stage 1 should have no predictive power in Stage 2. Under the model estimated in Table 1.2,

[25] This can be ascertained visually in Figure 1.4 as the weighted average value of the black KR-CPE prediction over the values of $E[\lambda]$.

columns (3) and (4), we construct KR-CPE predictions analogous to those for heterogeneous gain-loss attitudes. Under these predictions, $E[\lambda]$ is simply a misspecified object, which should be orthogonal to treatment effects given the random assignment in Stage 2. The calculations yield exactly this prediction, indicating a positive treatment effect of around 0.033 — consistent with the homogeneous $\lambda = 1.31$ — and no relation to the "misspecified" value of $E[\lambda]$.[26] Stated differently, had we misattributed Stage 1 behavior to gain-loss attitudes rather than the "true" model of heterogeneous utility, Stage 2 treatment effects should have corresponded to the dashed gray line. The data resoundingly reject this interpretation: we reject the null hypothesis of a constant treatment effect of 0.033, $F_{2,1020} = 9.20,\ (p < 0.01)$.

Reduced Form Exercise

Thus far, we have interpreted Stage 1 behavior through the lens of a structural model estimating the heterogeneity of gain-loss attitudes. In the final column of Table 1.3, we provide one additional reduced-form analysis to ensure our results are not a spurious product of our structural assumptions. Specifically, we first conduct principal components analysis on our three Stage 1 preference statements and reduce the data to the first principal component. This first component captures around 70 percent of the variation in the three preference statements. We regress this component on Stage 1 object assignment interacted with replication and use the residuals as our reduced form measure. These residuals capture the variation in preference statements across subjects taking into account their assigned object. An individual who disproportionately likes their assigned object relative to mean preferences is plausibly more loss averse than one who exhibits a residual in the opposite direction. Column (3) shows a close correspondence in our structural and reduced form results. A significant interaction effect of 0.077 (0.020) is observed between Condition F and the reduced form measure, echoing our structural results on heterogeneous treatment effects over gain-loss attitudes, $F_{1,1020} = 14.65,\ (p < 0.01)$.

[26]The slight variation in the dashed gray line prediction in Figure 1.4 is due to the assigned objects in Stage 1 and Stage 2 and their heterogeneous valuations.

1.4.3 Robustness Tests

Complementarities Between Stages

Our results indicate that gain-loss attitudes measured with one pair of objects in Stage 1 are predictive of exchange behavior for a distinct counterbalanced pair of objects in Stage 2. Though we attempted to choose Stage 1 and Stage 2 objects that would have no plausible complementarities, if some un-modeled, unintentional complementarity did exist, it might spuriously appear as predictive power across stages. For example, a subject might state a preference for or against both of their endowed objects in order to consume both endowed objects or both alternatives together. Note that this mechanism cannot explain the Stage 2 treatment effect, but could perhaps provide a rationale for the correlations documented between Stage 1 gain-loss attitudes and exchange in Stage 2, Condition B.

Importantly, our Stage 1 design was constructed with one piece of random variation that serves to break complementarities between objects across stages. After providing their preference statements, half of subjects have their endowed object replaced with the alternative. If our results are reproduced both for subjects who have their endowed object replaced and those who do not, then explanations based upon accidental complementarities cannot be relevant for our results. To explore this possibility, Table 1.4 reproduces the structural results of Table 1.3 separately by individuals who do and do not have their Stage 1 endowed object replaced. For both groups, our results are maintained. Appendix Table A.5 provides the same analysis with standard errors clustered at the session level and reaches the same statistical conclusions.

Replication Consistency and Additional Controls

Our results to here have combined the data from our initial and replication studies. Table 1.5 reproduces the structural results of Table 1.3 separately for the two samples. The null aggregate treatment effect and heterogeneous treatment effects over gain-loss attitudes are produced

Table 1.4: **Stage 2 Behavior and Stage 1 Experience**

	Dependent Variable: Exchange (=1)			
	Stage 1 Object Not Replaced		Stage 1 Object Replaced	
	(1)	(2)	(3)	(4)
Condition F	0.013	-0.255	-0.019	-0.418
	(0.044)	(0.126)	(0.043)	(0.122)
$E[\lambda]$		-0.121		-0.153
		(0.057)		(0.058)
Condition F * $E[\lambda]$		0.176		0.272
		(0.077)		(0.077)
Constant (Condition B)	0.386	0.569	0.374	0.600
	(0.033)	(0.094)	(0.034)	(0.095)
R-Squared	0.000	0.011	0.000	0.024
# Observations	511	511	513	513
H_0 : Zero Endowment Effect in B	$F_{1,509} = 11.73$	$F_{1,507} = 0.54$	$F_{1,511} = 13.96$	$F_{1,509} = 1.11$
	$(p < 0.01)$	$(p = 0.46)$	$(p < 0.01)$	$(p = 0.29)$
H_0 : Zero Treatment Effect (F-B)	$F_{1,509} = .08$	$F_{1,507} = 4.08$	$F_{1,511} = 0.19$	$F_{1,509} = 11.76$
	$(p = 0.77)$	$(p = 0.04)$	$(p = .67)$	$(p < 0.01)$
H_0 : Gain-Loss Attitudes $\perp$ Exchange in B		$F_{1,507} = 4.52$		$F_{1,509} = 6.96$
		$(p = 0.03)$		$(p < 0.01)$
H_0 : Gain-Loss Attitudes $\perp$ Treatment Effect		$F_{1,507} = 5.25$		$F_{1,509} = 12.53$
		$(p = 0.02)$		$(p < 0.01)$

Notes: Ordinary least square regression. Robust standard errors in parentheses. Null hypotheses tested for 1) zero baseline endowment effect regression (Constant coefficient = 0.5); 2) zero treatment effect (Condition F coefficient= 0); 3) no relationship between gain-loss attitudes and behavior in Condition B behavior ($E[\lambda]$ coefficient = 0); 4) constant treatment effect over gain-loss attitudes (Condition F * $E[\lambda]$ = 0). F-statistics and two-sided p-values reported.

in both our initial and replication studies. Quantitatively the observed relationships between gain-loss attitudes and exchange behavior are broadly consistent, though the replication has less precise estimates due to the smaller sample size.

Our replication study was conducted to assure confidence in our previously obtained heterogeneous treatment effects. The registration of our pre-analysis plan, including power calculations, can be found at `https://www.socialscienceregistry.org/trials/3124`. The analysis proposed there carries one important difference to that conducted here: our proposed methodology for identifying gain-loss attitudes was based on standard logit, rather than mixed logit methods. This was the methodology used in our initial draft posted at `https://papers.ssrn.com/sol3/papers.cfm?abstractid=3170670`. Advice from an anonymous referee highlighted the value of the mixed logit methods that we currently conduct. For completeness, in Appendix A.1 we provide the pre-registered replication analysis. There, as well, we find a striking consistency between the results obtained in our initial and replication samples.

Table 1.5: Replication Consistency and Additional Controls

	Dependent Variable: Exchange (=1)				
	Initial Study		Replication Study		
	(1)	(2)	(3)	(4)	(5)
Condition F	0.004	-0.409	-0.010	-0.239	-0.805
	(0.040)	(0.117)	(0.048)	(0.137)	(0.415)
$E[\lambda]$		-0.159		-0.103	-0.116
		(0.054)		(0.065)	(0.066)
Condition F * $E[\lambda]$		0.266		0.161	0.174
		(0.070)		(0.089)	(0.091)
Constant (Condition B)	0.365	0.616	0.399	0.542	0.917
	(0.032)	(0.093)	(0.036)	(0.099)	(0.321)
Additional Controls	No	No	No	No	Yes
Additional Interactions	No	No	No	No	Yes
R-Squared	0.000	0.023	0.000	0.008	0.060
# Observations	607	607	417	417	417
H_0 : Zero Endowment Effect in B	$F_{1,605} = 18.32$	$F_{1,603} = 1.55$	$F_{1,415} = 7.97$	$F_{1,413} = 0.18$	$F_{1,393} = 1.70$
	$(p < 0.01)$	$(p = 0.21)$	$(p < 0.01)$	$(p = 0.67)$	$(p = 0.19)$
H_0 : Zero Treatment Effect (F-B)	$F_{1,605} = 0.01$	$F_{1,603} = 12.29$	$F_{1,415} = 0.05$	$F_{1,413} = 3.06$	$F_{1,393} = 3.76$
	$(p = 0.91)$	$(p < 0.01)$	$(p = 0.83)$	$(p = 0.08)$	$(p = 0.05)$
H_0 : Gain-Loss Attitudes $\perp$ Exchange in B		$F_{1,603} = 8.84$		$F_{1,413} = 2.50$	$F_{1,393} = 3.08$
		$(p < 0.01)$		$(p = 0.12)$	$(p = 0.08)$
H_0 : Gain-Loss Attitudes $\perp$ Treatment Effect		$F_{1,603} = 14.66$		$F_{1,413} = 3.28$	$F_{1,393} = 3.69$
		$(p < 0.01)$		$(p = 0.07)$	$(p = 0.06)$
H_0 : Add'l Controls & Interactions $\perp$ Exchange					$F_{20,393} = 1.21$
					$(p = 0.24)$

Notes: Ordinary least square regression. Robust standard errors in parentheses. Null hypotheses tested for 1) zero baseline endowment effect regression (Constant coefficient = 0.5); 2) zero treatment effect (Condition F coefficient= 0); 3) no relationship between gain-loss attitudes and behavior in Condition B behavior ($E[\lambda]$ = 0); 4) constant treatment effect over gain-loss attitudes (Condition F * $E[\lambda]$ = 0); 5) no effect of additional controls or interactions (all coefficients = 0). Additional controls include: gender, age, educational status, monthly income bracket, knowledge of economics, composite Raven matrices score, composite CRT score, and fixed effects for experimental assistant. Interactions include all controls interacted with Condition F. F-statistics and two-sided p-values reported.

Our results demonstrate heterogeneous treatment effects over measured gain-loss attitudes. Stage 1 gain-loss attitudes represent the sole dimension of individual differences in this exercise and are closely predictive of Stage 2 behavior. A plausible critique of this approach is that it may be subject to omitted variable bias, with the documented correlations driven by other dimensions of heterogeneity. Importantly, in our replication data we have access to a rich set of covariates, and can control for age, gender, household income (in one of seven brackets), educational status, knowledge of economics, a composite score on the Cognitive Reflection Test (Frederick, 2005), and a composite score on a series of Raven matrices. Additionally, in our replication data, four different experimental assistants conducted the sessions providing another potential dimension of heterogeneity. In Table 1.5, column (5) we also interact each of these covariates with Condition F and include the main and interacted effects in regressions for our replication sample. The coefficients of interest for gain-loss attitudes and heterogeneous treatment effects are quite similar with and without these additional controls. If anything the results grow stronger when controlling for the rich set of control variables between columns (4) and (5). Importantly, we identify no other dimension of heterogeneity as being important for explaining Stage 2 behavior and fail to reject the null hypothesis that all other main effects and interactions are zero. Appendix Table A.6 provides the same analysis with standard errors clustered at the session level and reaches even stronger statistical conclusions.

1.5 Discussion and Conclusion

Expectations-based reference-dependent preferences (Kőszegi and Rabin, 2006, 2007) (KR) represent a key advance in behavioral economics, but a host of conflicting evidence for the theory exists. In this paper, we aimed to reconcile this conflicting evidence by explicitly recognizing and evaluating heterogeneity in gain-loss attitudes. Heterogeneity is critical both because the model's comparative statics change sign depending on the level of gain-loss attitudes and because prior work has noted that loss aversion is by no means a universal characteristic.

We measure gain-loss attitudes by evaluating preference statements for a first pair of objects and then place subjects in an exchange environment where they make choices over a second, different pair of objects. We show that explicitly accounting for the heterogeneity in gain-loss attitudes restores behavior in line with KR predictions. Individual gain-loss attitudes are predictive of exchange behavior in a standard exchange environment. Using a mechanism of probabilistic forced exchange, we then show that individuals who are measured to be loss-averse grow more willing to exchange when probabilistically forced to do so; and individuals who are measured to be gain-loving grow less willing to exchange. These findings, and the magnitudes of the observed treatment effects, are closely in line with the predictions of the KR model.

Our results may help to reconcile conflicting results in the empirical study of the KR model (Ericson and Fuster, 2011; Goette, Harms, et al., 2016; Heffetz and List, 2014) and follow naturally from the broad recognition of heterogeneity in gain-loss attitudes (Chapman, Dean, et al., 2017; Erev et al., 2008; Harinck et al., 2007; J. Knetsch and Wong, 2009; Nicolau, 2012; Sokol-Hessner et al., 2009; Sprenger, 2015). If we are to recognize that loss aversion is not a universal trait, we must also recognize it as a confound of first-order importance for empirical tests of expectations-based reference dependence.

Two factors are central to the confounding effects of heterogeneity for aggregate studies of KR preferences. First, we show that treatment effects that have been used to test the KR model change sign at $\lambda = 1$ and do not aggregate linearly over gain-loss attitudes. Even with loss aversion on average, gain-loving individuals can have a substantial effect on the aggregate treatment effect. Under our estimates for gain-loss attitudes with loss aversion on average, the average treatment effect in our sample should be approximately 5.9 percentage points under logit choice. Second, if aggregation over gain-loss types reduces predicted treatment effects, it will also induce larger required sample sizes for reliably-powered studies. To identify a 5.9 percentage point treatment effect with 80 percent power, a sample of 2250 is required. This value is more than twice as large as the sample size of Goette, Harms, et al. (2016), and ten times as large as the sample sizes of Heffetz and List (2014) and Ericson and Fuster (2011). A reasonable conclusion would be that no prior aggregate test of the KR model is appropriately powered.

Though we provide replication of our results for exchange behavior, future work should examine other domains of application and other methodologies for identifying gain-loss attitudes. Relevant steps include examination of reference-dependent labor supply, lottery choice, and financial behavior.

By design, our findings cannot be attributed to specific models for Stage 1 behavior such as heterogeneous valuations for objects or a 'grass-is-greener' rationale for preferring a non-endowed object. Both of these views may rationalize heterogeneous behavior in Stage 1, but deliver no heterogeneity in treatment effects in Stage 2. Rationales based on confusion, noisy response, or sampling variation are similarly ruled out by our data, as they cannot account for the predictive power of measured gain-loss attitudes for subsequent behavior in both conditions of Stage 2. Naturally, one may wish to seek alternate modeling degrees of freedom to rationalize the data presented here. Such models should be formulated to not only rationalize our data, but also make testable predictions.

1.6 Acknowledgements

This work is currently being prepared for submission for publication. Goette, Lorenz; Graeber, Thomas; Sprenger, Charles. The dissertation author was a principal investigator in this analysis.

Chapter 2

Gain-Loss Distributions in KR Exchange Paradigms: Bayesian Analysis of Experimental Literature

2.1 Introduction

Reference-dependent preferences are a core contribution of behavioral economics, with applications to labor supply, job search, and market entry and exit decisions (Abeler et al., 2011; Backus et al., 2017; DellaVigna et al., 2017; Fehr and Goette, 2007). Popularized by Kahneman and Tversky (1979) and Tversky and Kahneman (1992), initial models left reference points largely unspecified, granting researchers an important degree of freedom in selecting a reference point. A prominent advance by Kőszegi and Rabin (2006, 2007) (henceforth KR) sought to discipline these early models of reference dependence by endogenizing the reference point as rational expectations; in closing the model, this formulation yields novel, testable implications across various domains.

The KR mechanism has been the focus of several experimental investigations, notably (for this paper) in the endowment effect paradigm, with mixed results. The canonical design revolves around a manipulation of the endogenous reference point: fixing an endowment, variation in the possibility of exchanging one good for another should influence the fraction of individuals who wish to exchange. If reference points are grounded in rational expectations, experimental manipulations affecting the set of expected outcomes should alter attitudes towards exchange. In an initial piece of evidence, Ericson and Fuster (2011) (N=45) found behavior largely in line with KR. Shortly thereafter, Heffetz and List (2014) provided a series of complementary experiments, with results varying broadly over minor experimental design features; across three experiments (N=458), the authors find a null effect from their experimental manipulation, casting doubt on the findings in Ericson and Fuster (2011). Finally, a similar manipulation of the reference point was introduced by the design in Cerulli-Harms et al. (2019), which again demonstrated that participants were largely unaffected by the exogenous changes to exchange probabilities (N = 861).

In this manuscript, we evaluate what can be learned from this body of existing empirical results from tests of KR. A dominant interpretation suggests that KR is incompatible with this observed behavior in the lab; instead, it could be that some individuals hold the status quo as their referent, while others hold aspirations, and still others have some alternate formulation of expectations. However, we take up a secondary interpretation – that the KR formulation is correct, but experimental results hinge crucially on variation in the draws from the distribution of gain-loss attitudes. Viewed through this lens, the body of evidence speaks to the underlying distribution of gain-loss attitudes that prevails in the population.

There are several reasons to consider this alternative perspective. First, heterogeneity in gain-loss preferences is well documented, with gain-lovers (those who weight gains more than commensurately-sized losses) composing between 22% and 53% of the individuals across various populations (Chapman, Snowberg, et al., 2018). Importantly, treatment effects in the experimental tests of KR preferences theoretically change signs in accordance with gain-loss types, such that null effects are possible even with loss aversion on average. This hypothesis was put to the test by Goette, Graeber, et al. (2018) in a (replicated) experiment, providing evidence that out-of-sample behavior in a second stage respects gain-loss classification in a first stage, with treatment effects by gain-loss type hewing closely to KR theory.

Operating under the assumption that individuals behave according to KR leaves the distribution of gain-loss preferences as the central object governing observed variation in experimental behavior. As such, we can interpret the body of mixed experimental evidence through the lens of the model as informative about the underlying distribution of gain-loss attitudes, asking what distributions of gain-loss preferences would rationalize the data at hand. We conduct this empirical exercise by adapting Bayesian methods, to which these analyses are well suited, with structural KR preferences.

First, we deploy the estimator on synthetic datasets to ensure they reliably recover the distribution of gain-loss attitudes when the true model is known. Two candidate priors are considered to ensure robustness: one where roughly 5% of individuals are gain-loving and one derived from the Goette, Graeber, et al. (2018) results where about 35% are gain-loving. The former prior represents what we assess to be the popular academic opinion on the relative frequency of loss aversion and gain-lovingness in laboratory samples. In at least 90% of our simulated cases, regardless of the prior, the true distribution lies in the credible interval of the estimated distribution of gain-loss attitudes. Having demonstrated this proof of concept, we deploy our estimator on each of the data sets in turn. Our posterior estimates based on the examination of the body of experimental data suggest that gain-lovers make up between 35% and 55% of these lab populations, regardless of prior. These findings are in line with the evidence from Chapman, Snowberg, et al. (2018) and Goette, Graeber, et al. (2018), who report a proportion of 53% and 38% in their respective populations.

The paper proceeds as follows. In Section 2.2, we outline the KR model, providing examples within the context of exchange experiments. Section 2.3 discusses some nuances of our estimator, and highlights its appealing properties through synthetic recovery exercises. Section 2.4 describes the existing body of experimental evidence for KR in the exchange context. We apply our estimator to the existing data in Section 2.5 and conclude in Section 2.6.

2.2 KR Model

The models in Kőszegi and Rabin (2006, 2007) represent an important advance on earlier models of reference-dependent preferences, endogenizing reference points as rational expectations. Our estimator relies on the assumption that individuals behave according to KR, so it is important to discuss the details of the model in the context of the exchange experiments.

Formally, let $\mathbf{c} = (m_X, m_Y)$, $\mathbf{r} = (r_X, r_Y)$ represent the vectors of intrinsic utility and reference utility from objects X, Y. The KR model specifies a utility function composed of the sum of two functions: intrinsic utility, $m(\mathbf{c}) = m_X + m_Y$, and gain-loss utility, $n(\mathbf{c}|\mathbf{r}) = \mu(m_X - r_X) + \mu(m_Y - r_Y)$, where $\mu(z) = \mathbf{1}(z \geq 0)[\eta z] + \mathbf{1}(z < 0)[\eta\lambda z]$. The parameter η determines the relative weight on gain-loss utility compared to intrinsic utility, whereas the parameter λ indicates the gain-loss preference, with $\lambda > 1$ indicating loss aversion and $0 < \lambda < 1$ indicating gain lovingness. Altogether, KR utility is defined as

$$u(\mathbf{c}|\mathbf{r}) = m_X + m_Y + \mu(m_X - r_X) + \mu(m_Y - r_Y).$$

An important component of the KR specification is that the reference point is determined endogenously as part of a forward-looking plan, and can be stochastic. That is, KR utility often takes as input a lottery (F) over consumption outcomes ($\mathbf{c}$) and evaluates them relative to a lottery (G) over reference points ($\mathbf{r}$), so

$$U(F|G) = \int \int u(\mathbf{c}|\mathbf{r}) dF(\mathbf{c}) dG(\mathbf{r}).$$

Throughout this analysis, we appeal to a particular equilibrium notion known as *Choice-Acclimating Personal Equilibrium* (CPE), and assume individuals behave accordingly. Under the concept of CPE, reference points catch up to the planned consumption, so that ex-ante utility is used as a basis for selecting among options in the choice set. Gains and losses therefore only arise when the consumption bundle (thus the reference point) is stochastic – otherwise, there is no potential for consumption to differ from expectations.

Choice-Acclimating Personal Equilibrium: Given a choice set, $\mathcal{D}$, of lotteries, F, over consumption outcomes $\mathbf{c}$, F is a CPE if

$$U(F|F) \geq U(F'|F')\ \forall F' \in \mathcal{D}.$$

2.2.1 Exogenous Shocks to Reference Points

Throughout the experiments we analyze, participants are endowed one of two goods – X or Y – and asked to choose their preferred good. If agents are allowed to choose among the two objects without any restrictions, the CPE model does not depend on λ. To see this, consider the CPE utilities for each choice: choosing X yields $U(X,0|X,0) = m_X$ while choosing Y yields $U(0,Y|0,Y) = m_Y$, so the unrestricted comparison between the two goods in CPE is purely a function of their intrinsic utilities.

In order to test the comparative statics of KR, experiments have introduced a permission-to-exchange (or relatedly, probabilistic forced exchange) mechanism that acts on the reference point of would-be traders (keepers) of their endowments. In the permission-to-exchange framework, as implemented in Ericson and Fuster (2011) and Heffetz and List (2014), the experimenter informs participants that with probability q, they are allowed to exchange their good should they chose to do so. If $q = 1$, a participant who planned to keep/exchange their endowment can do so with no difference, so CPE utility is still only a function of intrinsic utilities.

However, for $q < 1$, CPE utility for an individual planning to exchange hinges crucially on gain-loss preferences. To see this, consider an individual endowed good X but planning to exchange for good Y – their CPE utility is given by

$$U(q(0,Y)+(1-q)(X,0)|q(0,Y)+(1-q)(X,0)) =$$
$$q*m_Y+(1-q)m_X+q(1-q)\eta(1-\lambda)(m_X+m_Y).$$

To determine whether exchanging is optimal for them, the individual would compare the above utility to the CPE utility from keeping their endowment, $U(X,0|X,0) = m_X$. For a fixed q, the individual would prefer to exchange so long as

$$\frac{m_Y}{m_X} \geq \frac{1-(1-q)(1-\lambda)}{1+(1-q)(1-\lambda)},$$

and prefers to keep otherwise.

This choice depends critically on the ratio of intrinsic utilities of the two goods as well as λ, and leads to the key comparative static tested in Ericson and Fuster (2011) and Heffetz and List (2014): as the permission-to-exchange increases, loss-averse individuals are predicted to become more willing to attempt an exchange. Importantly, the comparative static is *reversed* for gain-lovers – these types should grow less willing to attempt an exchange as the permission-to-exchange increases.[1]

The probabilistic forced exchange paradigm (Cerulli-Harms et al., 2019) provides an analogous set of considerations for participants; whereas the permission-to-exchange paradigm acts on those intending to exchange (individuals wanting to retain their endowment are unaffected), probabilistic forced exchange instead acts on individuals intending to keep their endowment – in other words, those individuals who generate the endowment effect. Participants in these experiments are told that they may choose to hold onto their endowment or to exchange

[1]This underlying heterogeneity of treatment effects, discussed in great length by Goette, Graeber, et al. (2018), is an important confounder of average treatment effects in these studies, which were conducted under an assumption of universal loss aversion.

for the alternative, but regardless of their choice, they will be forced to relinquish their good for the alternative with probability t. Thus, for a fixed t, an individual endowed X would prefer to exchange in CPE so long as

$$\frac{m_Y}{m_X} \geq \frac{1+t(1-\lambda)}{1-t(1-\lambda)}.$$

Once again, manipulating t yields comparative static predictions for KR CPE – namely that loss-averse individuals should grow more willing to exchange with t while gain-loving individuals should grow more hesitant to exchange.[2]

2.2.2 Logit Choice Formulation

Having considered the KR comparative statics explored in prior work, we next describe our mapping from the KR CPE utility considerations to observed choice data using standard methods from discrete choice (Train, 2009). In particular, we assume that individuals face Type I Extreme Value error, ε, so that the CPE utility of a lottery F is perceived as $\tilde{U}(F|F) = U(F|F) + \varepsilon$. An individual would then prefer the lottery F to the lottery F' in CPE as long as

$$U(F|F) - U(F'|F') \geq \nu,$$

where $\nu = \varepsilon' - \varepsilon \sim Logistic(0,1)$.

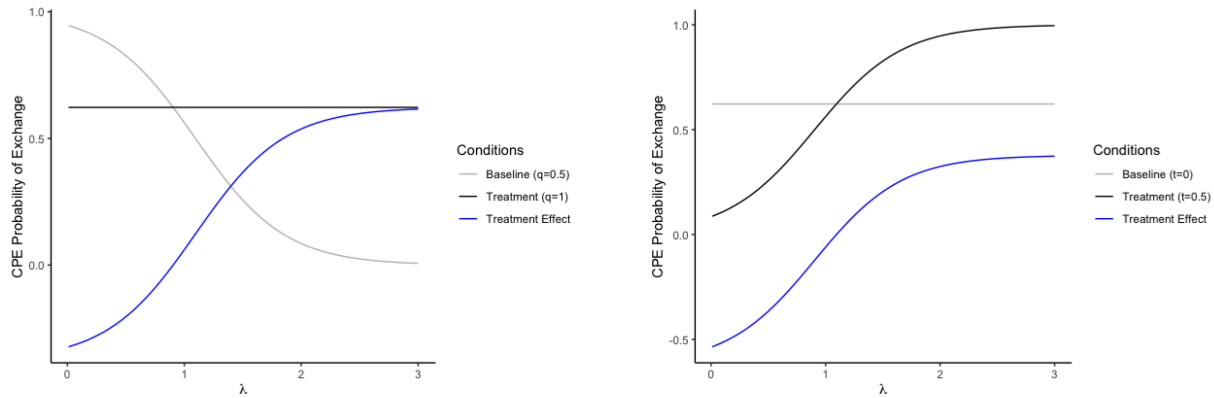
Under this set of assumptions, we can write the probability that an individual prefers F more succinctly as

$$p = logit^{-1}[U(F|F) - U(F'|F')],$$

where $logit^{-1}(x) = \frac{1}{1+\exp(-x)}$, relying on the CDF of the logistic distribution.

Plugging the permission-to-exchange CPE utility in the discrete choice model, the prob-

[2] KR comes equipped with several alternative equilibrium notions, notably Personal Equilibrium and Preferred Personal Equilibrium. For an analysis of the robustness of these comparative statics across equilibrium notions, see Goette, Graeber, et al. (2018).



(a) Permission Design (b) Forced Exchange Design

Figure 2.1: **Logit Choice Probability of Exchange**
Notes: Panel A presents CPE treatment effects in exchange probability in the permission-to-exchange design, while Panel B displays the probabilistic forced exchange design.

ability that an agent endowed X with permission q voluntarily exchanges their good is then

$$p = logit^{-1}[U(q(0,Y)+(1-q)(X,0)|q(0,Y)+(1-q)(X,0)) - U(X,0|X,0)]$$

$$p = logit^{-1}[qm_Y + (1-q)m_X + q(1-q)\eta(1-\lambda)(m_X+m_Y) - m_X].$$

In the forced exchange paradigm, the relevant probability of voluntary exchange at forced exchange probability t is given by

$$p = logit^{-1}[U(0,Y|0,Y) - U(t(0,Y)+(1-t)(X,0)|t(0,Y)+(1-t)(X,0))]$$

$$p = logit^{-1}[m_Y - (tm_Y + (1-t)m_X + t(1-t)(1-\lambda)(m_X+m_Y))].$$

We plot this object over values of $q \in \{0.5, 1\}$ and $t \in \{0, 0.5\}$ in Figure 2.1, clearly showing that exchange probabilities and treatment effects across conditions theoretically depend upon the value of λ. As such, under a heterogeneous distribution of λ, the average treatment effect should depend on more than just the first moment of the distribution of gain-loss attitudes. Indeed it is simple to construct examples with loss aversion on average and null effects from the manipulations of these experiments.

2.2.3 Estimator Overview

Under the assumption that individuals behave according to KR CPE and gain-loss attitudes are heterogeneous, observed exchange probabilities are informative about both the average gain-loss attitudes and their dispersion. From the Bayesian perspective we adopt in this manuscript, what is inferred from any set of choices depends on the prior that is adopted for the distribution of gain-loss attitudes – discussed in section 2.3.1. Our Bayesian estimator takes observed choice data and a prior distribution over gain-loss preferences and provides a best-fitting posterior distribution given the data.

The estimator hews closely to the logit choice set up described in section 2.2.2; we model observed exchange choice as stemming from a Binomial(p) process, with the probability of exchange defined as the CPE utils described therein. Given priors over the intrinsic utils and the gain-loss distribution, we sample from the posterior distribution of our relevant parameters resulting from the typical Bayesian update after observing the data. We rely on *rstan* to efficiently draw samples from the posterior distribution, the code for which can be found in Appendix B.1 (Stan Development Team, 2020).

2.3 Estimator Evaluation

Before deploying our estimator, we follow a series of critical steps in the Bayesian workflow – understanding the implications of our prior assumptions and exploring the recovery properties of our estimator (Schad et al., 2021). This section therefore focuses on a systematic approach to verifying that our estimator is reliable, particularly when it comes to recovering the distribution of λ from a given dataset. We emphasize the permission-to-exchange paradigm for the various examples throughout this exercise, noting here that the implications for probabilistic forced exchange are analogous.[3]

[3] More information on these exercises can be found in Appendix B.2.

2.3.1 Prior Beliefs

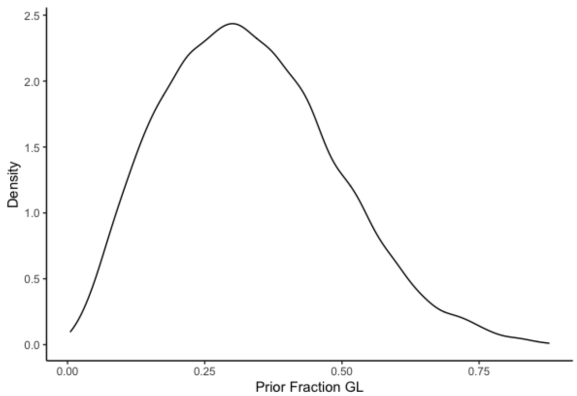
An important consideration in all Bayesian estimators is the prior, especially given how important it can be with small sample sizes. Because beliefs about the distribution of gain-loss preferences have shifted so dramatically following the work Chapman, Snowberg, et al. (2018), we specify two sets of priors – *high-GL* and *low-GL*. We introduce an additional assumption, limiting our search for the optimal distribution to the family of lognormals (i.e., $\lambda \sim LogNormal(\bar{\lambda}, \sigma_{\lambda}^2)$).[4]

Our *high-GL* prior relies on setting hyper-priors of $\bar{\lambda} \sim \mathcal{N}(0.25, 0.25^2)$ and $\log(\sigma_{\lambda}) \sim \mathcal{N}(-0.65, 0.15^2)$.[5] The density plot in Figure 2.2a depicts the implied fraction of gain-lovers associated with these draws from our priors; it suggests that the implied fraction of the population with $\lambda_i < 1$ has most of its mass between 10% and 50%, peaking around 30%. Clearly, this prior assumes substantial heterogeneity in gain-loss attitudes. Indeed, we plot the distribution of the LogNormal at its mean parameter values in Figure 2.2b to illustrate what a candidate population of λ looks like beyond the summary statistic of fraction gain loving; this distribution has a median (mean) of 1.29 (1.48).

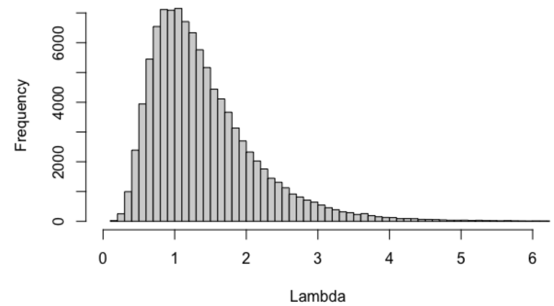
For our *low-GL* prior, we adjust the hyper-priors of the LogNormal distribution to generate gain-loss preferences that are more in line with the traditional view that gain-loving individuals are quite rare. This adjustment is made by shifting the distribution of the location, $\bar{\lambda} \sim \mathcal{N}(0.75, 0.25^2)$, as well as the scale, $\log(\sigma_{\lambda}) \sim \mathcal{N}(-0.8, 0.15^2)$. The resulting distributional simulations are shown in Figure 2.2c, indicating the effect of our shifting hyper-parameters. The fraction of gain-lovers in the population has its peak at around 2%, and the vast majority of the mass lies below 10%. At the mean values of these hyper-priors, the distribution of gain-loss

[4]We adopt the LogNormal following Goette, Graeber, et al. (2018); this is a natural candidate as λ is assumed non-negative, though certainly other distributions could apply.

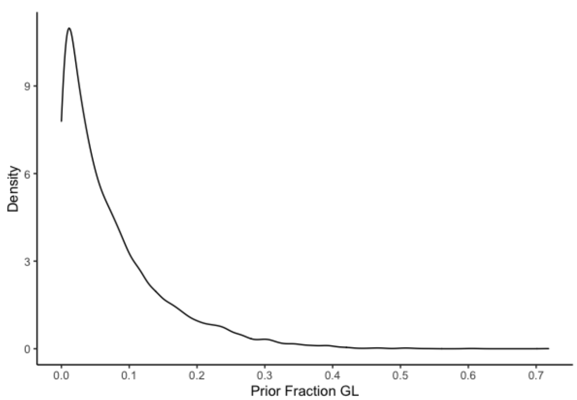
[5]One reason to use the normal, from McElreath (2020), is that it is the best distribution to use (in terms of Entropy maximization) when you know only that a variable has a mean and a scale – "If all we are willing to assume about a collection of measurements is that they have a finite variance, then the Gaussian distribution represents the most conservative probability distribution to assign to those measurements" (pg 306).



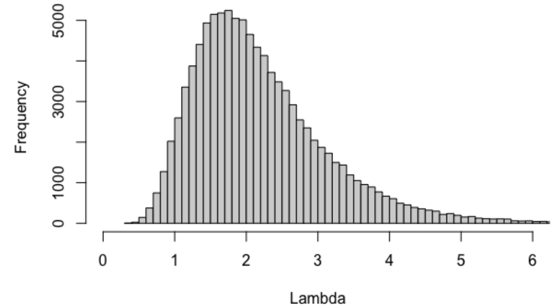
(a) *High-GL* Prior Fraction Gain-Loving



(b) *High-GL* Prior Draws of λ_i



(c) *Low-GL* Prior Fraction Gain-Loving



(d) *Low-GL* Prior Draws of λ_i

Figure 2.2: **Prior Draws of Gain-Loss Parameters**

Notes: The draws are based on our priors of $\lambda \sim LogNormal(\bar{\lambda}, \sigma_\lambda^2)$, where *High-GL* refers to $\bar{\lambda} \sim \mathcal{N}(0.25, 0.25^2)$ and $\log \sigma_\lambda \sim \mathcal{N}(-0.65, 0.15^2)$ and *Low-GL* refers to $\bar{\lambda} \sim \mathcal{N}(0.75, 0.25^2)$ and $\log \sigma_\lambda \sim \mathcal{N}(-0.85, 0.15^2)$.

preferences is represented in Figure 2.2d, with median (mean) λ_i of 2.08 (2.27).

Besides the implied fraction of gain-lovers, the priors we specify for our estimator also have regularizing properties; the main consideration involves an exploration of the implied probabilities of exchanging given a randomized endowment and permission-to-exchange. As shown in Appendix B.2, both the *high-GL* and *low-GL* priors make reasonable predictions about the probabilities of exchange.

2.3.2 Synthetic Recovery

The next key step in the Bayesian workflow is to validate our estimator's ability to recover the central parameters. Here, we focus on the recovery of the population distribution of λ by conducting a 2x2 synthetic recovery exercise. The aim is to demonstrate our ability to recover the true data generating process (DGP) parameters $\left(\bar{\lambda}, \log(\sigma_\lambda)\right)$ when we begin with different underlying DGPs (either *high-GL* or *low-GL*) and different priors (*high-GL* or *low-GL*). Recovering the true distribution – regardless of the similarity between DGP and prior – allows us to confidently apply our model to the existing body of experimental data and update our belief about λ.

For this section, we focus on the convergence properties under a single dataset with one choice per individual as the number of individuals in the dataset increases. We focus on the *high-GL* DGP in the main text, and leave the details of our synthetic DGP, as well as the discussion of the *low-GL* DGP and the probabilistic forced exchange paradigm, to Appendix B.2.

High-GL Prior

The first model we use to estimate the population distribution relies on our *high-GL* prior described in section 2.3.1. The priors are not perfectly in line with the DGP in this setting, but they have qualitatively similar implied fractions of gain-lovers. The posterior draws of $\bar{\lambda}$ and $\log(\sigma_\lambda)$ are plotted (with 95% credible intervals) in blue alongside the implied prior 95%

bands (from our Normal distributions on the hyper-priors) in Figure 2.3. The results show that, even with one choice per individual, our posterior estimates are quite accurate, with uncertainty greatly reduced relative to our prior – growing much tighter with N. Interestingly, our uncertainty regarding $\log(\sigma_\lambda)$ is only slightly reduced relative to the prior.

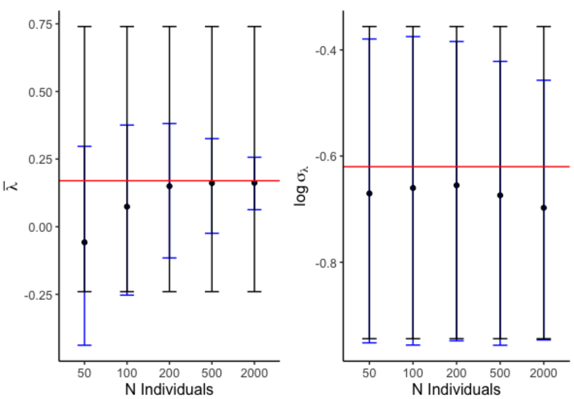
The convergence plots suggest our estimator recovers the true population distribution robustly over sample size. Given this recovery, we can take our posterior estimates of the hyper-parameters and compute the posterior (versus prior) fraction of gain-lovers. In the example shown in Figure 2.3, where we look specifically at $N = 2000$, the two distributions overlap quite a bit, though our posterior is much more certain that the fraction is between about 25% and 40%. That is, while the prior fraction of gain-lovers has substantial weight below 25% and above 50%, the posterior seems to rule those as quite unlikely for this synthetic dataset.

Low-GL Prior

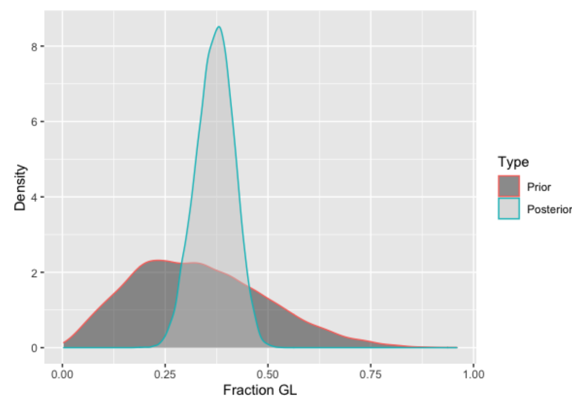
Next, we supply the estimator with our *low-GL* prior under a true *high-GL* DGP. This recovery exercise sheds light on whether our estimator is able to recover the population distribution when the prior is misspecified relative to the data. The results illustrate that recovery of $\bar{\lambda}$ is near perfect with this misspecified prior, consistently reaching values outside of its prior 95% confidence interval. The posterior for $\log(\sigma_\lambda)$ contains the true value in its 95% credible interval, but uncertainty is hardly reduced and the posterior mean does not move much towards the true value as N increases.

Nonetheless, turning our attention to the posterior versus prior fraction of gain-lovers, we see that our estimator shifts dramatically away from the misspecified prior and towards the truth. While the prior placed almost no mass above 20% of gain-lovers in the population, the posterior places the majority of the mass on the fraction being between 20% and 35%, which is approaching the true fraction (about 37%).

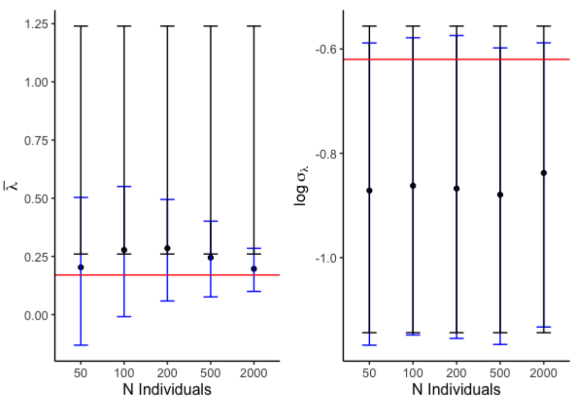
Overall, this synthetic recovery exercise demonstrates that our Bayesian structural esti-



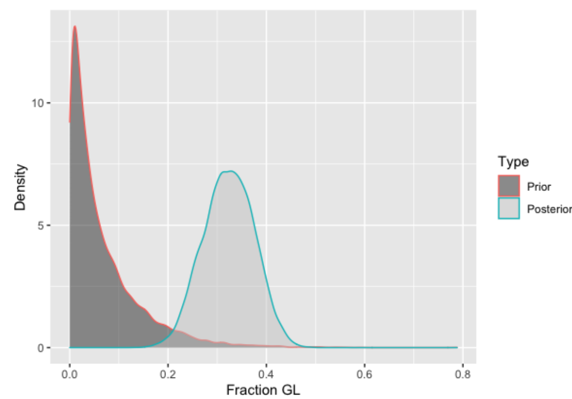
(a) *High-GL* 95% Credible Intervals



(b) *High-GL* Fraction Gain-Lovers



(c) *Low-GL* 95% Credible Intervals



(d) *Low-GL* Fraction Gain-Lovers

Figure 2.3: **Recovery Results by Prior: High-GL DGP**

Notes: Panel A contains Posterior 95% credible intervals (blue) for $\bar{\lambda}$ and $\log(\sigma_\lambda)$ overlaid upon the implied 95% confidence interval based on our Normal hyper-priors of $\bar{\lambda} \sim \mathcal{N}(0.25, 0.25^2)$ and $\log \sigma_\lambda \sim \mathcal{N}(-0.65, 0.15^2)$ (black). Panel B contains the traction of gain-lovers in the population, evaluated by computing the LogNormal CDF, $LogNormal(\bar{\lambda}, \sigma_\lambda^2)$ at a value of 1 under either prior or posterior draws of $\bar{\lambda}$ and $\log(\sigma_\lambda)$. Panel C contains the Posterior 95% credible intervals (blue) for $\bar{\lambda}$ and $\log(\sigma_\lambda)$ overlaid upon the implied 95% confidence interval based on our Normal hyper-priors of $\bar{\lambda} \sim \mathcal{N}(0.75, 0.25^2)$ and $\log \sigma_\lambda \sim \mathcal{N}(-0.85, 0.15^2)$ (black). Panel D contains the fraction of gain-lovers in the population, evaluated by computing the LogNormal CDF, $LogNormal(\bar{\lambda}, \sigma_\lambda^2)$ at a value of 1 under either prior or posterior draws of $\bar{\lambda}$ and $\log(\sigma_\lambda)$.

mator is robustly capable of recovering the latent gain-loss preference distribution, even with small N, 1 choice per person, and misspecified priors. Comparing the posteriors under the two priors emphasizes that misspecified priors can affect the recovered fraction of gain-lovers through small bias in the scale term, but nevertheless converge in the correct direction.

2.3.3 MCMC: Bias and Coverage

Our synthetic analysis of the estimator has so far focused on a single dataset for a given number of individuals; in this section, we introduce a Monte Carlo exercise over 500 datasets, intended to speak to the coverage and bias properties of our estimator. From each dataset, we store the posterior mean and 95% credible intervals for $\bar{\lambda}$ and $\log(\sigma_\lambda)$ for sample sizes of $N \in \{100, 2000\}$ to better explore the empirical properties and their dependence on N.

The results are presented in Table 2.1 below. Regardless of whether the prior is misspecified relative to the DGP, our estimator has almost no bias in estimating $\bar{\lambda}$ with large sample sizes; moreover, the coverage on the 95% credible intervals converges towards the nominal 95%, ensuring our estimator consistently recovers $\bar{\lambda}$ over the 500 datasets. For $\log(\sigma_\lambda)$, our bands are too conservative, with 100% coverage in all cases. However, the implications of $\log(\sigma_\lambda)$ on the fraction of gain-lovers are fairly negligible, and are certainly less important than the $\bar{\lambda}$ hyperparameter (especially given the magnitudes of the bias observed).[6]

2.4 Existing Experimental Data

Having established our estimator's ability to recover the ground truth, we next discuss the structure of the existing experimental data from Cerulli-Harms et al. (2019), Ericson and Fuster

[6] In one example, where the true DGP $\log(\sigma_\lambda) = -0.62$ but it is estimated at -0.85, there is about a 1 pp difference in the fraction of gain-lovers; in terms of other summary statistics, the medians are essentially identical under this error, but the more negative -0.85 leads to an (expected) compression in the distribution, with a smaller IQR (0.89 - 1.58 for the incorrect versus 0.83 - 1.71 for the true).

Table 2.1: **Monte Carlo, High-GL Prior**

		Coverage, Mean	Bias, Mean	Coverage, SD	Bias, SD
High-GL DGP					
	N = 100	99%	0.022	100%	-0.0365
	N = 2000	96.6%	0.008	100%	-0.0756
Low-GL DGP					
	N = 100	89.8%	-0.139	100%	0.135
	N = 2000	94.2%	-0.0149	100%	-0.0153

Notes: Monte Carlo simulations were conducted over 500 datasets for each N. Posterior samples for each dataset were drawn from *rstan*, from which Bias was computed by comparing the posterior means to the true DGP values, and coverage was computed by determining the fraction of datasets for which the 95% credible intervals on the parameters $\bar{\lambda}$ and $log(\sigma_\lambda)$ contained the true value. Samples are drawn from our *high-GL* model, and data is simulated from both our *high-GL* and *low-GL* DGP.

(2011), and Heffetz and List (2014) and how it shapes our analysis.[7] All of these experiments were designed to test the same question – does exchange behavior change in the KR predicted direction under exogenous shocks to the reference point?

Ericson and Fuster (2011) and Heffetz and List (2014) approach this by analyzing exchange behavior as they experimentally vary the permission-to-exchange probability from $q = 0.1$ to $q = 0.9$. Cerulli-Harms et al. (2019) develop the probabilistic forced exchange paradigm, analyzing behavior as they vary the forced exchange probability over $t \in \{0, 0.25, 0.5, 0.75\}$. As discussed in section 2.2, the hypothesized sign on the treatment effect varies by gain-loss type; however, these papers implicitly assume universal loss aversion because heterogeneity in gain-loss preferences was not widely recognized at the time.

Participants in Ericson and Fuster (2011) and Heffetz and List (2014) only make one choice – attempt to exchange for the alternative good or keep their endowment – given their randomly assigned endowment and permission probability. Thus, it is impossible to accurately identify an individual's λ_i. Nevertheless, as demonstrated in section 2.3.3, the experimental

[7] We do not to use the replication of Ericson and Fuster (2011) from C. F. Camerer et al. (2016) as they focus on a different protocol than the pure exchange design we've described. We also omit data from several sessions of these papers to avoid estimating distributions when forced exchange or permission-to-exchange are extreme (i.e. $t = 0.99; q = 0.01$. In these cases, choices are more indicative of intrinsic utils than gain-loss preferences. This does not have a noticeable effect on our posterior distributions, however.

variation in permission-to-exchange and endowment allows us to identify a *distribution* of λ that best fits the observed data given that individuals behave according to KR CPE. In addition, because individuals select between two goods, we do not directly observe the intrinsic utilities, and thus need to estimate these utilities from the data.

The data from Cerulli-Harms et al. (2019) is quite different from that in Ericson and Fuster (2011) and Heffetz and List (2014) in that (1) it relies on probabilistic forced exchange for identification, yielding distinct CPE equations; (2) participants trade off a mug for money, so under the standard assumption of linear utility, we know the utils from one of the goods; and (3) choices are elicited in Multiple Price Lists, wherein sellers (buyers) of the mug state their Willingness-to-Accept (Willingness-to-Pay) by indicating whether they prefer the mug or \$0.50 to \$10 (in \$0.50 increments). While this richer data provides an avenue through which to estimate individuals' λ_i, this requires a fundamentally different estimator and will not be explored in this paper.

In order to provide a comparison using a similar number of subject choices to that of Ericson and Fuster (2011) and Heffetz and List (2014), we extract several choices from these price lists.[8] We select prices around the median WTA and WTP for each endowment, returning us to our familiar logit choice setup, and deploy a similar estimation strategy at the choice level.[9]

The data from each of the three experiments we consider are presented in Table 2.2. Under the assumption of universal loss aversion, the KR comparative statics derived in section 2.2 imply that exchange should increase with probability (regardless of design). However, across

[8]The vast majority of rows will not contain any information about the gain-loss preferences. For example, suppose that we applied our estimator only on the row of the data when participants compare the mug to \$1; regardless of the forced exchange probabilities and endowments, individuals would (and do) almost universally prefer the mug to \$1. Because our estimator is identified off of the variation in exchange probabilities, these edge prices where there is no treatment effect would imply either highly gain-loving or highly loss averse population distributions (depending on the endowment). As such, we would not learn anything about the gain-loss distribution by including these prices.

[9]By extracting the median price (+/- 0.5) for buyers and sellers under a forced exchange probability of 0.5, we ensure variation in exchange behavior since we know by construction that half of the population (under forced exchange of 0.5) will prefer the money at the chosen price. Note that when we take multiple rows, we treat them as independent.

Table 2.2: Experimental Data: Proportion Exchange

		Probability					
		0%	10%	25%	50%	75%	90%
	Endowment		*Permission-to-Exchange Design*				
Ericson and Fuster (2011)							
	Mug		22.7% (5/22)				56.5% (13/23)
							33.8% (+)
Heffetz and List (2014)							
Exp 1	Mug		36.0% (9/25)				18.8% (6/32)
Exp 1	Pen		52.9% (18/34)				50.0% (13/26)
Subtotal							-13.0% (-)
Exp 2	Mug		32.3% (10/31)				37.5% (12/32)
Exp 2	Pen		66.7% (18/27)				80.8% (21/26)
Subtotal							8.6% (+)
Exp 3	Mug		38.6% (17/44)				30.4% (17/56)
Exp 3	Pen		53.1% (34/64)				54.1% (33/61)
Subtotal							-4.5% (-)
	Endowment		*Probabilistic Forced Exchange Design*				
Cerulli-Harms et al. (2019)							
Exp 1	Mug	36.7% (22/60)		46.7% (28/60)	34.5% (19/55)	50.0% (27/54)	
Exp 1	Money	43.3% (26/60)		46.7% (28/60)	48.3% (29/60)	48.2% (27/56)	
Subtotal				6.7% (+)	1.7% (+)	9.1% (+)	
Exp 2	Mug			57.5% (23/40)	52.5% (21/40)	44.7% (17/38)	
Exp 2	Money			62.9% (22/35)	48.7% (19/39)	30.6% (11/36)	
Subtotal					-9.4% (-)	-22.2% (-)	
Exp 3	Mug			44.8% (13/29)	42.3% (11/26)	42.9% (12/28)	
Exp 3	Money			37.0% (10/27)	55.2% (16/29)	37.9% (11/29)	
Subtotal					8.0% (+)	-0.72% (-)	

Notes: Data from Heffetz and List (2014) is broken down from 3 experiments, described in Table 2 and Table 3 in their paper. Exp 1 refers to the **More Endowment** condition of their *Experiment 2*, Exp 2 refers to the **Less Endowment** condition of their *Experiment 2*, and Exp 3 refers to *Experiment 3*. For the data in Cerulli-Harms et al. (2019), the subsession follow from the definitions in their data. Importantly, the Cerulli-Harms et al. (2019) data presented in this table considers the exchange behavior implied by comparing the WTA or WTP to the median price (per session and endowment) under a forced exchange of 50%. Aggregate differences in exchange by condition are presented below each session, always relative to the lowest probability for that session. $(+)$ or $(-)$ indicate whether the observed exchange behavior is in line with KR comparative statics under the assumption of universal loss aversion.

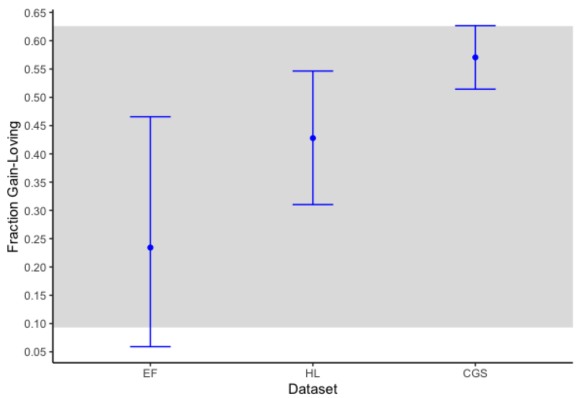
all experiments, no consistent pattern emerges; 6 of the 11 experimental datasets have a sign consistent with KR, and the remaining 5 report treatment effects of the opposite sign. While some interpret these inconsistencies as a shortcoming of KR, in the next section, we instead ask what draws from the distribution of gain-loss preferences could lead to these varying results.

2.5 Application to Existing Data

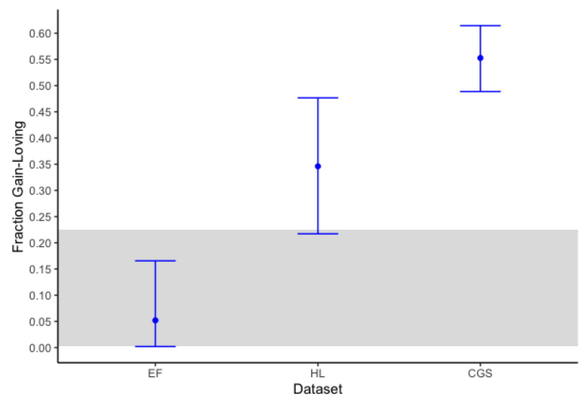
We turn our attention to estimating the distribution of gain-loss preferences from the body of existing experimental studies. Under both the *high-GL* and *low-GL* priors, we analyze the posterior 95% credible interval for the fraction of the population estimated as gain-loving. We present our results in two related ways, estimating the posterior intervals for each dataset separately, as well as under a sequential approach where we add each new experimental dataset to the estimator.

First, we consider the data from Ericson and Fuster (2011), where the relatively small sample size yields a posterior that is closely tied to the prior. To the extent that the data is able to move the prior, the large and positive treatment effects observed in this data would align with a loss-averse population per KR CPE. These results are clearly demonstrated in Figure 2.4, where the posterior mean for the fraction of gain-lovers is highly sensitive to the prior in the EF dataset, and the small movements shift the posterior closer to 0.

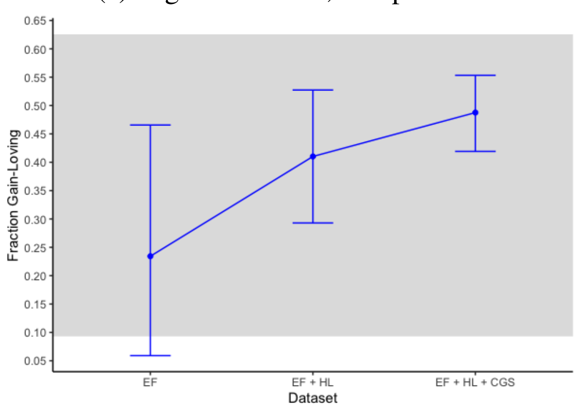
Next, we apply our estimator on three sessions run by Heffetz and List (2014). Because each session has slight differences, including small nudges in the instructions to strengthen the sense of possession as well as the transparency of the randomization procedures, we allow for a separate estimate of the intrinsic utility for each session (but a common distribution of gain-loss preferences). Posterior estimates illustrated in the HL column of Figure 2.4 suggest substantial heterogeneity in gain-loss attitudes, with a mean fraction of gain-lovers between 34% and 42% depending on the prior, and 95% credible intervals indicating a likely range of 20% to 54% of



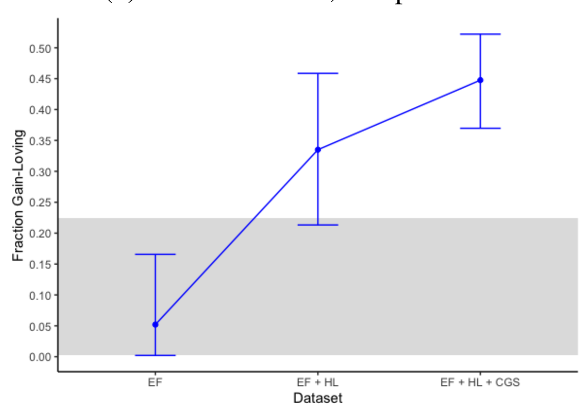
(a) High-GL Model, Independent



(b) Low-GL Model, Independent



(c) High-GL Model, Sequential



(d) Low-GL Model, Sequential

Figure 2.4: **Posterior Results**

Notes: Panels A and B contain the posterior and prior results for the fraction of gain-lovers in the data when analyzing each dataset separately. Panels C and D instead add new data sequentially. The data from Cerulli-Harms et al. (2019) in panels (c) and (d) only consists of one choice per participant, namely their decision at the median WTA (WTP) to avoid overweighting that data relative to the other studies.

the experimental participants as gain-loving under KR CPE.

Last among the independent data sets comes Cerulli-Harms et al. (2019), which also consists of three experiments with slightly different protocols. To flexibly estimate the distribution of gain-loss preferences across these sessions, we allow for distinct valuations of intrinsic utils per session *and* endowment.[10] As seen in Figure 2.4b, the entire 95% posterior credible interval for CGS lies far away from the 95% confidence interval implied by our prior. Under either the *high-GL* or *low-GL* prior, our estimator rationalizes the observed choice data with a distribution of gain-loss preferences consisting of roughly 50% to 62% gain-lovers.

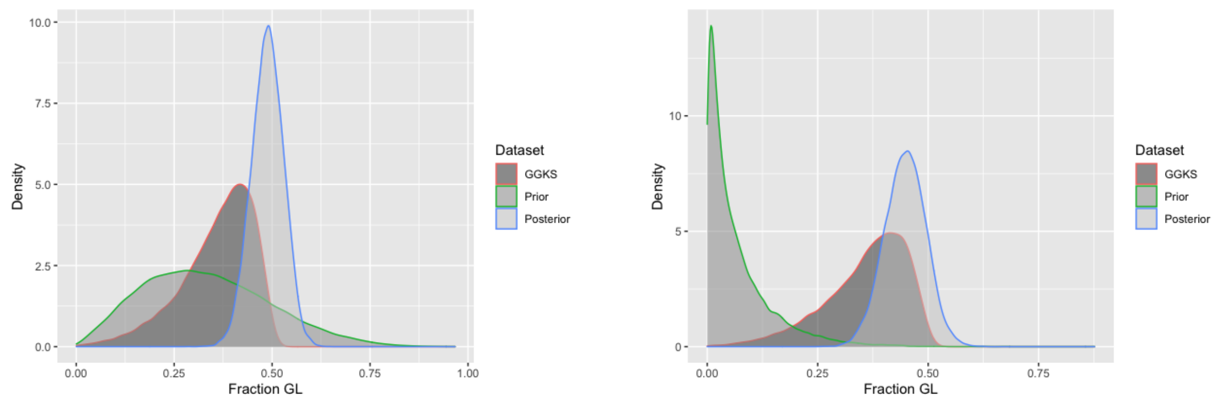
Figure 2.4 also displays the results when applying our estimator to the chronological accumulation of data.[11] When we combine the data from the permission-to-exchange experiments to form our posterior on EF+HL, the result is quite similar to HL alone given the relative sample sizes. Taken together, this implies that regardless of the prior, the body of work in the permission-to-exchange paradigm is best rationalized under KR CPE by a distribution of gain-loss preferences containing between 20% to 50% gain-lovers. Incorporating the evidence gathered under the probabilistic forced exchange paradigm (EF + HL + CGS), the fraction of gain-lovers implied by the data under CPE shifts up to 35% to 55%.[12]

Altogether, under the assumption that gain-loss attitudes drive the observed variation in exchange behavior from existing studies, the distribution that best rationalizes the body of evidence consists of around 45% gain lovers. This estimate is robust to prior beliefs over the likely distribution of gain-loss preferences, as the pool of data is large enough to reduce our reliance on the prior. As shown in Figure 2.5, even when our prior places almost no weight on

[10]Allowing for the intrinsic utility of the mug by session and endowment helps our estimator overcome the substantial endowment effect across all conditions – where WTA is everywhere larger than WTP. In so doing, the intrinsic utility parameter helps fit the observed level of exchange within endowment instead of forcing λ to rationalize the broad differences in valuation. Note that removing this degree of freedom would result in a more gain-loving posterior.

[11]Naturally, the results for EF are the same since it comes first.

[12]The 95% credible interval for EF + HL + CGS in the sequential analysis is wider than the 95% credible interval for CGS because we only use the median price (one choice per person) rather than the median + 50 cents. Because the data from CGS only are about twice the size as that of EF + HL, EF + HL + CGS is still less than half the size of the 3 rows from CGS only.



(a) High-GL Model (b) Low-GL Model

Figure 2.5: **Proportion of Gain-Lovers**
Notes: The distribution for the proportion of gain-lovers from Goette, Graeber, et al. (2018) (GGKS) is computed following their bootstrapped approach, simulating values for $\mu_\lambda, \sigma^2_\lambda$ from the joint normal distribution implied by the estimates in their Table 2, columns (1) and (2).

more than 10% of the population as gain-loving, the resulting posterior peaks close to 45% with little uncertainty. The bootstrapped proportion of gain-lovers from Goette, Graeber, et al. (2018) is plotted alongside, demonstrating that while our posteriors may be slightly more gain-loving, these results are highly consistent with an independent study in this same context of exchange experiments. Similarly, Chapman, Snowberg, et al. (2018) estimate 53% of the U.S. population to be gain-loving in their survey, falling well within our posterior estimates.[13]

2.6 Conclusion

On the surface, the mixed evidence for expectations-based reference dependence (KR) from prior experimental studies makes it difficult to draw conclusions about the model's ability to describe behavior in the lab. However, this interpretation relies on an often ignored assumption that agents are universally loss-averse. In light of recent evidence to the contrary (Chapman, Snowberg, et al., 2018; Goette, Graeber, et al., 2018), this paper reexamines the data from these

[13]Interestingly, when implementing their methodology on lab participants, Chapman, Snowberg, et al. (2018) find closer to 10% of the population as gain-loving, smaller than their reported weighted average across 8 existing studies (22%).

experiments with an eye towards understanding the degree of heterogeneity in gain-loss attitudes implied by the results.

Under several assumptions, most critically the logit choice framework and that individuals behave according to KR CPE, we demonstrate that regardless of prior beliefs over the fraction of gain-lovers, the data pushes our optimal posterior towards a substantial 35% to 55%. These findings are in line with an increasingly prevalent set of papers estimating roughly 38% to 53% of participants as gain-loving (Chapman, Snowberg, et al., 2018; Goette, Graeber, et al., 2018). To verify these results, we carefully proceed along an established Bayesian workflow to ensure our estimator is capable of discerning the true population distribution of gain-loss preferences in a synthetic recovery exercise (Schad et al., 2021). Moreover, our Monte Carlo simulations demonstrate remarkable coverage and bias properties given the limitations of the data.

One interesting question not discussed in this paper is what a similar exercise would uncover about the distribution of gain-loss preferences in other domains – namely effort provision (Abeler et al., 2011; Gill and Prowse, 2012; Gneezy et al., 2017) and monetary lotteries (Sprenger, 2015). Surprisingly little is known about the relationship between gain-loss distributions and context, both on an aggregate and individual level.

Chapter 3

Reference-Dependent Effort Provision under Heterogeneous Gain-Loss Preferences

3.1 Introduction

A central tenet of behavioral economics, models of reference-dependent preferences have been widely adopted to explain anomalous results under Expected Utility theory (EUT) (Kahneman and Tversky, 1979; Rabin, 2000). The core intuition of these models is that, rather than evaluating outcomes over final wealth, individuals consider how prospective outcomes will position them relative to some reference point – often taken to be the status quo, a goal, or an expectation. In addition to a reference point, models traditionally posit asymmetric treatment of gains and losses relative to the reference point. Losses having greater utility consequences than commensurate gains has been termed 'loss aversion'.

One prominent domain of application for reference-dependent preferences is labor supply. A reference-dependent individual making labor market decisions at the daily level may exhibit anomalous labor supply decisions, working less when wages are high and the reference point is easily reached than when wages are low. A leading example comes from the seminal work by C. Camerer et al. (1997), who document evidence of a negative wage elasticity among New York City cab drivers, consistent with a daily income target.[1]

A common criticism of early formulations of reference-dependent preferences is that the reference point itself is unspecified. When applying the model in different environments, researchers may make use of this degree of freedom to rationalize behavior. While this may aid the model's explanatory ability, lack of discipline in choosing the reference point limits the value of the model for making predictions. In response to this critique, recent theoretical contributions have focused on closing the model by endogenizing the reference point, most prominently in the Kőszegi and Rabin (2006, 2007) (KR) model of expectations-based reference dependence. Within expectations-based reference-dependent models, the referent is a distribution of expected outcomes. As in other areas of economics, requiring expectations to be rational disciplines what agents can expect and hence what reference-dependent behaviors are predicted.[2]

The KR development spurred new research into the relationship between this expectations-based mechanism and the documented labor supply findings. Crawford and Meng (2011), for instance, advanced the cab driver literature through specifying target earnings as a function of recent expectations (i.e., average daily earnings on prior days) as well as introducing a reference point over hours worked. Thakral and To (Forthcoming) analyze how the timing of earnings on a given day interacts with the probability of stopping a shift, developing a model where the reference point adjusts throughout the day so that recent hours receive more weight in stopping decisions.

[1]Though there was some debate about the mechanism behind these results (Farber, 2005), the reference-dependent interpretation was bolstered by Fehr and Goette (2007), who provide experimental variation in the wages of bicycle messengers and link a measure of loss aversion to the observed reduction in shift earnings.

[2]KR provide several different rational expectations constructions. The most prominent of these is Choice-Acclimating Personal Equilibrium, which we assume in this manuscript. Recent work has noted a tight connection between this work and models of rank-dependent utility (Masatlioglu and Raymond, 2016). Intuitively, this connection derives from the fact that changing the ranking of two outcomes changes gain-loss comparisons in the KR model and also changes rank-dependent probability distortions in rank-dependent models.

Beyond the study of cab drivers, a number of laboratory experiments have been designed to test KR comparative static predictions in the real effort domain. Abeler et al. (2011) manipulate expected earnings by offering the lottery $(0.5, F; 0.5, we)$ with $F \in \{3, 7\}$ depending on the condition – that is, a 50% chance that payment is a piece rate, w, for each completed task, e, and a 50% chance of receiving F regardless of effort; Gneezy et al. (2017) expand the set of contracts by offering the lottery $(p, F; 0.5 - p, 0; 0.5, we)$ under various F and p. This design is based on the KR comparative static that loss-averse individuals will exert more effort as F or p increase – holding the lottery as a referent, the marginal benefit of effort is higher when the considered piece-rate earnings, we, fall below the fixed F, as it introduces the possibility of losses. While Abeler et al. (2011) find evidence supporting the KR hypothesis under loss aversion, the results in Gneezy et al. (2017) are inconsistent with KR, and exhibit no consistent pattern. In a far-reaching replication study focusing on experimental economics, C. F. Camerer et al. (2016) provide a confidence interval that contains the point estimate from Abeler et al. (2011), but that is not statistically significant.[3]

While the empirical evidence accumulated from observing the behavior of cab drivers is encouraging, the mixed body of experimental results casts doubt on the predictive validity of the KR model. However, a series of recent papers has highlighted a potential confound in the experimental design: heterogeneity of gain-loss preferences.[4] While loss aversion is a well documented and well understood phenomenon, work by Chapman, Snowberg, et al. (2018) reports that somewhere between 22% and 53% of individuals may instead be gain loving, weighting gains more than commensurately-sized losses. Importantly, the canonical experimental manipulations yield oppositely signed treatment effects when comparing gain-loving and loss-averse

[3]An additional experimental design in this domain is explored in Gill and Prowse (2012), who consider a sequential game in which the payoff is a function of how many more tasks a player competes relative to their opponent. The payoff is structured such that the second mover has a theoretical discouragement effect under KR (e.g., their effort should decrease with player 1's effort), which is exactly what they document in experimental data.

[4]Heterogeneity could be similarly relevant in field applications, especially because estimates of the population distribution of gain-loss attitudes is much more gain loving in representative samples of the U.S. relative to laboratory samples (Chapman, Snowberg, et al., 2018).

individuals. Moreover, aggregate treatment effects need not represent the average gain-loss preference.

Thus, individual gain-loss preferences must be taken into account to provide a true test of the KR comparative statics. In the domain of exchange behavior, Goette, Graeber, et al. (2018) provide such an experimental test, pre-pending the canonical KR endowment effect design of Cerulli-Harms et al. (2019), Ericson and Fuster (2011), and Heffetz and List (2014) with a first stage intended to capture gain-loss types. Their results demonstrate that null aggregate effects mask substantial heterogeneity of treatment effects over the gain-loss types.

In this manuscript we implement a two-stage experimental design to test the KR model in the real effort setting, accounting for heterogeneity in gain-loss attitudes. Our experiment is conducted on a sample of 265 subjects. Our first stage measure of gain-loss preferences is based on a series of 30 effort decisions elicited when offering standard wages (e.g., 20 cents per task) alongside mean-preserving spreads of these wages (e.g., 50% chance of 30 cents per task, 50% chance of 10 cents per task). Under standard wages, the marginal considerations for effort allow us to identify the cost-of-effort function under an assumed functional form; the stochastic wages then identify individuals' gain-loss preferences because the induced uncertainty introduces losses and gains relative to each potential outcome. We develop empirical methods to recover the distribution of gain-loss preferences for our entire sample, as well recovering the expected gain-loss parameter for each individual given their choices.

In a second stage, we explore behavior in two treatment conditions akin to Abeler et al. (2011) and Gneezy et al. (2017). Our conditions elicit effort when payments are structured as $(p,20;0.5-p,0;0.5,0.2e)$, so that payoffs are a lottery that may or may not depend on effort. Our baseline condition sets p to 5% so that individuals have a 50% chance of earning 20 cents per task, a 5% chance of \$20, and a 45% chance of \$0 (regardless of effort). For our treatment condition, p increases to 45%, meaning participants have a 50% chance of earning 20 cents per task, but now have a 45% chance of earning a fixed \$20 and a 5% of a fixed \$0. As we describe in section 3.2, KR predicts that loss-averse individuals increase their effort whereas gain-lovers decrease their effort in treatment relative to baseline. Intuitively, a higher chance of a fixed $20 payment positions participants below their reference point, to which loss-averse individuals respond by increasing effort to close the gap whereas gain-lovers respond by decreasing effort in anticipation of a positive surprise.

We document three key results in our sample of 265 participants. First, we measure a sizable minority (28%) of participants to be gain-loving in our first stage, providing evidence of heterogeneity in gain-loss preferences in the effort domain; the estimated distribution is similar to that reported in the exchange environment (Goette, Graeber, et al., 2018). Second, these estimated gain-loss preferences are predictive of baseline effort choices in our out-of-sample second stage, with increased loss aversion associated with significantly lower effort. Third, we find interaction effects between gain-loss preferences and our treatment condition that follow the theoretically predicted sign, but these are not statistically significant in all specifications.[5]

Our analysis provides evidence in support of the KR expectations-based mechanism. Alongside Goette, Graeber, et al. (2018), this suggests a new interpretation of the mixed results from the extant experimental tests of KR across exchange and real effort domains: there is stronger empirical foundation for the KR model of individual decision making than previously appreciated when heterogeneity in gain-loss attitudes are taken into account. Moreover, researchers considering applying reference-dependent models in the labor supply context now have an additional observation that lab participants behave broadly in line with KR. Finally, our documented distribution of gain-loss attitudes provides another piece of evidence for the existence and extent of heterogeneity, this time in the domain of real effort. Given the well-

[5] Our application of the two-stage paradigm from Goette, Graeber, et al. (2018) to the context of labor supply offers a number of critical advantages. A common critique of Goette, Graeber, et al. (2018) is that only 1-3 preference statements are elicited to measure gain-loss preferences, and the experimental variation is such that individual parameters of gain-loss preferences are not identified. We overcome this obstacle by measuring gain-loss preferences over 30 decisions at the individual level, thereby reducing the level of noise inherent to our gain-loss categorization and sharpening our test of KR. In addition, we fill a gap in the literature by eliciting individual decisions over monetary lotteries as an alternative measure of gain-loss preferences, allowing us to link these preferences across context.

established prevalence of gain-lovers across all major domains in which reference dependence is applied, future theoretical or empirical work in this area must recognize the fact that gain-loss attitudes are heterogeneous. Heterogeneity can have a significant impact on the number of observations required for adequately-powered empirical studies; average treatment effects aggregating over gain-loss types may differ both qualitatively and quantitatively from the treatment effect of the average preference and require remarkably large samples to recover.

The paper proceeds as follows. In Section 3.2, we outline the KR model, highlighting the comparative statics in the real effort domain. Section 3.3 describes our experimental protocol, section 3.4 discusses our first stage measurement and our heterogeneous treatment effects, followed by our conclusion in section 3.5

3.2 Theoretical Background

We begin by laying out the theoretical framework of Kőszegi and Rabin (2006, 2007) as applied to an individual's labor supply decision. An agent's utility consists of two components – consumption utility derived from earned wages and the (negative) cost of exerting effort, as well as psychological utility derived from comparing the realized wage and effort level to the agent's expectations. Formally, this is represented by

$$u(w,e|r_w,r_e) = m(we) - c(e) + \mu(m(we) - m(r_w e)) + \mu(c(e) - c(r_e)),$$

where, as is standard in practice, $\mu(\cdot)$ is assumed to be piece-wise linear,

$$\mu(z) = \begin{cases} \eta z & z \geq 0 \\ \eta\lambda z & z < 0. \end{cases}$$

The first component of utility, $m(we) - c(e)$, is the standard consumption utility from working e and receiving wage we. This is added to the reference-dependent, psychological component of utility, where the utility from realized earnings $m(we)$ is compared to utility of expected earnings $m(r_w e)$ such that falling short of expectations leads to a reduction of utility by $\eta\lambda$ times the difference, while exceeding expectations increases utility by η times the difference (and analogous for cost of effort). Thus, λ represents an additional multiplier on losses. Throughout most of the literature, loss aversion (disliking losses more than commensurate gains, i.e. $\lambda > 1$) was assumed to hold universally; recent work, however, suggests that as much as 50% of the population could have a $\lambda < 1$, implying positive surprises are more enjoyable than same-sized shortfalls (Chapman, Snowberg, et al., 2018).

The distinctive feature of the KR model is its handling of the reference point. Specifically, KR propose that agents hold the entire distribution of the outcome space as their expectation, so that ex-ante, each potential realization is compared to every other potential realization and weighted by the relevant densities. In the labor supply context, decision-makers face a potentially stochastic schedule of wages and must commit to an effort level prior to the realization of wages. Thus, when considering the utility of an effort level e', the agent computes the expected consumption utility given the known wage distribution as well as the expected gain-loss utility. Mathematically, this is represented as a double integral over the stochastic reference points ($\mathbf{r} = (r_w, e')$) and the stochastic consumption realizations ($\mathbf{c} = (w, e')$):

$$U(F|G) = \int \int u(\mathbf{c}|\mathbf{r}) dG(\mathbf{r}) dF(\mathbf{c}),$$

where F, G represent the lotteries over the wage-outcome space at a fixed level of effort.

In order to close the model, KR equip it with a rational equilibrium concept known as CPE:

Choice-Acclimating Personal Equilibrium (CPE): A choice $F \in \mathcal{D}$, where $\mathcal{D}$ is the possible outcome space, is a choice-acclimating personal equilibrium if

$$U(F|F) \geq U(F'|F') \; \forall \, F' \in \mathcal{D}.$$

In our context, the effort level e^* is a CPE if its associated ex-ante KR utility – given the distribution of wages it induces – is the largest of all the possible effort choices given the ex-ante distributions they respectively induce.[6] In deriving comparative static predictions throughout the following sections, we will assume that agents seek to maximize their CPE utility.

3.2.1 Measuring Gain-Loss Preferences

We first discuss how to identify gain-loss preferences in the context of real effort. Consider how the introduction of mean-preserving spreads over wages affects individuals across the gain-loss types: loss-averse individuals, suddenly exposed to gains and losses at each potential effort level, would prefer to work fewer tasks under this wage structure because losses loom larger than gains. Gain-lovers instead weight the losses relatively less than the gains, so that additional effort can generate even more positive surprises, leading to increases in effort provision.

Theoretically, a CPE agent facing a deterministic wage maximizes the following utility function:

$$u(we_i|we_i) = we_i - c_i(e_i),$$

so that the optimal effort choice, $e_i^*(w)$, satisfies the first order condition $w = c_i'(e_i)$. Variation in w identifies a cost-of-effort curve, which we assume takes the functional form $c_i(e_i) = \frac{1}{\alpha\gamma_i}(e_i +$

[6]Because the effort level is decided in advance and induces the referent distribution, there is no gain-loss consideration in the effort domain as the realized effort is assumed to equal the referent effort under CPE.

$10)^{\gamma_i}$ as in Augenblick and Rabin (2018).[7] The marginal consideration is now

$$\frac{1}{\alpha}(e_i + 10)^{(\gamma_i - 1)} = w.$$

By introducing a mean-preserving spread over wages, we are able to identify the gain-loss parameter λ_i. Consider the piece-rate $(0.5, w_l; w_h)$, where $w_h > w_l$, representing a contract under which the agent exerts effort e_i knowing that with 50% probability they will earn either $e_i \times w_l$ or $e_i \times w_h$. The associated CPE utils for such an effort choice, e_i, is then

$$0.5w_l e_i + 0.5w_h e_i - 0.25\eta(\lambda_i - 1)(w_h e_i - w_l e_i) - c_i(e_i),$$

where $c_i(e_i)$ is as described above. The optimal effort choice under this wage structure, e_i^*, must then satisfy the first order condition:

$$0.5w_l + 0.5w_h - 0.25\eta(\lambda_i - 1)(w_h - w_l) = \frac{1}{\alpha}(e_i + 10)^{\gamma_i - 1}.$$

By the definition of a mean-preserving spread, $0.5w_l + 0.5w_h = w$ (the deterministic wage). For $\lambda > 1$,

$$0.5w_l + 0.5w_h - 0.25\eta(\lambda_i - 1)(w_h - w_l) < w,$$

so loss-averse individuals reduce their effort in response to a mean-preserving spread of wages. The inequality flips when $\lambda < 1$, indicating a theoretical increase in effort for gain-lovers under the stochastic wage.

[7] 10 is the required number of tasks that all participants must complete in order to receive their completion fee, regardless of how many tasks participants choose to perform at the various rates (which is constrained to be between 0 and 100). Quoting from Augenblick and Rabin (2018): " The parameter α is necessary and represents the exchange rate between effort and the payment amount. If instead $c_i(e_i) = \frac{1}{\gamma_i}(e + 10)^{\gamma_i}$, a requirement such as linear marginal costs (which necessitates $\gamma_i = 2$), would also imply that the marginal cost of e_i tasks is exactly e_i monetary units, regardless of the task type or the payment currency."

3.2.2 Manipulating Reference Points

This section demonstrates how KR comparative static predictions under the canonical labor supply experimental paradigm (Abeler et al., 2011; Gneezy et al., 2017) differ on either side of $\lambda = 1$.

In particular, we consider how KR CPE individuals behave when offered a wage $(p,H;q,L;0.5,w)$ where $L < H$; that is, individuals have a 50% chance of earning a piece-rate, w, per unit of effort, a p% chance of earning \$$H$, and a $q = (0.5 - p)$% chance of earning \$$L$ regardless of effort. The CPE utility induced by a prospective effort level, e, is given by

$$U((p,H;q,L;0.5,we)|(p,H;q,L;0.5,we)) =$$

$$\begin{cases} pH + qL + 0.5we + \eta(1-\lambda)\left[pq(H-L) + 0.5p(H-we) + 0.5q(L-we)\right] - c(e) & we < L < H \\ pH + qL + 0.5we + \eta(1-\lambda)\left[pq(H-L) + 0.5p(H-we) + 0.5q(we-L)\right] - c(e) & L < we < H \\ pH + qL + 0.5we + \eta(1-\lambda)\left[pq(H-L) + 0.5p(we-H) + 0.5q(we-L)\right] - c(e) & L < H < we. \end{cases}$$

Following the development of Gneezy et al. (2017), we study the effects of an increase in p by signing the derivative $\frac{\partial e^*}{\partial p}|_{p+q=0.5}$ when $L \leq we \leq H$.[8] When the considered level of effort yields earnings between the low and high fixed fees, the optimal level of effort can be found by studying the first order condition of

$$0.5w\left[1 + (p-q)\eta(\lambda - 1)\right] = c'(e).$$

Defining $\bar{P} = p + q = 0.5$ and $p - q = 2p - \bar{P} = 2p - 0.5$, we can sign the partial derivative as:

[8]For all other cases, the derivative equals zero so there is no predicted treatment effect; moreover, our design ensures this is the only relevant condition by choosing $H = 20, L = 0, w = 0.2$ while constraining $e \in [0,100]$. Refer to appendix C.1 for more details on the other cases.

$$\frac{\partial e^*}{\partial p}|_{p+q=0.5} = (c'^{-1})'(0.5w[1+(2p-0.5)\eta(\lambda-1)]) * \eta(\lambda-1)w.$$

By the inverse function theorem, $(c'^{-1})'(0.5w[1+(2p-0.5)\eta(\lambda-1)]) * \eta(\lambda-1)w = \frac{1}{c''(e^*)}$ where $0.5w[1+(2p-0.5)\eta(\lambda-1)] = c'(e^*)$. Thus,

$$\frac{\partial e^*}{\partial p}|_{p+q=0.5} = \frac{\eta(\lambda-1)w}{c''(e^*)}$$

and by the assumed convexity of $c(\cdot)$, we know $c''(e^*) > 0$ so that – fixing $\eta = 1$ (any positive number would also hold):

$$\lambda > 1 \implies \frac{\partial e^*}{\partial p}|_{p+q=0.5} > 0$$
$$\lambda < 1 \implies \frac{\partial e^*}{\partial p}|_{p+q=0.5} < 0.$$

Thus, under KR CPE, loss-averse agents are predicted to increase their effort whereas gain-loving individuals decrease their effort in response to an increasing probability of the high fixed payment.

3.2.3 Theoretical Summary

Our theoretical analysis of KR comparative statics across gain-loss types highlights the importance of measuring individual gain-loss preferences when testing for heterogeneous treatment effects. We address this in our experimental design through the use of two stages: the first in which we identify individual gain-loss preferences, and the second in which we measure heterogeneous effects by type. Our first stage offers a series of wages with varying structures, identified off of loss-averse individual's dislike for mean-preserving spreads because of the potential losses introduced for a fixed expected wage. By analyzing the difference in effort choices across these wage structures, we are able to both recover a cost-of-effort function as well as individual level gain-loss parameters.

Our second stage is designed to test the comparative statics when payoffs are structured with high and low fixed payments. By varying the probability of these fixed payments, we alter the reference point and thus the marginal benefit of effort. In this setting, shifting the probability mass onto the high fixed amount leads loss-averse individuals to increase their effort in order to avoid falling short of their expectations. Gain lovers, on the other hand, relish the possibility of a positive wage surprise relative to their level of effort, so shifting the probability mass onto the high fixed amount leads these individuals to reduce their effort.

3.3 Experiment Design

Based on our derivations in section 3.2, we develop a two-stage experimental paradigm to recover individual gain-loss preferences and test for heterogeneous effects within the context of real effort. In stage 1, we present participants with a number of wages, asking them how many tasks they are willing to complete for each wage; the objective is to recover an individual's cost-of-effort function (assuming a functional form) as well as their gain-loss parameter. In stage 2, we present participants with the baseline and treatment conditions in random order, allowing us to test the KR comparative static predictions derived in section 3.2 over gain-loss types. A full set of screenshots for our experiment, designed and implemented in *oTree* (Chen et al., 2016), can be found in Appendix C.3.

3.3.1 Stage 1: Measuring Gain-Loss Preferences

Participants enter the (virtual) lab and are presented with an overview of the experiment's various parts. They are then informed about the task they will be asked to complete – transcribing a row of blurry Greek text, as shown in Figure 3.1 – and go on to complete two practice tasks to

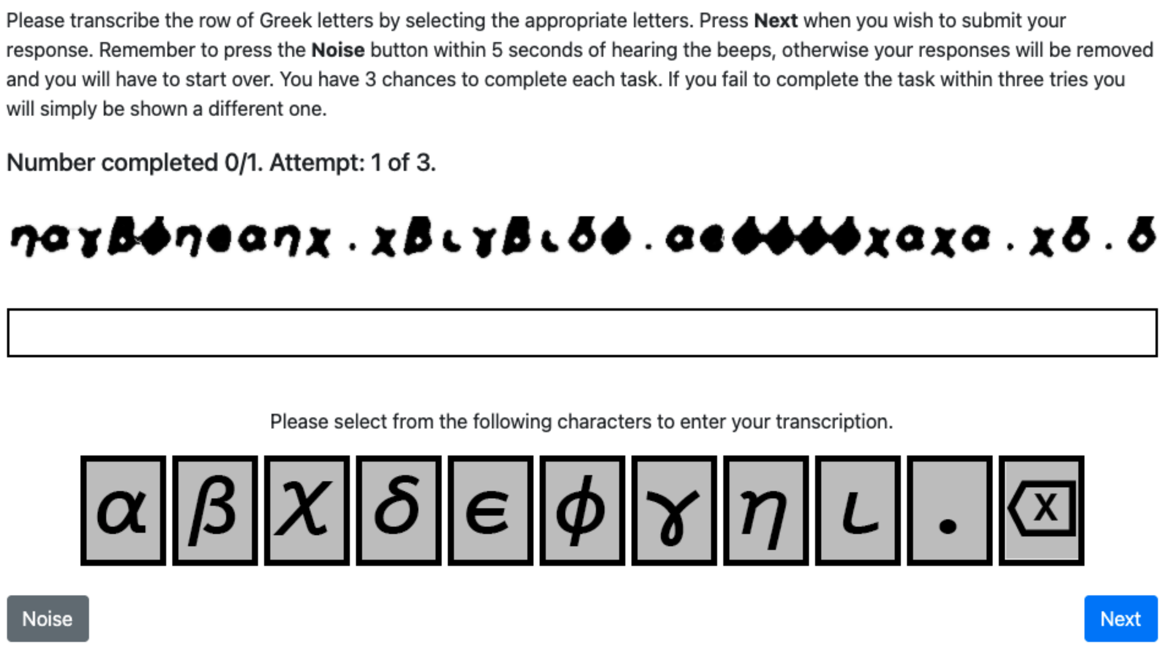


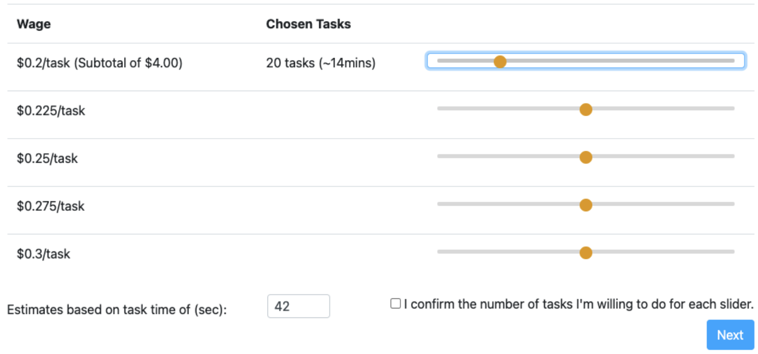
Figure 3.1: **Reduced Form Stage 1**
Notes: Example of a Greek transcription task.

familiarize themselves with the process.[9]

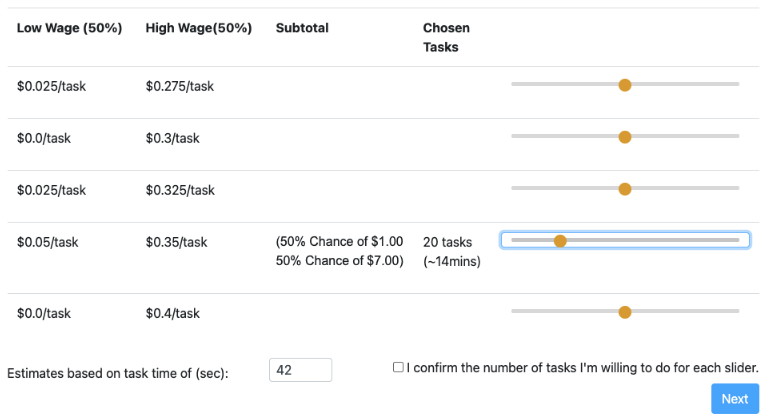
Next, they are presented with 6 screens, each consisting of a list of 5 wages. The screens are ordered randomly, but each list contains either deterministic wages or stochastic wages, never a combination. An example of two such screens is presented in Figure 3.2, highlighting the variation in wage structure. We rely on these two types of wages to jointly identify a cost-of-effort function and gain-loss preferences for each individual, as described in section 3.2.

Beside each wage offering is a slider bar which participants use to indicate how many tasks they are willing to complete at a given wage; as they move the slider, they are shown the earnings subtotal derived from that number of tasks, as well as an estimate of the time to complete these tasks. Similar to Augenblick and Rabin (2018), we select (expected) wages from between \$0.05/task and \$0.30/task (an hourly wage rate between approximately \$4.00 and \$26.00, according to their average time of completion). This range was able to generate sufficient

[9]The task is borrowed from Augenblick and Rabin (2018).



(a) Deterministic Wage Menu



(b) Stochastic Wage Menu

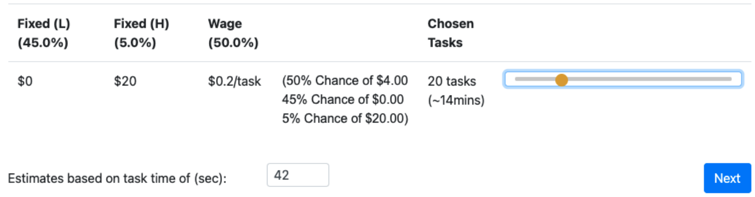
Figure 3.2: **Stage 1 Decision Screens**

Notes: Panel A provides an example of a deterministic wage screen. Panel B provides an example of the stochastic wage screen.

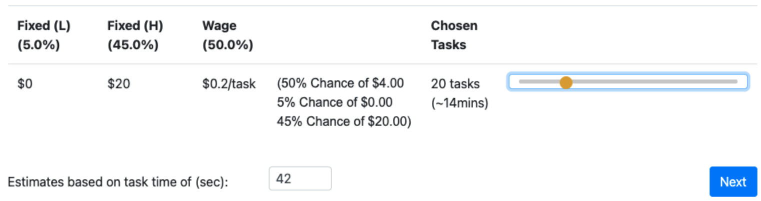
variation in exerted effort to identify a reasonable cost-of-effort function in their study. To ensure incentive compatibility, we remind participants that each choice is equally likely to be selected as the *decision-that-counts*. They are explicitly reminded throughout the experiment that they may be asked complete the number of tasks they indicate.

3.3.2 Stage 2: KR Treatments

After participants make this series of 30 choices, they are presented with an instruction screen informing them about two final effort decisions with slightly different wage structures. We describe our two treatment conditions, explaining that there will be a 50% chance a piece-rate



(a) Baseline Wage



(b) Treatment Wage

Figure 3.3: **Stage 2 Decision Screens**
Notes: Panel A shows Baseline choice screen. Panel B shows the Treatment choice screen.

is paid out, a 45% chance that a fixed payment $\$X$ is paid out regardless of the number of tasks, and a 5% chance of a fixed payment $\$Y$.[10] We then present them with the *baseline* condition – $(0.05, \$20; 0.45, \$0, 0.5, 0.2*e)$ and the *treatment* condition – $(0.45, \$20; 0.05, \$0, 0.5, 0.2*e)$ in random order, eliciting their desired number of tasks for each. Figure 3.3 presents the screens for the treatment and baseline conditions.

3.3.3 Additional Protocols

Once these critical choices have been made, we present subjects with Multiple Price Lists (MPLs), a commonly used technique to estimate gain-loss preferences in the monetary lottery domain (Sprenger, 2015). More specifically, we implement two Probability Equivalent tasks through MPLs, in which we hold fixed a sure payoff of \$5 [\$0] as Option B and offer the gamble $(p, \$10; 0)$ [or $(p, \$3; -\$3.5)$] for p ranging from 0% to 100% in increments of 5% as Option A. Requiring subjects to have standard preferences over money at both extremes (e.g., they are forced to prefer \$5 for sure to a 0% chance of \$10 and prefer a 100% chance of \$10

[10]We are purposefully vague about the amounts of money involved as well as any variation over the two choices because our aim is to obtain within-individual effects and we do not want to tell them about the treatment beforehand.

to \$5 for sure) as well as a single switch-point, the p at which they switch from Option B to Option A informs us about their gain-loss preferences. These measurements are intended to shed light on the relationships between gain-loss preferences across different contexts, and are thus incentivized. Participants know in advance that their payment may come from one of these 42 MPL choices or from the 32 effort provision questions.

Next, we randomly select the *decision-that-counts* for each participant, and regardless of which decision or how many tasks were selected, each participant completes a mandatory 10 transcriptions. If the decision is from one of the MPLs, the computer generates a random number and determines the outcome of the lottery, and the participants receive their payment upon completion of the mandatory tasks and an ensuing survey. If one of the effort decisions is selected, participants first complete the mandatory 10 tasks and then the additional number they indicated in that decision; if the relevant rate is stochastic, uncertainty in wages is not resolved until after they have completed all of the additional tasks.[11]

After all the tasks are completed, participants are presented a series of Raven's matrices (John Raven and Jean Raven, 2003) to measure cognitive ability, followed by a demographic survey (gender, major, age, parental income, risk attitudes). Finally, we resolve any remaining wage lotteries, and pay participants privately.

3.4 Experimental Results

We discuss our results throughout three sections. First, we describe our reduced form data from Stage 1, followed by an overview of our structural estimator along with summary statistics for our recovered distribution of gain-loss preferences. Second, we link our individual

[11]They will have been informed of this in the instructions. This does lead to differential levels of uncertainty when completing both the mandatory tasks and the additional tasks; those with MPL decisions know their payment, those with deterministic wages know their payment, but those with stochastic wages do not. This is intentional, and aimed to mitigate the risk that subjects strategically indicate a large amount of tasks at a highly uncertain wage of $(0.5, \$0.00; 0.5, \$0.60)$ with the intention of leaving if the rate is determined to be low.

measures of gain-loss preferences to behavior in Stage 2, exploring the extent to which our data adheres to the predictions of KR CPE. Finally, we link our MPL measures of gain-loss preferences with our real-effort measures to understand how the distributions vary and the extent to which they are correlated.

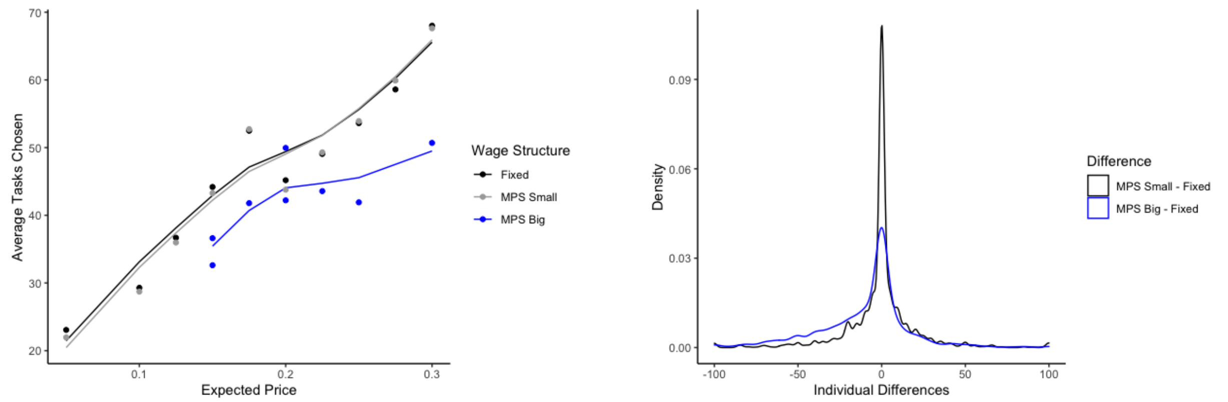
3.4.1 Stage 1

Reduced Form Results

We plot the smoothed average task choices across expected wages in Figure 3.4a, demonstrating our 30 wages are able to generate sufficient variation while avoiding censoring.[12] On average, the task choices across fixed wages and small mean-preserving spreads (MPS) is negligible, but there is a clear distaste for the large MPS wage structure. We explore the relationship between the three wage structures further in Figure 3.4b, wherein we compute the distribution of within-ndividual effort differences for a fixed expected wage. The labor supply curves of *Fixed* and MPS Small clearly mask individual differences – with a peak close to zero and a substantial mass just above and below zero. These individual differences are more pronounced in the *MPS Big* structure, where fewer participants choose the same number of tasks as under the fixed wage, and the negative tail has more mass at large differences.

We interpret these individual differences as instructive about gain-loss preferences, relying on the intuition that loss-averse individuals dislike mean-preserving spreads and are thus predicted to have negative individual differences (while gain-lovers have positive predicted differences). Viewed in this way, Figure 3.4b is suggestive of a heterogeneous underlying distribution of gain-loss attitudes with loss aversion on average, given the relative mass in each tail. In the next section, we build out a structural estimator relying on these individual differences and the mechanics of KR CPE described in section 3.2.

[12]There is a slight dip for some of the wages, which could be related to the relative nature of our wage lists and their random ordering. For instance, a wage of \$0.175 shows up as the largest within its list, but the wage of \$0.20 is the smallest in its comparison set.



(a) Labor Supply

(b) Distribution of Individual Differences

Figure 3.4: **Reduced Form Stage 1**

Notes: Panel A provides a smoothed labor supply function by expected wage and wage structure. Panel B provides a density plot of individual-level differences in effort choices fixing an expected wage, comparing small/large mean-preserving spreads to a deterministic wage.

Structural Estimation

Following our work in section 3.2, we use the variation in both deterministic and stochastic wages, as well as the functional form assumptions, to estimate the cost of effort and gain-loss parameters $(\hat{\alpha}, \hat{\gamma}_i, \hat{\lambda}_i)$ using standard Bayesian methods.[13] Letting e_i^* be the optimal level of effort for an agent facing the wage bundle W, we model the experimentally observed effort, e_i, as normally distributed around the KR CPE optimal, e_i^*, with some noise: $e_i \sim \mathcal{N}(e_i^*, \sigma^2)$.[14]

However, because of the imposed limitations on task choices (they must fall between 0 and 100 tasks), we apply a Tobit correction to account for the fact that the choice of a corner solution may not satisfy the standard tangency conditions of the utility maximization problems. To compute the estimates, we draw samples from *rstan* (Stan Development Team, 2020), updat-

[13]Following standard practice when estimating λ in the KR model, we fix $\eta = 1$.

[14]For deterministic wages, $e_i^* = (\alpha w)^{\frac{1}{\gamma_i - 1}} - 10$; for stochastic wages,

$$e_i^* = (\alpha(0.5w_l + 0.5w_h - 0.25\eta(\lambda_i - 1)(w_h - w_l))^{\frac{1}{\gamma_i - 1}} - 10.$$

Note that there are important interactions between the CPE assumption and the cost-of-effort assumption. In particular, $\lambda > 3$ is ruled out under CPE since it has unrealistic implications – including violations of First Order Stochastic Dominance (see Masatlioglu and Raymond (2016) for more details). If we did not rule this out, we would have computational issues in our estimation as $\lambda \geq 3$ produces imaginary values (stemming from taking the square root of a negative number) unless $\gamma = 2$.

ing our likelihood given prior beliefs and the observed data. For our choice of prior, we rely on work from Goette, Graeber, et al. (2018) and Kellogg (n.d.).[15] Moreover, because our Bayesian estimator returns a posterior mean for each individual, regardless of how well λ is identified, we use the width of the resulting 95% credible interval to gauge how informative each individual's 30 choices are about their gain-loss parameter.[16] As described in the pre-analysis plan, we expect 5-20% of individuals to have poorly identified λ_i, so we conservatively drop observations with a credible interval width above the 80th percentile.[17]

3.4.2 Stage 1 In-Sample Fit

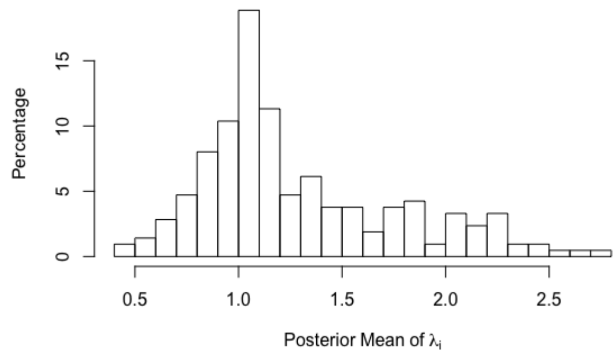
The resulting distribution of participants' gain-loss preferences is illustrated in Figure 3.5a, where we assign each individual the posterior mean given by our estimator.[18] We estimate 28% of participants to be gain-loving in the domain of labor supply, a similar number as the weighted average of 22% reported in Chapman, Snowberg, et al. (2018) in the monetary lottery domain. Importantly, this confirms that heterogeneity of gain-loss attitudes is ubiquitous across contexts in which reference dependence is prominently applied, with documented evidence indicating at least a quarter of the population is gain loving in the monetary, exchange, and effort domains (Chapman, Snowberg, et al., 2018; Goette, Graeber, et al., 2018; Kellogg, n.d.). Researchers planning to apply models of reference dependence must seriously consider heterogeneity in gain-loss attitudes and its implications in their context.

[15] Our estimator performed quite well in a synthetic recovery exercise (IQR of the estimated bias for λ_i was [-0.053,0.042], mean of -0.005), and was more computationally efficient relative to an analogous structural MLE estimator; for more information about these results, please contact the authors.

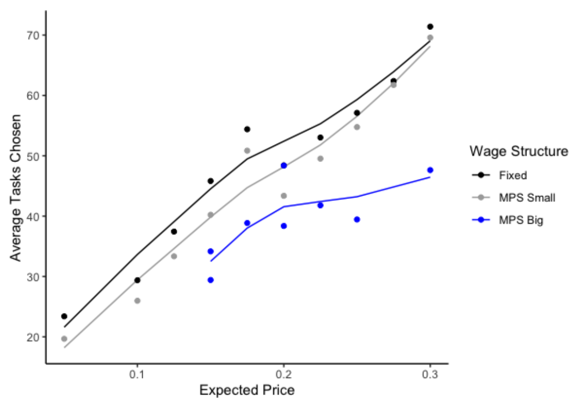
[16] We also estimate λ_i under a more loss averse prior and compare the distance between the posterior λ_i under the two priors; this is highly correlated with the width of the Credible Interval ($p < 0.001$), and thus serves as an alternate justification for our approach: the λ_i that we are least certain about are the ones that are most sensitive to the prior.

[17] All subsequent results in the main text rely on this trimmed sample, though full sample results can be found in Appendix C.2.

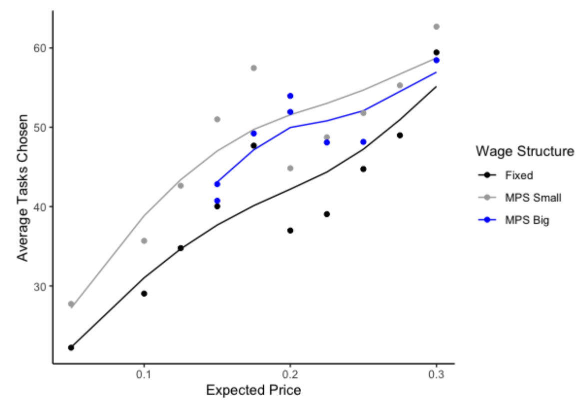
[18] Summary statistics including the mean and standard deviation of our posterior samples for the key parameters are described in Table C.1, with point estimates on the cost-of-effort function comparable to those in Augenblick and Rabin (2018).



(a) Sample Distribution



(b) Labor Supply, Loss Averse



(c) Labor Supply, Gain Loving

Figure 3.5: **Structural Estimation, Stage 1**

Notes: Given the 30 choices from each participant, we assign each individual a λ_i using the posterior mean from our estimator. Individuals with $\lambda_i > 1$ are categorized as loss averse. The lines represent smoothed labor supply curves over wage structure, and the points indicate means.

In Figure 3.5b and 3.5c, we plot the stage 1 labor supply curves by wage structure separately for each gain-loss type. As derived in section 3.2, loss-averse individuals dislike mean-preserving spreads, reducing their effort relative to the fixed wage for both *MPS Small* and *MPS Big*. Gain-lovers, on the other hand, respond to *MPS Small* with substantially more effort than under the fixed wage structure, while their average effort under *MPS Big* falls between the two curves. These plots provide further evidence that our average labor supply curves in Figure 3.4a mask underlying heterogeneity over gain-loss types, and demonstrate an in-sample fit that aligns with our results derived under KR CPE.

3.4.3 Stage 2: KR Comparative Statics

Given our estimate of each participant's gain-loss preference, we analyze effort provision in our second stage conditions. In random order, each participant is asked how many tasks they wish to complete in the baseline condition, with payment given by $(0.05, \$20; 0.45, \$0, 0.50, 0.2 * e)$, and the treatment condition, $(0.45, \$20; 0.05, \$0, 0.50, 0.2 * e)$.

Column 1 of Table 3.1 contains the aggregate treatment effects from our median regression on our trimmed sample.[19] When faced with a 45% chance of earning \$0 regardless of effort, participants select a median of 40 tasks. Shifting this probability mass to the high fixed payment of \$20 leads to a median increase of 10 tasks ($p < 0.05$), consistent with the results in Abeler et al. (2011).

The interaction between posterior λ_i and treatment is presented in Column 2, informing us about the predictive ability of our estimated λ_i on effort in both baseline and treatment conditions. Our coefficient on λ of -12.82 ($p < 0.05$) indicates a statistical relationship between our stage 1 measure of gain-loss preference and stage 2 baseline effort, where effort decreases with loss aversion. Similarly, the coefficient on our interaction term of 11.72 (*n.s.*) is in line with the positive slope predicted by KR CPE, indicating that loss-averse individuals increase the number of tasks they work in the treatment relative to gain-lovers.

Column 3 provides an interaction specification under a coarser measure of gain-loss attitudes – relying on an indicator for loss aversion, $\mathbf{1}(\lambda_i > 1)$, rather than the individual's estimated λ_i. Once again, our stage 1 measure of gain-loss attitudes is predictive of out-of-sample effort in stage 2, with participants categorized as loss averse choosing to work fewer tasks in baseline ($p < 0.01$) but more tasks in the treatment condition ($p < 0.05$).

Finally, Column 4 presents a reduced form specification, where loss aversion is defined

[19] Because the individual-level treatment effects can have high variance, our analysis focuses on median treatment effects to avoid the influence of outliers within bins of λ_i. Average treatment effect figures and tables corresponding to those in the main text can be found in Appendix C.2.

as choosing to work for fewer total tasks under stochastic compared to deterministic wages.[20] The coefficients from this regression have the same sign as those under our coarse structural measure of loss aversion, but are diminished in magnitude and statistically insignificant.

Table 3.1: **Treatment Effects over Gain-Loss Preference**

	Dependent Variable:	*Effort Choice*		
	(Agg)	(Fine)	(Coarse)	(RF)
Constant	40.00	53.15	45.00	40.00
	(3.75)	(6.91)	(3.40)	(4.96)
λ		-12.82		
		(5.32)		
$\mathbf{1}(\lambda > 1)$			-17.00	
			(6.40)	
$\mathbf{1}(\lambda > 1)_{\text{RF}}$				-4.00
				(6.64)
Treatment	10.00	-2.15	5.00	5.00
	(4.26)	(9.91)	(5.42)	(6.23)
$\lambda \times$ Treatment		11.72		
		(8.55)		
$\mathbf{1}(\lambda > 1) \times$ Treatment			17.00	
			(8.69)	
$\mathbf{1}(\lambda > 1)_{\text{RF}} \times$ Treatment				9.00
				(8.04)
Observations	212	212	212	212

Notes: Quantile ($\tau = 0.5$) regression with bootstrapped standard errors in parentheses. Agg represents the regression of effort on a constant and a treatment indicator (over all gain-loss types). Fine represents the regression using individual posterior λ_i from our trimmed sample based on the credible interval width. Coarse instead relies on an indicator for whether the posterior λ_i indicates loss aversion. RF stands for our reduced form measure of loss aversion, computed by summing up over all individual differences between MPS structured and Fixed wages (for a given expected wage). Loss aversion is defined as having a negative sum here, following the intuition that loss-averse individuals dislike MPS.

We present these results graphically in Figure 3.6, binning individuals within λ buckets and showing median effort across the gain-loss space. First, we plot the median number of tasks chosen in the baseline condition alongside the CPE predictions in Figure 3.6a. Selected effort in this baseline condition is tightly linked to λ_i, reflecting our findings from Table 3.1 where more loss-averse individuals exert less effort under a 45% chance of earning \$0. This pattern demonstrates the predictive validity of our stage 1 measure of λ_i in this out-of-sample exercise. Interestingly, while the observed data has a similar slope as the CPE prediction, the

[20]For each expected wage, we compute the individual difference between MPS Small (and/or MPS Big when relevant) and Fixed, and sum over all of these differences within individual. When this sum is positive, individuals are defined as Reduced Form Loss Averse.

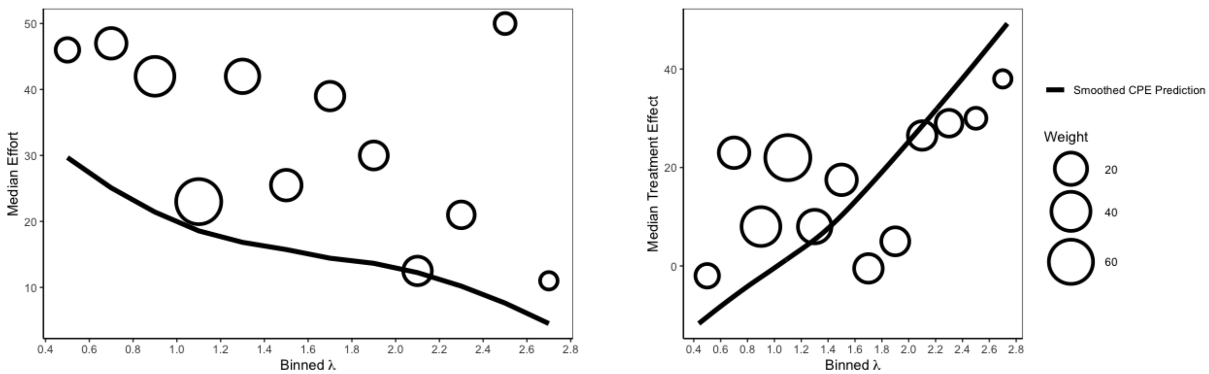
level is shifted upwards, indicating higher observed effort in the baseline condition relative to theoretical predictions.

The link between treatment effects and λ_i is depicted in Figure 3.6b. Our smoothed CPE Prediction line demonstrates the theoretical heterogeneous treatment effects described in section 3.2, where gain-loving participants are predicted to work fewer tasks under the treatment condition while loss-averse participants are predicted to work more tasks (relative to baseline). The observed effects hew closely to this predicted line, particularly when considering the subset of highly loss averse individuals. While individuals almost universally exert more effort when there is a 45% chance of earning $20 (as opposed to a 45% chance of $0), the magnitude of this difference seems to trend with λ_i.

Our experimental results demonstrate that previously measured gain-loss preferences predict behavior in our out-of-sample baseline and treatment conditions, with observed behavior displaying similar patterns as KR predictions. This helps align the mixed results in the laboratory studies with those from the NYC cab driver literature, providing a strong empirical foundation for KR in the domain of labor supply. While aggregate treatment effects loosely reflect the average gain-loss preference of loss aversion in our study, this masks the underlying heterogeneity in behavior that is linked with individual gain-loss types.

3.4.4 Gain-Loss Preferences Across Domains

In addition to our effort-based measure of gain-loss preferences, we also provide two measures based on choices from a set of Probability Equivalent Multiple Price Lists. Following Sprenger (2015), we construct our estimate of λ_i from each of the two MPLs, relying on the switching probability to identify gain-loss preferences. Because the identified λ_i vary considerably as a function of the parameters of the MPL, we coarsen our measure to an indicator for gain lovingness based on whether participants switch to the risky option before or after the Expected Utility equivalence point.



(a) Baseline Effort

(b) Treatment Effects

Figure 3.6: **Stage 2 Results**

Notes: Smoothed CPE predictions are based on posterior λ_i means in our sample, while our Binned λ_i takes the median individual effort (in panel a) and treatment effect (panel b) for those with $\lambda_i \in [0.4, 0.6], [0.6, 0.8] \dots$ and plots them based on the midpoint.

Table 3.2 presents the correlation between our indicator for gain lovers across the lottery and effort domains. For both the mixed and gains-only lottery MPLs, there is no statistically significant correlation between our measures of gain lovingness. Interestingly, each of our three measures yields a sizable minority of gain-lovers – 28% from our effort indicator, 26% in our mixed MPL, and 16% in the gains MPL – but classification as gain-loving in the lottery MPLs is not predictive of gain lovingness in the effort domain.

Table 3.2: **Gain-Loss Preferences Across Domains**

	Dependent Variable: *Effort Gain Lover* $(=1)$		
	(Mixed)	(Gains)	(Both)
$\mathbf{1}(\lambda < 1)_{Mixed}$	0.060		0.067
	(0.073)		(0.073)
$\mathbf{1}(\lambda < 1)_{Gains}$		-0.057	-0.067
		(0.081)	(0.079)
Constant	0.27	0.29	0.28
	(0.035)	(0.034)	(0.038)
Observations	212	212	212

Notes: OLS regression where the dependent variable is a participant's classification as gain loving according to our structural effort estimator ($\lambda_{i,effort} < 1$). $\mathbf{1}(\lambda < 1)_{Mixed}$ represents the gain loving classficiation from our mixed MPL as the independent variable, whereas $\mathbf{1}(\lambda < 1)_{Gains}$ represents the measure from our gains only MPL.

3.5 Conclusion

Reference-dependent preferences are a central contribution of behavioral economics, with expectations-based reference dependence representing an important advance by endogenizing the reference point. In the domain of labor supply, these models have consistently predicted behavior among cab drivers and bike messengers (C. Camerer et al., 1997; Crawford and Meng, 2011; Fehr and Goette, 2007; Thakral and To, Forthcoming). In the lab, where researchers are able to more precisely manipulate reference points to test comparative statics, evidence for KR is mixed (Abeler et al., 2011; Gneezy et al., 2017).

Rather than interpreting these past results as a shortcoming of KR, recent work by Goette, Graeber, et al. (2018) provides an alternative explanation: KR comparative statics change signs across gain-loss types in such a way that aggregate treatment effects need not respect aggregate gain-loss preferences. As such, a compelling lab test of KR comparative statics requires the researcher to measure gain-loss preferences prior to analyzing treatment effects.

We adapt this two-stage paradigm to the canonical real effort KR paradigm, relying on 30 effort choices across stochastic and deterministic wage structures to identify individual gain-loss preferences. In our second stage, we demonstrate the strong predictive power of λ_i in both baseline and treatment conditions, with median treatment effects hewing closely to KR theory.

Interestingly, when connecting our measure of λ_i from the real effort task and our secondary probability equivalent MPLs, we find no correlation between the two. This highlights the importance of measuring individual gain-loss attitudes in the relevant domain in order to obtain predictive validity, as well as emphasizing the need for a greater understanding of how gain-loss attitudes vary over context or even over time.

Our investigation has a number of implications for future studies. First, considered alongside the results in Chapman, Snowberg, et al. (2018), Goette, Graeber, et al. (2018), and Kellogg (n.d.), there is now ample evidence that heterogeneity of gain-loss preferences is a fixture within the core contexts to which reference-dependent models have been applied. Second, tests of KR comparative statics that account for individual gain-loss attitudes provide a strong empirical foundation for the model, furthering the interpretation that the prior mixed results may be confounded by the underlying heterogeneity in gain-loss attitudes. Taken together, this suggests future developments and applications of reference-dependent models must carefully consider how heterogeneity of gain-loss attitudes applies in their context.

3.6 Acknowledgements

This work is currently being prepared for submission for publication. Campos-Mercade, Pol; Goette, Lorenz; Sprenger, Charles. The dissertation author was a principal investigator in this analysis.

Appendix A

A.1 Replication and Reconciliation with Pre-Analysis Plan

In this section we report the methodology and corresponding analyses from earlier versions of this paper (https://papers.ssrn.com/sol3/papers.cfm?abstractid=3170670 and https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3589906) as specified in the pre-registration plan of our replication study (https://www.socialscienceregistry.org/trials/3124). The key difference is that while our approach in the present version of the paper relies on a mixed-logit methodology following a suggestion of an anonymous referee, our previous approach employed standard logit methods. All our previous results are closely in line with those obtained using the new methodology. Here we provide a summary of the central exercises conducted in prior versions of the manuscript. For the complete analysis please see https://papers.ssrn.com/sol3/papers.cfm?abstractid=3170670 and https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3589906.

A.1.1 Stage 1: Identifying Gain-Loss Attitudes

Our previous methodology relied on the same preference statements that we introduced in Section 4.1, but focused only on the liking preference statements. As noted in the main text, the liking data indicate both a substantial endowment effect and potential differences in utility across objects. We construct a simple structural model of the liking preference statement based upon standard random utility methods (McFadden, 1974) with the objective of capturing the source of both of these features: gain-loss attitudes and differences in intrinsic utility for the two objects.

Consider an individual endowed with X that is asked to provide ratings statements for both X and Y. Under the KR model, an individual evaluates their endowment, X, based upon $U(X,0|X,0)$. Given that the agent is endowed with X and is uninformed of the possibility of confiscation at the time of the ratings, they plausibly evaluate Y based upon $U(0,Y|X,0)$. With standard logit shocks, ε_X and ε_Y, the parameters associated with these KR utilities are easily estimated. We assume subjects will provide a higher rating for their endowed object, X, if

$$U(X,0|X,0)+\varepsilon_X > U(0,Y|X,0)+\varepsilon_Y+\delta,$$

where δ is a discernibility parameter which accounts for the fact that the goods may be given identical ratings (for use of such methods, see, e.g., Cantillo et al., 2010). Similarly, subjects provide a higher rating for the alternative object, Y, if

$$U(0,Y|X,0)+\varepsilon_Y > U(X,0|X,0)+\varepsilon_X+\delta,$$

and provide the same rating if the difference in utilities falls within the range of discernibility,

$$|U(X,0|X,0)+\varepsilon_X-(U(0,Y|X,0)+\varepsilon_Y)| \leq \delta.$$

Under the functional form assumptions of section 2 with $\eta = 1$, for someone endowed with object X, we obtain familiar logit probabilities for the ranking of ratings $R(X)$ and $R(Y)$,

$$
\begin{aligned}
P(R(X) > R(Y)) &= \frac{\exp(U(X,0|X,0))}{\exp(U(X,0|X,0)) + \exp(U(0,Y|X,0) + \delta)} = \frac{\exp(X)}{\exp(X) + \exp(2Y - \lambda X + \delta)} \\
P(R(Y) > R(X)) &= \frac{\exp(U(0,Y|X,0))}{\exp(U(0,Y|X,0)) + \exp(U(X,0|X,0) + \delta)} = \frac{\exp(2Y - \lambda X)}{\exp(X + \delta) + \exp(2Y - \lambda X)} \\
P(R(X) = R(Y)) &= 1 - P(R(X) > R(Y)) - P(R(Y) > R(X)),
\end{aligned}
$$

where the intrinsic utility values, X and Y, the discernibility parameter δ, and the gain-loss parameter, λ, are the desired estimands.[1] We normalize one of the good's values to be $Y = 1$, and estimate the remaining parameters via maximum likelihood.

Table A.1 provides aggregate estimates of intrinsic utilities, λ and δ, separately for each pair of goods in both the initial study and our replication. In each case we find aggregate support for loss aversion, $\lambda > 1$, though less pronounced in our replication study.

Individual Gain-Loss Attitudes

The aggregate estimates show evidence of loss aversion. To construct bounds for estimates of individual gain-loss attitudes, we evaluate individual choices assuming average utility and discernibility values. For example, consider an individual endowed with the pen set in Pair 1 in the initial study. At the aggregate estimates of δ and X for Pair 1, if this individual were to state a higher rating for the pen set than for the USB stick, it would imply $0.632 > 2 - \hat{\lambda} * 0.632 + 0.549$ or $\hat{\lambda} > 3.03$. Similarly, stating a higher rating for the USB stick

[1] For someone endowed with the alternative object, Y, these same probabilities are

$$
\begin{aligned}
P(R(X) > R(Y)) &= \frac{\exp(U(X,0|0,Y))}{\exp(U(X,0|0,Y)) + \exp(U(0,Y|0,Y) + \delta)} = \frac{\exp(2X - \lambda Y)}{\exp(Y + \delta) + \exp(2X - \lambda Y)} \\
P(R(Y) > R(X)) &= \frac{\exp(u(0,Y|0,Y))}{\exp(U(0,Y|0,Y)) + \exp(U(X,0|0,Y) + \delta)} = \frac{\exp(Y)}{\exp(Y) + \exp(2X - \lambda Y + \delta)} \\
P(R(X) = R(Y)) &= 1 - P(R(X) > R(Y)) - P(R(Y) > R(X)).
\end{aligned}
$$

Table A.1: **Prior Analysis: Aggregate Parameter Estimates**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Initial Study				*Replication Study*			
	Est.	(Std. Err.)	Est.	(Std. Err.)	Est.	(Std. Err.)	Est.	(Std. Err.)
	Pair 1		*Pair 2*		*Pair 1*		*Pair 2*	
Gain-Loss Attitudes:								
$\hat{\lambda}$	1.56	(0.14)	1.29	(0.12)	1.18	(0.15)	1.12	(0.13)
Utility Values:								
$\hat{X}_1$ *(Pen Set)*	0.63	(0.05)			0.66	(0.06)		
$\hat{Y}_1$ *(USB Stick)*	1	-			1	-		
$\hat{X}_2$ *(Picnic Mat)*			0.84	(0.05)			1.05	(0.07)
$\hat{Y}_2$ *(Thermos)*			1	-			1	-
Discernibility:								
$\hat{\delta}$	0.55	(0.06)	0.45	(0.05)	0.45	(0.06)	0.62	(0.07)

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

would imply $\hat{\lambda} < 1.30$,[2] and stating the same rating implies $\hat{\lambda} \in [1.30, 3.03]$. Of these three possible cases, two demonstrate evidence of loss aversion $\hat{\lambda} > 1$, while the other case is plausibly loss neutral as $\hat{\lambda} = 1$ can rationalize the ratings.[3] In total, there exist twelve cases of endowments and relative liking statements.

Overall, in our initial study 217 subjects (35.7 percent) are categorized as loss-averse, 240 (39.5 percent) are categorized as potentially loss-neutral, and 150 (24.7 percent) are categorized as gain-loving. In our replication study, 124 subjects (29.7 percent) are categorized as loss-averse, 185 (44.4 percent) are categorized as potentially loss-neutral, and 108 (25.9 percent) are categorized as gain-loving. These are the taxonomies of individual gain-loss types used in our analysis.

[2]To state a higher rating for the USB implies $2 - \hat{\lambda} * 0.632 > 0.632 + 0.549$ or $\hat{\lambda} < 1.30$.

[3]It may seem prima-facie surprising that providing the same rating in this case is consistent with loss aversion. The logic is simple: given that the pen set has substantially lower intrinsic utility than the USB stick, one must be loss-averse to rate them equally.

A.1.2 Stage 2: Heterogeneous Treatment Effects

Table A.2, presents linear probability models for Stage 2 behavior with dependent variable *Exchange (=1)*. Panels A and B provide separate results for our initial and replication studies. Beginning with the initial study, we find a null average treatment effect in Column (1). In Condition B, 36.5 percent of subjects choose to exchange, demonstrating a significant endowment effect relative to the null hypothesis of 50 percent exchange, $F_{1,605} = 18.32,\ (p < 0.01)$. Probabilistic forced exchange has a null average treatment effect, increasing trading probabilities by only 0.4 percentage points on aggregate. Columns (2) through (4) conduct the same regressions separately for subjects categorized as loss-averse, loss-neutral, and gain-loving, based on their Stage 1 liking statements. Panel A of Table A.2 shows a dramatic heterogeneous treatment effect. Loss-averse subjects exhibit a statistically significant endowment effect in Condition B, and grow more approximately 16 percentage points more willing to exchange in Condition F. Gain-loving subjects exhibit no endowment effect in Condition B, and grow approximately 25 percentage points less willing to exchange in Condition F. The heterogeneous treatment effect over gain-loving and loss-averse subjects of roughly 40 percentage points closely follows our theoretical development on the sign of comparative statics, and is significant at all conventional levels, $F_{1,363} = 15.76,\ (p < 0.01)$.

As detailed in the main text, we registered and conducted an exact replication in the summer of 2018 with 417 subjects, again at the University of Bonn. The registration of our pre analysis plan, including power calculations, can be found at `https://www.socialscienceregistry.org/trials/3124`. The number of subjects for the replication was guided by a requirement of 80 percent power for the 40 percentage point difference in treatment effect between gain-loving and loss-averse subjects noted above. Ex-post our initial study was slightly over-powered and the replication was thus conducted with around 400 subjects. Panel B of Table A.2 provides the replication analysis analogous to that presented in Panel B. The null average treatment effect, positive treatment effect for loss-averse subjects, and negative treatment effect for gain-loving

Table A.2: **Prior Analysis: Exchange Behavior and Probabilistic Forced Exchange**

	(1)	(2)	(3)	(4)
	Dependent Variable: Exchange (=1)			
	Full Sample	Loss Averse	Loss Neutral	Gain Loving
Panel A: Initial Study				
Condition F	0.004	0.158	0.027	-0.248
	(0.034)	(0.067)	(0.066)	(0.078)
Constant (Condition B)	0.365	0.330	0.361	0.429
	(0.028)	(0.049)	(0.053)	(0.067)
R-Squared	0.000	0.025	0.001	0.072
# Observations	607	217	240	150
H_0: Zero Endowment Effect in B	$F_{1,605}$=18.32	$F_{1,215}$=12.21	$F_{1,238}$=6.85	$F_{1,148}$=1.15
	$(p < 0.01)$	$(p < 0.01)$	$(p < 0.01)$	$(p = 0.29)$
H_0: Zero Treatment Effect (F-B)	$F_{1,605} = 0.01$	$F_{1,215} = 5.64$	$F_{1,238} = 0.17$	$F_{1,148} = 10.18$
	$(p = 0.90)$	$(p = 0.02)$	$(p = 0.68)$	$(p < 0.01)$
H_0: Constant (col. 2) = Constant (col. 4)				$F_{1,363} = 1.44$
				$(p = 0.23)$
H_0: Condition F (col. 2) =Condition F (col. 4)				$F_{1,363} = 15.76$
				$(p < 0.01)$
Panel B: Replication Study				
Condition F	-0.010	0.206	-0.073	-0.160
	(0.044)	(0.085)	(0.075)	(0.094)
Constant (Condition B)	0.399	0.271	0.444	0.474
	(0.030)	(0.058)	(0.059)	(0.067)
R-Squared	0.000	0.045	0.005	0.027
# Observations	417	124	185	108
H_0: Zero Endowment Effect in B	$F_{1,415}$=7.97	$F_{1,122}$=15.40	$F_{1,183}$=0.89	$F_{1,106}$=0.16
	$(p < 0.01)$	$(p < 0.01)$	$(p = 0.35)$	$(p = 0.69)$
H_0: Zero Treatment Effect (F-B)	$F_{1,415} = 0.05$	$F_{1,122} = 5.79$	$F_{1,183} = 0.95$	$F_{1,106} = 2.92$
	$(p = 0.83)$	$(p = 0.02)$	$(p = 0.33)$	$(p = 0.09)$
H_0: Constant (col. 2) = Constant (col. 4)				$F_{1,228} = 5.22$
				$(p = 0.02)$
H_0: Condition F (col. 2) =Condition F (col. 4)				$F_{1,228} = 8.33$
				$(p < 0.01)$

Notes: Ordinary least square regression. Robust standard errors in parentheses. Null hypotheses tested for 1) zero baseline endowment effect, regression (Constant = 0.5); 2) zero treatment effect (F-B); 3) Identical Condition B behavior across loss-averse and gain-loving subjects (Constant (col. 2) = Constant (col. 4)); 4) Identical treatment effects of forced exchange across loss-averse and gain-loving subjects (Forced Exchange (col. 2) = Forced Exchange (col. 4)). Hypotheses 3 and 4 tested via interacted regression with observations from columns (2) and (4).

subjects are all reproduced with accuracy. Indeed, the 40 percentage point heterogeneous treatment effect in our initial study is echoed in a 37 percentage point difference between gain-loving and loss-averse subjects in our replication study.

Our replication study reproduces with precision the heterogeneous treatment effect over gain-loss types obtained in our initial study under our prior methods. Subjects classified as loss-averse respond to Condition F by increasing their willingness to exchange; subjects classified as gain-loving respond by decreasing their willingness to exchange.

A.2 Additional Theoretical Analysis: PE and PPE

This appendix provides additional theoretical development for heterogeneity in response to probabilistic forced exchange under Personal Equilibrium (PE) and the PE refinement, Preferred Personal Equilibrium, PPE. Throughout, our maintained assumptions will be $X, Y, \lambda, \eta > 0$. We begin with the restrictions on behavior implied by PE. To begin, we focus on Condition B and a choice set consisting of pure strategy choices $\mathcal{D} = \{(X,0),(0,Y)\}$. In this setting, there are two potential PE selections, $[\mathbf{c},\mathbf{r}] = [(X,0),(X,0)]$ and $[\mathbf{c},\mathbf{r}] = [(0,Y),(0,Y)]$. The individual can support not exchanging, $[\mathbf{c},\mathbf{r}] = [(X,0),(X,0)]$, in a PE if

$$U(X,0|X,0) \geq U(0,Y|X,0),$$

or

$$X \geq \frac{1+\eta}{1+\eta\lambda}Y. \tag{A.1}$$

Note that the smallest value of X at which the individual can support not exchanging, $\underline{X}_{B,PE} = \frac{1+\eta}{1+\eta\lambda}Y$, is inferior to Y if $\lambda > 1$. As such, loss-averse individuals with $\lambda > 1$ may be able support not exchanging X for Y even if Y would be preferred on the basis of intrinsic utility alone. This describes the mechanism by which the KR model generates an endowment effect in PE.

Similarly, the individual can support exchanging, $[\mathbf{c},\mathbf{r}] = [(0,Y),(0,Y)]$, if

$$U(0,Y|0,Y) \geq U(X,0|0,Y),$$

or

$$X \leq \frac{1+\eta\lambda}{1+\eta}Y.$$

The highest value of X at which the agent can support exchanging, $\overline{X}_{B,PE} = \frac{1+\eta\lambda}{1+\eta}Y$, increases linearly with λ. For $\underline{X}_{B,PE} \leq X \leq \overline{X}_{B,PE}$, there will be multiple equilibria, with the agent able to support both exchanging and not exchanging as a PE.

Note that for gain-loving individuals with $\lambda < 1$ it is also possible for $\overline{X}_{B,PE} < X < \underline{X}_{B,PE}$, such that no pure strategy PE selection from the assumed $\mathcal{D}$ exists. In this region, if $\mathcal{D}$ were to include all mixtures of exchanging and not exchanging, there would be a mixed strategy PE of not exchanging with a given probability, p. Below, we provide this analysis. Figure A.1 provides the pure strategy PE cutoffs associated with exchanging not exchanging in Condition B.

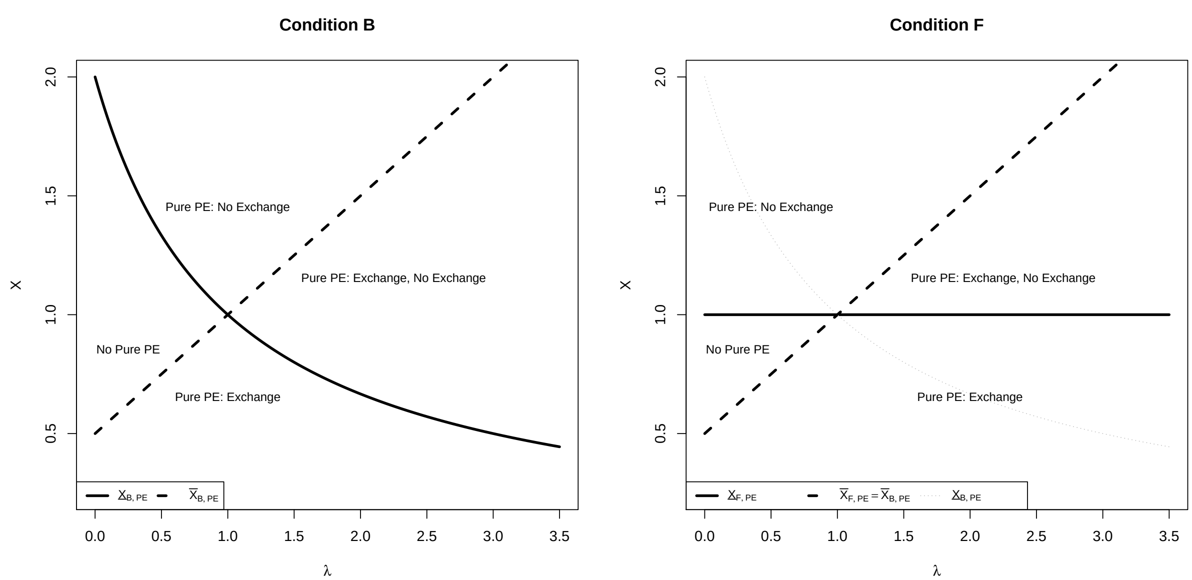


Figure A.1: **Gain-Loss Attitudes and Theoretical Pure PE Strategy Thresholds**
Notes: Threshold values for pure strategy PE for agent endowed with X, assuming $Y = 1$ and $\eta = 1$.

Now, consider Condition F. The potential selections for someone endowed with X are $\mathcal{D} = \{0.5(X,0) + 0.5(0,Y), (0,Y)\}$, with the first element reflecting attempting not to exchange and the second reflecting exchange, as before. The individual can support attempting not to exchange in a PE if

$$U(0.5(X,0)+0.5(0,Y)|0.5(X,0)+0.5(0,Y)) \geq U(0,Y|0.5(X,0)+0.5(0,Y)),$$

or

$$X \geq Y. \tag{A.2}$$

Under forced exchange, the individual can support attempting to retain X in PE only on the basis of intrinsic utility values, regardless of the level of λ.

Though probabilistic forced exchange alters the PE considerations associated with not exchanging, it leaves unchanged the PE considerations associated with exchanging. The individual can support exchanging in PE if

$$U(0,Y|0,Y) \geq U(0.5(X,0)+0.5(0,Y)|0,Y),$$

which as before is

$$X \leq \frac{1+\eta\lambda}{1+\eta} Y.$$

Hence, $\overline{X}_{F,PE} = \overline{X}_{B,PE}$.

The manipulation of probabilistic forced exchange changes the PE cutoff for not exchanging from $\underline{X}_{B,PE} = \frac{1+\eta}{1+\eta\lambda} Y$ to $\underline{X}_{F,PE} = Y$. There is no longer any possibility in PE for a loss-averse individual to support keeping their object if $Y > X$. A loss-averse individual with $\lambda > 1$ and valuation $\underline{X}_{B,PE} < X < \underline{X}_{F,PE}$ moves from a position of multiple PE in Condition B, to having a unique PE to exchange in Condition F. Such an individual plausibly grows more willing to exchange when moving from Condition B to Condition F. Similarly, a gain-loving individual with $\lambda < 1$ and valuation $\underline{X}_{F,PE} < X < \underline{X}_{B,PE}$ moves from a position of no pure strategy PE in Condition B to having a unique PE of exchange in Condition F. Such an individual plausibly grows less willing to exchange when moving from Condition B to Condition F. Figure A.1, illustrates these changing pure strategy PE considerations from Condition F to Condition B. The direction of these comparative statics is identical to that of our CPE analysis in the main text.

A.2.1 PE Mixed Strategy Analysis

To provide more complete analysis, particularly when there is no pure strategy PE, we now elaborate PE and PPE formulations when the choice set $\mathcal{D}$ includes all available mixtures of exchanging and not exchanging. For Condition B , we assume $\mathcal{D}_B = \{p \in [0,1] : p(X,0)+(1-p)(0,Y)\}$, allowing all mixtures of exchange and no exchange to be chosen. A given mixture, p, will be PE if

$$U(p(X,0)+(1-p)(0,Y)|p(X,0)+(1-p)(0,Y)) \geq$$
$$U(q(X,0)+(1-q)(0,Y)|p(X,0)+(1-p)(0,Y)) \ \forall\, q \in [0,1],$$

or

$$pX+(1-p)Y+p(1-p)\eta(1-\lambda)(X+Y) \geq$$
$$qX+(1-q)Y+(1-q)p\eta(Y-\lambda X)+q(1-p)\eta(X-\lambda Y) \ \forall\, q \in [0,1].$$

For a given p, let $\mathbf{q}^*(p) \equiv \{argmax_q U(q,p)\} \equiv \{argmax_q U(q(X,0)+(1-q)(0,Y)|p(X,0)+(1-p)(0,Y))\}$. The brackets indicate that $\mathbf{q}^*(p)$ may be a set. A mixture, $p \in [0,1]$, is PE if $p \in \mathbf{q}^*(p)$.

Note that

$$
\begin{aligned}
\frac{\partial U(q,p)}{\partial q} &= X - Y - p\eta(Y - \lambda X) + (1-p)\eta(X - \lambda Y) \\
&= (1+\eta)X - (1+\eta\lambda)Y - p\eta(1-\lambda)(Y+X)
\end{aligned}
$$

is constant for a given p, as $U(q,p)$ is linear in q. If $\frac{\partial U(q,p)}{\partial q} > (<)\ 0$, then it will attain a unique maximum $\mathbf{q}^*(p) = \{1\}(\{0\})$. As such, any strict mixtures, $p \in (0,1)$, for which $\frac{\partial U(q,p)}{\partial q} \neq 0$ cannot be PE. Note that this development implies that not exchanging with certainty, $p = 1$, will be PE if $\frac{\partial U(q,1)}{\partial q} \geq 0$, or

$$
\begin{aligned}
(1+\eta)X - (1+\eta\lambda)Y - \eta(1-\lambda)(Y+X) \geq 0, \\
X \geq \frac{(1+\eta)}{(1+\eta\lambda)}Y,
\end{aligned}
$$

which corresponds to the pure strategy threshold noted above, $\underline{X}_{B,PE}$. Similarly, exchanging with certainty, $p = 0$, will be PE if $\frac{\partial U(q,0)}{\partial q} \leq 0$, or

$$
\begin{aligned}
(1+\eta)X - (1+\eta\lambda)Y \leq 0 \\
X \leq \frac{(1+\eta\lambda)}{(1+\eta)}Y,
\end{aligned}
$$

which corresponds to the pure strategy threshold, $\overline{X}_{B,PE}$. For values of X such that

$$
\frac{(1+\eta)}{(1+\eta\lambda)}Y \leq X \leq \frac{(1+\eta\lambda)}{(1+\eta)}Y,
$$

$p = 1$ and $p = 0$ will be PE.

Strict mixtures, $p \in (0,1)$, for which $\frac{\partial U(q,p)}{\partial q} = 0$, $p \in \mathbf{q}^*(p)$, as all values of q, including $q = p$, attain the maximum. For each parameter constellation, X, Y, η, λ, if there exists a candidate mixture

$$p \in (0,1)\ s.t\ p = \frac{(1+\eta)X - (1+\eta\lambda)Y}{\eta(1-\lambda)(Y+X)}$$

such a p is PE. Note that there will be at most one strict mixture PE. This strict mixture will be a proper probability provided $\frac{(1+\eta)X-(1+\eta\lambda)Y}{\eta(1-\lambda)(Y+X)} \in (0,1)$. For such a proper mixture probability to exist for $\lambda > 1$, it must be that

$$\frac{(1+\eta)}{(1+\eta\lambda)}Y < X < \frac{(1+\eta\lambda)}{(1+\eta)}Y.$$

That is, if $\lambda > 1$, both pure strategies, $p = 0$ and $p = 1$, are PE, and the required preferences are strict, there will also be a strict mixture PE. In contrast, for such a proper probability mixture to exist for $\lambda < 1$, it must be that

$$\frac{(1+\eta\lambda)}{(1+\eta)}Y < X < \frac{(1+\eta)}{(1+\eta\lambda)}Y.$$

That is, if $\lambda < 1$, and neither pure strategy, $p = 0$ or $p = 1$, are PE, there will be a strict mixture PE.

Figure A.2 summarizes the PE considerations in Condition B recognizing the possibility of mixed strategy equilibria with the corresponding value of the mixture probability noted. In contrast to the pure strategy analysis of Figure A.1, for $\lambda < 1$ within the bounds $\frac{(1+\eta\lambda)}{(1+\eta)}Y < X < \frac{(1+\eta)}{(1+\eta\lambda)}Y$, there is now a mixed strategy PE. Further, for $\lambda > 1$ and $\frac{(1+\eta)}{(1+\eta\lambda)}Y < X < \frac{(1+\eta\lambda)}{(1+\eta)}Y$ there are three equilibria when accounting for potential mixtures.

Having elaborated the PE restrictions for Condition B, we proceed to Condition F. Condition F alters the choice set from $\mathcal{D}_B = \{p \in [0,1] : p(X,0) + (1-p)(0,Y)\}$ to $\mathcal{D}_F = \{p \in [0,0.5] : p(X,0) + (1-p)(0,Y)\}$. This alteration induces two potential changes to the PE calculus. First, potential PE choices from Condition B may not be available in Condition F.

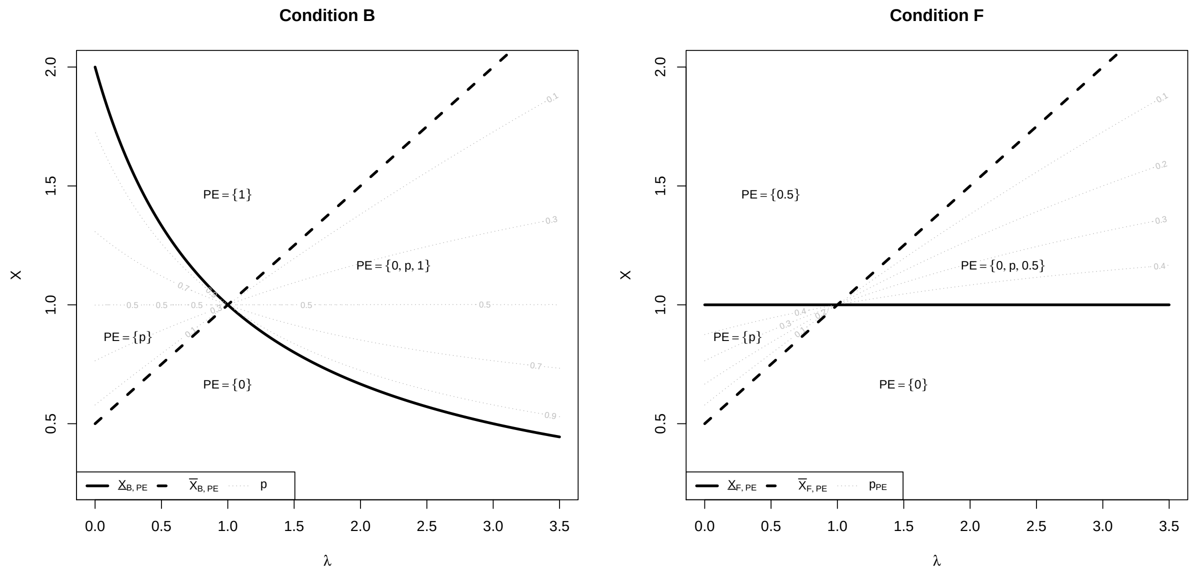


Figure A.2: **Gain-Loss Attitudes and Theoretical PE Strategy Thresholds**
Notes: Threshold values for mixed strategy PE for agent endowed with X, assuming $Y = 1$ and $\eta = 1$.

Second, lotteries, q, that prevent a specific p from being PE may potentially be eliminated.

In Condition F, a given mixture $p \in [0, 0.5]$ will be PE if

$$U(p(X,0)+(1-p)(0,Y)|p(X,0)+(1-p)(0,Y)) \geq$$
$$U(q(X,0)+(1-q)(0,Y)|p(X,0)+(1-p)(0,Y)) \; \forall \, q \; \in [0,0.5].$$

As before $U(q,p)$ is linear in q, and so a boundary strategy of attempting to keep one's object, $(p = 0.5)$ will be PE if

$$\begin{aligned} \frac{\partial U(q,0.5)}{\partial q} = (1+\eta)X - (1+\eta\lambda)Y - 0.5\eta(1-\lambda)(Y+X) \geq 0 \\ (1+0.5\eta(1+\lambda))X \geq (1+0.5\eta(1+\lambda))Y \\ X \geq Y, \end{aligned}$$

which corresponds to the pure strategy threshold, $\underline{X}_{F,PE}$. Similarly, exchanging with certainty,

$p = 0$, will be be PE if

$$\frac{\partial U(q,0)}{\partial q} = (1+\eta)X - (1+\eta\lambda)Y \leq 0$$
$$X \leq \frac{(1+\eta\lambda)}{(1+\eta)}Y,$$

which corresponds to the pure strategy threshold, $\overline{X}_{F,PE} = \overline{X}_{B,PE}$.

Again strict mixtures, $p \in (0, 0.5)$, for which $\frac{\partial U(q,p)}{\partial q} = 0$, $p \in \mathbf{q}^*(p)$, as all values of q, including $q = p$, attain the maximum. For each parameter constellation, X, Y, η, λ, if there exists a candidate mixture

$$p \in (0, 0.5) \; s.t \; p = \frac{(1+\eta)X - (1+\eta\lambda)Y}{\eta(1-\lambda)(Y+X)}$$

such a p is PE. Note that there will be at most one strict mixture PE. This strict mixture will be a proper probability and within the choice set provided $\frac{(1+\eta)X-(1+\eta\lambda)Y}{\eta(1-\lambda)(Y+X)} \in (0, 0.5)$. For such a proper mixture probability to exist for $\lambda > 1$, it must be that

$$Y < X < \frac{(1+\eta\lambda)}{(1+\eta)}Y$$

That is, if $\lambda > 1$, both pure strategies, $p = 0$ and $p = 0.5$, are PE, and the required preferences are strict, there will also be a strict mixture PE. In contrast, for such a proper probability mixture to exist for $\lambda < 1$, it must be that

$$\frac{(1+\eta\lambda)}{(1+\eta)}Y < X < Y.$$

That is, if $\lambda < 1$, and neither pure strategy, $p = 0$ or $p = 0.5$, are PE, there will be a strict mixture PE.

Figure A.2 summarizes the PE considerations in Condition F recognizing the possibility of mixed strategy equilibria with the corresponding value of the mixture probability noted. Moving from Condition B to Condition F all mixed strategy PE with $p \in (0.5, 1)$ are eliminated from the choice set. Individuals with $\lambda > 1$ and multiple equilibria, $PE = \{0, p > 0.5, 1\}$ in Condition B have a unique $PE = \{p = 0\}$ in Condition F. Such individuals may exchange less than 100 percent of the time in Condition B and do so 100 percent of the time in Condition F, growing more willing to exchange. In contrast, individuals with $\lambda < 1$ and a unique $PE = \{p > 0.5\}$ in Condition B, have a unique $PE = \{p = 0.5\}$ in Condition F. Such individuals would attempt to retain their object less than 100 percent of the time in Condition B and would do so 100 percent of the time in Condition F, growing less willing to exchange. This analysis highlights exactly the intuition laid out with our prior pure strategy analysis and that for the CPE concept. We next turn to PPE analysis to select among multiple PE selections.

Preferred Personal Equilibrium Analysis

Where there exist multiple PE selections, the KR model is equipped with an equilibrium selection mechanism, *Preferred Personal Equilibrium* (PPE). PPE selects among PE values on the basis of ex-ante utility. Having elaborated the PE values in the Figure A.2, it is straightforward to identify the selection, p, with the highest value of $U(p(X,0) + (1-p)(0,Y)|p(X,0) + (1-p)(0,Y)) = pX + (1-p)Y + p(1-p)\eta(1-\lambda)(X+Y)$. In the case of Condition B, there is a region of multiplicity for $\lambda > 1$ where the set of $PE = \{0, p \in (0,1), 1)\}$. In this region it is clear that not exchanging, $p = 1$, will yield higher ex-ante utility than exchanging ,$p = 0$, if

$$X > Y.$$

If $X > Y$, $p = 1$ will also yield higher ex-ante utility than any PE mixture $p \in (0,1)$ as all mixtures will both lower intrinsic utility (as $X > Y \to X > pX + (1-p)Y \forall p \in (0,1)$) and expose the individual to the overall negative sensations of gain loss embodied in the term $p(1-p)\eta(1-$

$\lambda)(X+Y) < 0$ for $\lambda > 1$. Following this logic, in Condition B, multiplicity is resolved via PPE by selecting either $p = 1$ if $X > Y$ or $p = 0$ if $X < Y$.

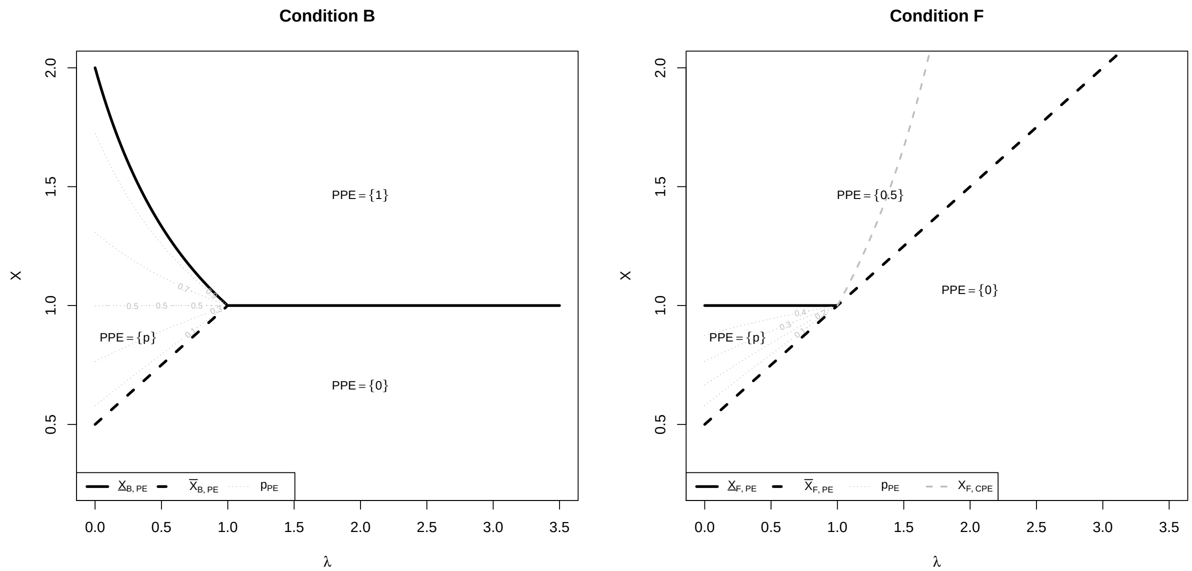


Figure A.3: **Gain-Loss Attitudes and Theoretical PPE Strategy Thresholds**
Notes: Threshold values for PPE for agent endowed with X, assuming $Y = 1$ and $\eta = 1$.

Similarly, in Condition F, there is a region of multiplicity for $\lambda > 1, Y < X < \frac{(1+\eta\lambda)}{(1+\eta)}Y$ where the set of $PE = \{0, p \in (0, 0.5), 0.5\}$. Note that for $\lambda > 1$, if $X < \frac{(1+\eta\lambda)}{(1+\eta)}Y$, then $X < \frac{(1+0.5\eta(\lambda-1))}{(1+0.5\eta(1-\lambda))}Y = \frac{(1+\eta\lambda-0.5\eta(\lambda+1))}{(1+\eta-0.5\eta(\lambda+1))}$. That is, in this region of multiplicity, X is below the $X_{F,CPE}$ cutoff noted in the main text. Hence, we know that exchanging, $p = 0$, yields higher ex-ante utility than attempting not to exchange, $p = 0.5$, in this region. It suffices to check which of the remaining PE selections $\{0, p = \frac{(1+\eta)X-(1+\eta\lambda)Y}{\eta(1-\lambda)(Y+X)} \in (0, 0.5)\}$ provide higher utility. For this key mixture,

$$p = \frac{(1+\eta)X-(1+\eta\lambda)Y}{\eta(1-\lambda)(Y+X)}$$
$$(1-p) = \frac{\eta(1-\lambda)(Y+X)}{\eta(1-\lambda)(Y+X)} - \frac{(1+\eta)X-(1+\eta\lambda)Y}{\eta(1-\lambda)(Y+X)}$$

The PPE selection will be $p = 0$ provided

$$
\begin{aligned}
Y &> pX + (1-p)Y + p(1-p)\eta(1-\lambda)(X+Y) \\
Y &> X + (1-p)\eta(1-\lambda)(X+Y) \\
Y &> X + \left[\frac{\eta(1-\lambda)(Y+X)}{\eta(1-\lambda)(Y+X)} - \frac{(1+\eta)X - (1+\eta\lambda)Y}{\eta(1-\lambda)(Y+X)}\right]\eta(1-\lambda)(X+Y) \\
Y &> X + [\eta(1-\lambda)(Y+X) - (1+\eta)X + (1+\eta\lambda)Y] \\
Y - (1+\eta\lambda)Y - \eta(1-\lambda)Y &> X + \eta(1-\lambda)(X) - (1+\eta)X \\
-\eta Y &> -\eta\lambda X \\
X &> \frac{1}{\lambda}Y,
\end{aligned}
$$

Which is satisfied as $X > Y$ and $\lambda > 1$ in this region.

Figure A.3 summarizes the PPE considerations in Conditions B and F recognizing the possibility of a mixed strategy PPE with the corresponding value of the mixture probability noted. Also graphed in Figure A.3 is the relevant CPE cutoff for $\lambda > 1$ in Condition F to reinforce both that in the region of multiplicity exchanging, $p = 0$, yields higher ex-ante utility than attempting not to exchange, $p = 0.5$, and that the restrictions on behavior differ meaningfully between CPE and PPE. Nonetheless, both solution concepts share the same directional comparative statics that individuals with $\lambda > 1$ should grow more willing to exchange moving from Condition B to Condition F, while individuals with $\lambda < 1$ should grow less-so.

A.3 Estimation Strategy

In this appendix, we provide the likelihood formulation for our mixed-logit methodology to estimate heterogeneity in gain-loss attitudes and utilities. There are three relative preference statements that subjects provide in Stage 1: relative wanting statements, relative liking statements, and hypothetical choice. Let $i = 1, ..., N$ represent the index for subjects, and let

$\{w, l, h\}$ represent the index of the three preference statements, referring to (w)anting, (l)iking, and (h)ypothetical choice, respectively. Let $w, l \in \{-1, 0, 1\}$ correspond to providing a higher rating for the alternative object, providing equal ratings for both objects, and providing a higher rating for the endowed object, respectively. Let $h \in \{-1, 1\}$ correspond to hypothetically choosing the alternative object or the endowed object, respectively.

We begin by presenting a standard logit formulation and then extend to the mixed logit case. Let $G(\cdot)$ represent the CDF of the logistic distribution. For each individual there are three potential probabilities associated with the three potential wanting ratings for those endowed with X, $Prob_{w_i,X}$,

$$\begin{aligned}
Prob_{w_i,X} &= G((1+\lambda) - 2\tfrac{Y}{X} - \delta_X) & if\ w_i = 1 \\
Prob_{w_i,X} &= G(2\tfrac{Y}{X} - (1+\lambda) - \delta_X) & if\ w_i = -1 \\
Prob_{w_i,X} &= 1 - G((1+\lambda) - 2\tfrac{Y}{X} - \delta_X) - G(2\tfrac{Y}{X} - (1+\lambda) - \delta_X) & if\ w_i = 0,
\end{aligned}$$

and three for those endowed with Y, $Prob_{w_i,Y}$,

$$\begin{aligned}
Prob_{w_i,Y} &= G(2 - (1+\lambda)\tfrac{Y}{X} - \delta_X) & if\ w_i = -1 \\
Prob_{w_i,Y} &= G((1+\lambda)\tfrac{Y}{X} - 2 - \delta_X) & if\ w_i = 1 \\
Prob_{w_i,Y} &= 1 - G(2 - (1+\lambda)\tfrac{Y}{X} - \delta_X) - G((1+\lambda)\tfrac{Y}{X} - 2 - \delta_X) & if\ w_i = 0.
\end{aligned}$$

Similarly, there are three potential probabilities associated with the three potential liking ratings for those endowed with X, $Prob_{l_i,X}$,

$$\begin{aligned}
Prob_{l_i,X} &= G((1+\lambda) - 2\tfrac{Y}{X} - \delta_X) & if\ l_i = 1 \\
Prob_{l_i,X} &= G(2\tfrac{Y}{X} - (1+\lambda) - \delta_X) & if\ l_i = -1 \\
Prob_{l_i,X} &= 1 - G((1+\lambda) - 2\tfrac{Y}{X} - \delta_X) - G(2\tfrac{Y}{X} - (1+\lambda) - \delta_X) & if\ l_i = 0,
\end{aligned}$$

and three for those endowed with Y, $Prob_{l_i,Y}$,

$$\begin{aligned}
Prob_{l_i,Y} = & \quad G(2-(1+\lambda)\tfrac{Y}{X}-\delta_X) & if\ l_i=-1 \\
Prob_{l_i,Y} = & \quad G((1+\lambda)\tfrac{Y}{X}-2-\delta_X) & if\ l_i=1 \\
Prob_{l_i,Y} = & \quad 1-G(2-(1+\lambda)\tfrac{Y}{X}-\delta_X)-G((1+\lambda)\tfrac{Y}{X}-2-\delta_X) & if\ l_i=0.
\end{aligned}$$

Lastly, there are two potential probabilities associated with the two hypothetical choice statements for those endowed with X $Prob_{h_i,X}$,

$$\begin{aligned}
Prob_{h_i,X} = & \quad G((1+\lambda)-2\tfrac{Y}{X}) & if\ w_i=1 \\
Prob_{h_i,X} = & \quad G(2\tfrac{Y}{X}-(1+\lambda)) & if\ w_i=-1,
\end{aligned}$$

and two for those endowed with Y, $Prob_{h_i,Y}$,

$$\begin{aligned}
Prob_{h_i,Y} = & \quad G(2-(1+\lambda)\tfrac{Y}{X}) & if\ w_i=-1 \\
Prob_{h_i,Y} = & \quad G((1+\lambda)\tfrac{Y}{X}-2) & if\ w_i=1.
\end{aligned}$$

Let $\mathbf{1}_X$ indicate an individual endowed with object X. A single individual's choice probability would thus be

$$L_i = (Prob_{w_i,X} \cdot Prob_{l_i,X} \cdot Prob_{h_i,X})^{\mathbf{1}_X} \cdot (Prob_{w_i,Y} \cdot Prob_{l_i,Y} \cdot Prob_{h_i,Y})^{(1-\mathbf{1}_X)},$$

and the grand log likelihood would be

$$\mathcal{L} = \sum_{i=1}^{N} log(L_i)$$

Moving from this logit formulation to our mixed logit formulation is straightforward and follows Train (2009). For estimating the heterogeneity of gain-loss attitudes, we assume that the value λ is drawn from a log-normal distribution with $log(\lambda) \sim N(\mu_\lambda, \sigma^2_\lambda)$. Let $\theta \equiv (\mu_\lambda, \sigma^2_\lambda)$, represent the parameters of this distribution, and let $f(\lambda|\theta)$ be the distribution of λ given these parameters. A single individual's choice probabilities are thus

$$L_i = \int L_i(\lambda) f(\lambda|\theta) d\lambda$$

where $L_i(\lambda)$ is the individual choice probability evaluated at a given draw of $f(\lambda|\theta)$. We construct these choice probabilities through simulation. Let $r = 1, ..., R$ represent simulations of λ from $f(\lambda|\theta)$ at a given set of parameters, θ. Let λ^r be the r^{th} simulant. We simulate L_i as

$$\check{L}_i = \frac{1}{R} \sum_{r=1}^{R} L_i(\lambda^r),$$

And these simulated probabilities replace the standard choice probabilities in the grand log likelihood to create a simulated log likelihood,

$$\mathcal{SL} = \sum_{i=1}^{N} log(\check{L}_i).$$

This simulated log likelihood is maximized to deliver estimates of μ_λ and σ^2_λ alongside the homogeneous utility ratio $\frac{X}{Y}$.

When considering heterogeneous utility, the exercise is analogous. We assume that the value $\frac{X}{Y}$ is drawn from a log-normal distribution with $log(\frac{X}{Y}) \sim N(\frac{X}{Y}, \sigma^2_{\frac{X}{Y}})$. Let $\theta' \equiv (\mu_{\frac{X}{Y}}, \sigma^2_{\frac{X}{Y}})$, represent the parameters of this distribution, and let $f(\frac{X}{Y}|\theta')$ be the distribution of $\frac{X}{Y}$ given these parameters. A single individual's choice probabilities are thus

$$L_i = \int L_i(\frac{X}{Y}) f(\frac{X}{Y}|\theta') d\frac{X}{Y}$$

where $L_i(\frac{X}{Y})$ is the individual choice probability evaluated at a given draw of $f(\frac{X}{Y}|\theta')$. We construct these choice probabilities through simulation. Let $r = 1, ..., R$ represent simulations of $\frac{X}{Y}$ from $f(\frac{X}{Y}|\theta')$ at a given set of parameters, θ'. Let $\frac{X}{Y}^r$ be the r^{th} simulant. We simulate L_i as

$$\check{L}_i = \frac{1}{R} \sum_{r=1}^{R} L_i(\frac{X}{Y}^r),$$

And these simulated probabilities replace the standard choice probabilities in the grand log likelihood to create a simulated log likelihood,

$$\mathcal{SL} = \sum_{i=1}^{N} log(\check{L}_i).$$

This simulated log likelihood is maximized to deliver estimates of $\mu_{\frac{X}{Y}}$ and $\sigma^2_{\frac{X}{Y}}$ alongside the homogeneous gain-loss parameter, λ.

Operationally for implementing both of our simulated likelihood techniques we use 1000 Halton draws for each heterogeneous parameter and implement the code in Stata. The code for our procedure estimating the distribution of gain-loss attitudes is presented below.

```
/* Estimator with MSL Portion for Distribution of Lambda */

capture program drop MSL_hetlambda
program define MSL_hetlambda
    * specifiy the arguments for the program
    args lnf l ratio12 ratio34 d12 d34 ln_sd

    * declare temporary variables
    tempvar choice choicetype endowed2 endowed3 endowed4 lambda delta firstval secondval sim_f sim_avef


    quietly {
        * initialize the data
        generate int `choice' = $ML_y1
        generate int `choicetype' = $ML_y2
        generate int `endowed2' = $ML_y3
        generate int `endowed3' = $ML_y4
        generate int `endowed4' = $ML_y5


        * initiate simulation average likelihood
        generate double `sim_avef' = 0

        * set seed equivalent to prior seed
        set seed 10101

        * simulate likelihood at each draw of lambda
        forvalues drawnum = 1/1000 {

            * draw lambda
            generate double `lambda' = exp(`l' + (exp(`ln_sd')*invnormal(draws1_`drawnum')))

            * evaluate the utilities
            generate double `firstval' = (1 + `lambda')
            replace `firstval' = (1 + `lambda')*`ratio12' if `endowed2' == 1
            replace `firstval' = (1 + `lambda') if `endowed3' == 1
            replace `firstval' =  (1 + `lambda')*`ratio34' if `endowed4' == 1

            generate double `secondval' =  2*`ratio12'
            replace  `secondval' =  2  if `endowed2' == 1
            replace  `secondval' = 2*`ratio34'  if `endowed3' == 1
            replace  `secondval' = 2  if `endowed4' == 1

            *indifference value
            generate double `delta' = exp(`d12')
            replace `delta' = exp(`d34')  if (`endowed3' == 1 | `endowed4' == 1)

            * construct simulated likelihood at current draw for ratings statements
            gen `sim_f' =   invlogit(`firstval' - `secondval' - `delta') if `choice' == 1  & (`choicetype' == 1 | `choicetype' == 2)
            replace `sim_f' = invlogit(`secondval' - `firstval' - `delta')  if `choice' == -1 & (`choicetype' == 1 | `choicetype' == 2)
            replace `sim_f' =  1- invlogit(`firstval' - `secondval' - `delta')  - invlogit(`secondval' - `firstval' - `delta')    if `choice' == 0 & (`choicetype' == 1 | `choicetype' == 2)


            * construct simulated likelihood  for hypothetical choice
            replace `sim_f' = invlogit(`firstval' - `secondval')   if `choice' == 1  & `choicetype' ==3
            replace `sim_f' =  1- invlogit(`firstval' - `secondval')  if `choice' == -1  & `choicetype' ==3

            *update average simulated likelihood
            replace `sim_avef' = `sim_avef' + (`sim_f'/1000)

            drop `lambda'  `firstval' `secondval' `sim_f' `delta'

            }

        * Establish log simulated likelihood

        replace `lnf' = ln(`sim_avef')

    }
end
```

A.4 Additional Tables

Table A.3: **Method of Simulated Likelihood Estimates: Sensitivity Analysis**

	(1)	(2)	(3)	(4)	(5)	(6)
	Estimate	(Std. Error)	Estimate	(Std. Error)	Estimate	(Std. Error)
	Heterogeneous λ		*Heterogeneous* λ		*Heterogeneous* λ	
Gain-Loss Attitudes:						
$\hat{\lambda}$	1.29	(0.04)	1.37	(0.08)	1.64	(0.21)
$\hat{\mu_{\lambda}}$	0.26	(0.03)	0.17	(0.07)	0.04	(0.08)
$\hat{\sigma^2_{\lambda}}$	0.00	(0.00)	0.29	(0.21)	0.91	(0.39)
Pair 1 Utilities (USB Stick (X) - Pen Set (Y)) *:*						
$\hat{\frac{Y}{X}}$ (Initial)	0.64	(0.03)	0.62	(0.04)	0.57	(0.04)
$\hat{\frac{Y}{X}}$ (Replication)	0.64	(0.04)	0.61	(0.04)	0.57	(0.05)
Pair 2 Utilities (Picnic Mat (X) - Thermos (Y))*:*						
$\hat{\frac{Y}{X}}$ (Initial)	1.10	(0.03)	1.11	(0.03)	1.13	(0.04)
$\hat{\frac{Y}{X}}$ (Replication)	0.90	(0.04)	0.88	(0.04)	0.87	(0.05)
Discernibility:						
δ_X	0.50	-	0.55	-	0.60	-
# Observations	3,072		3,072		3,072	

Notes: Maximum likelihood estimates. Standard errors in parentheses.

Table A.4: **Exchange Behavior and Probabilistic Forced Exchange, Clustered SE**

	Dependent Variable: Exchange (=1)		
	(1)	(2)	(3)
Condition F	-0.004	-0.340	-0.004
	(0.027)	(0.076)	(0.026)
$E[\lambda]$		-0.136	
		(0.036)	
Condition F * $E[\lambda]$		0.225	
		(0.046)	
Reduced Form Measure			-0.050
			(0.014)
Condition F * Reduced Form			0.077
			(0.018)
Constant (Condition B)	0.380	0.584	0.380
	(0.020)	(0.061)	(0.019)
R-Squared	0.000	0.017	0.014
# Observations	1024	1024	1024
# Clusters	53	53	53
H_0 : Zero Endowment Effect in B	$F_{1,52} = 34.96$	$F_{1,52} = 1.87$	$F_{1,52} = 38.26$
	$(p < 0.01)$	$(p = 0.18)$	$(p < 0.01)$
H_0 : Zero Treatment Effect (F-B)	$F_{1,52} = .02$	$F_{1,52} = 20.07$	$F_{1,52} = 0.02$
	$(p = 0.89)$	$(p < 0.01)$	$(p = 0.89)$
H_0 : Gain-Loss Attitudes $\perp$ Exchange in B		$F_{1,52} = 13.98$	$F_{1,52} = 13.19$
		$(p < 0.01)$	$(p < 0.01)$
H_0 : Gain-Loss Attitudes $\perp$ Treatment Effect		$F_{1,52} = 24.03$	$F_{1,52} = 19.48$
		$(p < 0.01)$	$(p < 0.01)$

Notes: Ordinary least square regression. Standard errors clustered at session level in parentheses. Null hypotheses tested for 1) zero baseline endowment effect regression (Constant coefficient = 0.5); 2) zero treatment effect (Condition F coefficient= 0); 3) no relationship between gain-loss attitudes and behavior in Condition B behavior ($E[\lambda]$ or Reduced Form Measure coefficient = 0); 4) constant treatment effect over gain-loss attitudes (Condition F * $E[\lambda]$ or Condition F * Reduced Form coefficient = 0). F-statistics and two-sided p-values reported.

Table A.5: **Stage 2 Behavior and Stage 1 Experience, Clustered SE**

	Dependent Variable: Exchange (=1)			
	Stage 1 Object Not Replaced		Stage 1 Object Replaced	
	(1)	(2)	(3)	(4)
Condition F	0.013	-0.255	-0.019	-0.418
	(0.035)	(0.120)	(0.044)	(0.124)
$E[\lambda]$		-0.121		-0.153
		(0.053)		(0.064)
Condition F * $E[\lambda]$		0.176		0.272
		(0.071)		(0.077)
Constant (Condition B)	0.386	0.569	0.374	0.600
	(0.027)	(0.092)	(0.032)	(0.104)
R-Squared	0.000	0.011	0.000	0.024
# Observations	511	511	513	513
# Clusters	53	53	53	53
H_0 : Zero Endowment Effect in B	$F_{1,52} = 17.82$	$F_{1,52} = 0.57$	$F_{1,52} = 15.78$	$F_{1,52} = 0.92$
	$(p < 0.01)$	$(p = 0.45)$	$(p < 0.01)$	$(p = 0.34)$
H_0 : Zero Treatment Effect (F-B)	$F_{1,52} = 0.13$	$F_{1,52} = 4.51$	$F_{1,52} = 0.18$	$F_{1,52} = 11.31$
	$(p = 0.72)$	$(p = 0.04)$	$(p = 0.67)$	$(p < 0.01)$
H_0 : Gain-Loss Attitudes $\perp$ Exchange in B		$F_{1,52} = 5.25$		$F_{1,52} = 5.81$
		$(p = 0.03)$		$(p = 0.02)$
H_0 : Gain-Loss Attitudes $\perp$ Treatment Effect		$F_{1,52} = 6.19$		$F_{1,52} = 12.62$
		$(p = 0.02)$		$(p < 0.01)$

Notes: Ordinary least square regression. Standard errors clustered at session level in parentheses. Null hypotheses tested for 1) zero baseline endowment effect regression (Constant coefficient = 0.5); 2) zero treatment effect (Condition F coefficient= 0); 3) no relationship between gain-loss attitudes and behavior in Condition B behavior ($E[\lambda]$ coefficient = 0); 4) constant treatment effect over gain-loss attitudes (Condition F * $E[\lambda]$ = 0). F-statistics and two-sided p-values reported.

Table A.6: Replication Consistency and Additional Controls, Clustered SE

	Dependent Variable: Exchange (=1)				
	Initial Study		Replication Study		
	(1)	(2)	(3)	(4)	(5)
Condition F	0.004	-0.409	-0.010	-0.239	-0.805
	(0.034)	(0.111)	(0.044)	(0.102)	(0.411)
$E[\lambda]$		-0.159		-0.103	-0.116
		(0.053)		(0.053)	(0.053)
Condition F * $E[\lambda]$		0.266		0.161	0.174
		(0.065)		(0.064)	(0.064)
Constant (Condition B)	0.365	0.616	0.399	0.542	0.917
	(0.028)	(0.093)	(0.030)	(0.081)	(0.343)
Additional Controls	No	No	No	No	Yes
Additional Interactions	No	No	No	No	Yes
R-Squared	0.000	0.023	0.000	0.008	0.060
# Observations	607	607	417	417	417
# Clusters	31	31	22	22	22
H_0 : Zero Endowment Effect in B	$F_{1,30} = 23.85$	$F_{1,30} = 1.53$	$F_{1,21} = 11.73$	$F_{1,21} = 0.26$	$F_{1,21} = 1.48$
	$(p < 0.01)$	$(p = 0.23)$	$(p < 0.01)$	$(p = 0.61)$	$(p = 0.24)$
H_0 : Zero Treatment Effect (F-B)	$F_{1,30} = 0.01$	$F_{1,30} = 13.44$	$F_{1,21} = 0.05$	$F_{1,21} = 5.51$	$F_{1,21} = 3.84$
	$(p = 0.90)$	$(p < 0.01)$	$(p = 0.82)$	$(p = 0.03)$	$(p = 0.06)$
H_0 : Gain-Loss Attitudes $\perp$ Exchange in B		$F_{1,30} = 9.09$		$F_{1,21} = 3.79$	$F_{1,21} = 4.78$
		$(p < 0.01)$		$(p = 0.07)$	$(p = 0.04)$
H_0 : Gain-Loss Attitudes $\perp$ Treatment Effect		$F_{1,30} = 16.61$		$F_{1,21} = 6.32$	$F_{1,21} = 7.47$
		$(p < 0.01)$		$(p < 0.01)$	$(p = 0.01)$

Notes: Ordinary least square regression. Standard errors clustered at session level in parentheses. Null hypotheses tested for 1) zero baseline endowment effect regression (Constant coefficient = 0.5); 2) zero treatment effect (Condition F coefficient= 0); 3) no relationship between gain-loss attitudes and behavior in Condition B behavior ($E[\lambda]$ = 0); 4) constant treatment effect over gain-loss attitudes (Condition F * $E[\lambda]$ = 0). The number of clusters in replication data does not permit test for effect of additional controls or interactions (all coefficients = 0), which would require. Additional controls include: gender, age, educational status, monthly income bracket, knowledge of economics, composite Raven matrices score, composite CRT score, and fixed effects for experimental assistant. Interactions include all controls interacted with Condition F. F-statistics and two-sided p-values reported.

A.5 Instructions and Material Presented to Participants

A.5.1 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 pair consisted of the USB stick and erasable pens, but this was counter-balanced at the session level.

Figure A.4: **Image 1 Projected on the Wall to Present Objects**
Notes: For Stage 1 with objects pair consisting of USB stick and erasable pens.

A.5.2 Original instructions in German (computer-based)

Willkommen in Teil 1 von 2 in diesem Experiment!

Part 2

Figure A.5: **Image 2 Projected on the Wall to Present Objects**
Notes: For Stage 2 with objects pair consisting of thermos and picnic mat.

Bitte schließen Sie den Vorhang Ihrer Kabine und lesen die folgenden Informationen. Alle Eingaben, die Sie in diesem Experiment am Computer machen, sind völlig anonym und können nicht mit Ihrer Person in Verbindung gebracht werden. Es geht an keiner Stelle in diesem Experiment um Schnelligkeit. Bitte nehmen Sie sich stets ausreichend Zeit, um die Anweisungen zu lesen und zu verstehen.

Sie besitzen nun das Produkt vor Ihnen. Sie können es jederzeit anfassen und inspizieren. Bitte öffnen Sie jedoch noch nicht die Verpackung und benutzen das Produkt nicht.

Die beiden Ihnen vorgestellten Produkte wurden zufällig und in gleichen Mengen auf die Kabinen verteilt. Ihre Kabinennummer hat sich ebenfalls rein zufällig aus der Wahl Ihres Sitzplatzes im Präsentationsraum ergeben.

Klicken Sie OK, wenn Sie diese Informationen gelesen haben. Falls Sie Fragen haben, rufen Sie bitte den Leiter des Experiments.

Bitte beantworten Sie die Fragen.

[USB stick / Thermoskanne]

Wie gut gefällt Ihnen das Produkt?

Wie gern würden Sie dieses Produkt mitnehmen?

[Radierbare Kugelschreiber / Picknick-Matte]

Wie gut gefällt Ihnen das Produkt?

Wie gern würden Sie dieses Produkt mitnehmen?

Wenn Sie sich für ein Produkt entscheiden müssten, welches würden Sie lieber behalten?

[USB stick / Thermoskanne] [Radierbare Kugelschreiber / Picknick-Matte

Bitte lesen Sie die folgenden Informationen aufmerksam.

Der Leiter des Experiments wird gleich mit einer Bingo-Trommel eine Zufallszahl zwischen 1 und 20 ziehen.Die gezogene Zahl wird danach laut durchgesagt. Wenn die gezogene Zahl eine Zahl [von 11 bis 20 / von 1 bis 10] ist, werden/wird [Ihr USB-Stick / Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] weggenommen und Sie erhalten stattdessen eine/einen [USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne]. Wenn die gezogene Zahl eine Zahl [von 1 bis 10 / von 11 bis 20] ist, behalten Sie [Ihr USB-Stick / Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] und es passiert nichts. Nachdem die Zahl gezogen wurde und gegebenenfalls ein Austausch der Produkte vollzogen wurde, passiert nichts mehr in diesem Teil des Experiments. Sie können das Produkt dann endgültig behalten.

Bitte bestätigen Sie erst, wenn Sie alles verstanden haben. Falls Sie Fragen haben, rufen Sie bitte den Leiter des Experiments und warten, bis er zu Ihnen kommt.

[Mood elicitation 1]

Bitte beantworten Sie die folgenden Fragen dazu, wie Sie sich gerade fülen. Welche Ausdrücke treffen auf Sie jetzt im Moment eher zu? Positionieren Sie den Schieberegler entsprechend. "Unglücklich, Wütend, Unzufrieden, Traurig, Verzweifelt" – "Glücklich, Begeistert, Zufrieden, Frühlich"

Es ist soweit! Bitte warten Sie, bis die Zahl gezogen wurde.

Zur Erinnerung: Wenn die gezogene Zahl [von 11 bis 20 / von 1 bis 10] ist, verlieren Sie [Ihr USB-Stick / Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] und erhalten stattdessen eine/einen [USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne]. Wenn die gezogene Zahl [von 1 bis 10 / von 11 bis 20] ist, behalten Sie [Ihr USB-Stick / Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne].

Die gezogene Zahl ist [1 / 2 / ... / 20].

Dies ist eine Zahl [von 1 bis 10 / von 11 bis 20]. Daher [verlieren Sie [Ihren USB-Stick / Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] und erhalten stattdessen eine/einen [USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne] / können Sie [Ihren USB-Stick"/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] behalten]. Bitte warten Sie, während der Leiter des Experiments den Austausch in den Kabinen durchführt.

[Mood elicitation 2 and control question.]

Bitte beantworten Sie die folgenden Fragen dazu, wie Sie sich gerade fülen. Welche Ausdrücke treffen auf Sie jetzt im Moment eher zu? Positionieren Sie den Schieberegler entsprechend. "Unglücklich, Wütend, Unzufrieden, Traurig, Verzweifelt" – "Glücklich, Begeistert, Zufrieden, Frühlich"

In der Lottoziehung die eben stattgefunden hat: Wie hoch war die Wahrscheinlichkeit (in Prozent), dass Sie Ihr ursprüngliches Produkt verlieren würden? Bitte geben Sie eine Zahl zwischen 0 und 100 ein. Please enter a number between 0 and 100.

Teil 1 des Experiments ist vorbei!

Bitte befolgen Sie die Anweisungen.

- Prägen Sie sich die Nummer Ihrer Kabine ein.
- Sie können jetzt zurück in den Präsentationsraum gehen.
- Bitte lassen Sie [Ihren USB-Stick/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] in der Kabine liegen. Sie werden in wenigen Minuten zurück in der gleichen Kabine sein.
- Zur Erinnerung: Das Produkt gehört nun endgültig Ihnen und Sie werden es mit aus dem Experiment nehmen.

Willkommen in Teil 2 in diesem Experiment!

Bitte schließen Sie den Vorhang Ihrer Kabine und lesen die folgenden Informationen.

Sie besitzen nun den/die [USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne] vor Ihnen. Sie können es jederzeit anfassen und inspizieren. Bitte öffnen Sie jedoch noch nicht die Verpackung und benutzen das Produkt nicht.

Die beiden für Teil 2 vorgestellten Produkte ([USB Stick und radierbare Kugelschreiber] / [Thermoskanne und Picknick-Matte]) wurden erneut zufällig und in gleichen Mengen auf die Kabinen verteilt.

Klicken Sie OK, wenn Sie diese Informationen gelesen haben. Falls Sie Fragen haben, rufen Sie bitte den Leiter des Experiments.

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

Bitte lesen Sie die folgenden Informationen aufmerksam. Der/Die [USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne] aus Teil 2 des Experiments gehört nun Ihnen und Sie können es behalten. Wenn Sie möchten, können Sie [Ihren USB-Stick/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] freiwillig gegen ein/eine [USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne] tauschen. Wie auch immer Sie sich entscheiden, Ihre Wahl ist endg¸ltig und Sie werden Ihr gewähltes Produkt danach mit aus dem Experiment nehmen.
Bitte bestätigen Sie erst, wenn Sie alles verstanden haben. Falls Sie Fragen haben, rufen Sie bitte den Leiter des Experiments und warten, bis er zu Ihnen kommt

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

Bitte lesen Sie die folgenden Informationen aufmerksam. Sie haben ein neues Produkt in Teil 2 des Experiments erhalten ([einen USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne]).

Sie erhalten gleich die Gelegenheit, [Ihren USB-Stick/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] freiwillig gegen [einen USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne] zu tauschen. Wenn Sie sich für einen Tausch entscheiden, erhalten Sie wie gewüscht [einen USB-Stick / radierbare

Kugelschreiber / eine Picknick-Matte / eine Thermoskanne] für [Ihren USB-Stick/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] und können [Ihren USB-Stick/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] endgültig behalten. Das Experiment ist damit abgeschlossen.

Wenn Sie sich gegen einen Tausch entscheiden, besteht danach eine Wahrscheinlichkeit von 50%, dass der Austausch dennoch erzwungen wird und sie trotzdem tauschen müssen.

Folgendes passiert konkret im Fall, dass Sie sich gegen einen freiwilligen Tausch entscheiden: Der Leiter des Experiments wird (wie in Teil 1 des Experiments) mit einer Bingo-Trommel eine Zufallszahl zwischen 1 und 20 ziehen. Die gezogene Zahl wird danach laut durchgesagt. Wenn die gezogene Zahl eine Zahl [von 11 bis 20 / von 1 bis 10] ist, wird/werden Ihnen [Ihr USB-Stick/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] weggenommen und Sie erhalten stattdessen [einen USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne]. Wenn die gezogene Zahl eine Zahl [von 1 bis 10 / von 11 bis 20] ist, behalten Sie [Ihren USB-Stick/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] und es passiert nichts. Nachdem die Zahl gezogen wurde und gegebenenfalls ein Austausch der Produkte vollzogen wurde, passiert nichts mehr in diesem Teil des Experiments. Sie können das Produkt dann endgültig behalten.

Bitte bestätigen Sie erst, wenn Sie alles verstanden haben. Falls Sie Fragen haben, rufen Sie bitte den Leiter des Experiments und warten, bis er zu Ihnen kommt

[Mood elicitation 3]

Bevor Sie die Möglichkeit erhalten, Ihr Produkt zu tauschen, beantworten Sie bitte die folgenden Fragen dazu, wie Sie sich gerade fülen. Welche Ausdrücke treffen auf Sie jetzt im Moment eher zu? Positionieren Sie den Schieberegler entsprechend. “Unglücklich, Wütend, Unzufrieden, Traurig, Verzweifelt” – “Glücklich, Begeistert, Zufrieden, Frühlich”

Möchten Sie [Ihren USB-Stick/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] gegen [einen USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne] tauschen?

Ja, ich möchte tauschen.

Nein, ich möchte nicht tauschen.

[ONLY BASELINE (p=0.0)]

Sie haben sich [für / gegen] einen freiwilligen Tausch enschieden. Bitte warten Sie, während der Leiter des Experiments den Austausch in den Kabinen durchführt.

[ONLY FORCED EXCHANGE (p=0.5)]

Sie haben sich [für / gegen] einen freiwilligen Tausch enschieden. Bitte warten Sie, während der Leiter des Experiments den Austausch in den Kabinen durchführt.

[ONLY NON-TRADERS] Danach entscheidet sich, ob Sie trotzdem tauschen müssen.

[ONLY TRADERS] Bitte warten Sie, bis das Experiment weitergeht. Es wird nun eine Zufallszahl für diejenigen gezogen, die sich gegen den freiwilligen Austausch entschieden haben. Danach geht das Experiment für Sie weiter.

[ONLY NON-TRADERS] Zur Erinnerung: Wenn die gezogene Zahl [von 11 bis 20 / von 1 bis 10] ist, verlieren Sie [Ihr USB-Stick"/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] und erhalten stattdessen eine/einen [USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne]. Wenn die gezogene Zahl [von 1 bis 10 / von 11 bis 20] ist, behalten Sie [Ihr USB-Stick"/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne].

[ONLY NON-TRADERS]

Die gezogene Zahl ist [1 / 2 / ... / 20].

Dies ist eine Zahl [von 1 bis 10 / von 11 bis 20]. Daher [verlieren Sie [Ihren USB-Stick / Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] und erhalten stattdessen eine/einen [USB-Stick / radierbare Kugelschreiber / eine Picknick-Matte / eine Thermoskanne] / können Sie [Ihren USB-Stick"/ Ihre radierbaren Kugelschreiber / Ihre Picknick-Matte / Ihre Thermoskanne] behalten]. Bitte warten Sie, während der Leiter des Experiments den Austausch in den Kabinen durchführt.

[Mood elicitation 4]

Bitte beantworten Sie die folgenden Fragen dazu, wie Sie sich gerade fülen. Welche Ausdrücke treffen auf Sie jetzt im Moment eher zu? Positionieren Sie den Schieberegler entsprechend. "Unglücklich, Wütend, Unzufrieden, Traurig, Verzweifelt" – "Glücklich, Begeistert, Zufrieden, Frühlich"

Das Experiment ist zu Ende!

Sie können beide Produkte behalten. Zudem erhalten Sie gleich eine Teilnahmevergütung von 4 Euro. Bitte warten Sie noch kurz in Ihrer Kabine, bis Sie der Experimentator herausruft. Vielen Dank für Ihre Teilnahme!

A.5.3 English translation of instructions

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.

You are currently in possession the product in front of you. You may touch it and inspect it anytime. However, please do not open the packaging and do not use the product 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?

If you had to choose one of the objects, which one would you prefer to keep?

[Erasable pens / picnic mat] [USB stick / thermos]

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 percent) 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. You are now also in possession of the [USB stick / erasable pens / thermos / picnic mat] in front of 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 percent 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”

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!

Appendix B

B.1 Stan Models

Our models contain a few assumptions beyond those discussed in the main text. In particular, we place bounds of $[0,9]$ for draws of λ_i, $[-1.5,1.5]$ for $\bar{\lambda}$, and $[-2,-0.1]$ for $\log(\sigma_\lambda)$. These work together to such that most of the draws of λ_i lie between (0,6), as larger values in CPE make little sense.[1] Moreover, we bound the intrinsic utils of the unknown good to be between $[0,12]$.

These constraints are shown in Figure B.1 and B.2, which provides screenshot of our *high-GL* and *low-GL* stan models.

B.2 Bayesian Estimator Workflow

We provide additional evidence for our Bayesian estimator, delving deeper into the prior predictive simulations and the synthetic recovery exercise.

[1] Already, a value of $\lambda = 3$ in CPE suggests that gain-loss utils are twice as important as the intrinsic utils to KR CPE utility.

```
data{
  int G;
  int group_size[G];
  int exchange[G];
  int endowment[G];
  int experiment_id[G];
  int Num_Mugs;
  real q[G];
  real utils_known[G];
  real utils_mean;
  real sd_utils;
  int permission[G];
}
parameters{
  vector<lower=0, upper = 9>[G] lambda;
  real<lower=-1.5, upper=1.5> lambda_bar;
  real<lower=-2,upper=-0.1> log_sig_lambda;
  vector<lower=0, upper=12>[Num_Mugs] utils_mug;
}

model{
  vector[G] kr_utils_mug;
  vector[G] kr_utils_money_pen;
  vector[G] p;
  utils_mug ~ normal(utils_mean, sd_utils);

  log_sig_lambda ~ normal( -0.65 , 0.15 );
  lambda_bar ~ normal( 0.25 , 0.25 );
  lambda ~ lognormal(lambda_bar , exp(log_sig_lambda));
  for ( i in 1:G ) {
    if(permission[i] == 1) {
    // Permission
    // CPE Version of utility given endowed mug
      if(endowment[i] == 1) {
         // U(Mug | Mug)
        kr_utils_mug[i]= utils_mug[experiment_id[i]];
         // U(Attempt Pen | Attempt Pen)
         kr_utils_money_pen[i]= (q[i] * utils_known[i] +
         (1 - q[i])*utils_mug[experiment_id[i]] +
        q[i] * (1 - q[i]) * (1 - lambda[i]) * (utils_mug[experiment_id[i]] + utils_known[i]));
        // Probability of Exchange
        p[i] = inv_logit((kr_utils_money_pen[i] - kr_utils_mug[i]));
     }
    // CPE Version of utility given endowed Pen
      if(endowment[i] == 2) {
        // U(Attempt Mug | Attempt Mug)
        kr_utils_mug[i]= (q[i] * utils_mug[experiment_id[i]] +
        (1 - q[i]) * utils_known[i] +
          q[i] * (1 - q[i]) * (1 - lambda[i]) * (utils_mug[experiment_id[i]] + utils_known[i]));
        // U(Pen | Pen)
        kr_utils_money_pen[i]= utils_known[i];
        // Probability of Exchange
        p[i] = inv_logit((kr_utils_mug[i] - kr_utils_money_pen[i]));
      }
    }

    if(permission[i] == 0) {
      // CPE Version of utility given endowed mug
      if(endowment[i] == 1) {
         // U(Mug | Mug)
        kr_utils_mug[i]= (q[i] * utils_known[i] +
        (1 - q[i])*utils_mug[experiment_id[i]] +
        q[i] * (1 - q[i]) * (1 - lambda[i]) * (utils_mug[experiment_id[i]] + utils_known[i]));
        // U(Attempt Pen | Attempt Pen)
        kr_utils_money_pen[i]= utils_known[i];
        // Probability of Exchange
        p[i] = inv_logit((kr_utils_money_pen[i] - kr_utils_mug[i]));
      }
      // CPE Version of utility given endowed Pen
      if(endowment[i] == 2) {
       // U(Attempt Mug | Attempt Mug)
        kr_utils_mug[i]= utils_mug[experiment_id[i]];
        // U(Pen | Pen)
        kr_utils_money_pen[i]= (q[i] * utils_mug[experiment_id[i]] +
        (1 - q[i]) * utils_known[i] +
         q[i] * (1 - q[i]) * (1 - lambda[i]) * (utils_mug[experiment_id[i]] + utils_known[i]));
        // Probability of Exchange
        p[i] = inv_logit((kr_utils_mug[i] - kr_utils_money_pen[i]));
    }
  }

  }
  exchange ~ binomial(group_size, p);
}
```

Figure B.1: **Stan Code, *High-GL* Model**

```
data{
  int G;
  int group_size[G];
  int exchange[G];
  int endowment[G];
  int experiment_id[G];
  int Num_Mugs;
  real q[G];
  real utils_known[G];
  real utils_mean;
  real sd_utils;
  int permission[G];
}
parameters{
  vector<lower=0, upper = 9>[G] lambda;
  real<lower=-1.5, upper=1.5> lambda_bar;
  real<lower=-2,upper=-0.1> log_sig_lambda;
  vector<lower=0, upper=12>[Num_Mugs] utils_mug;
}

model{
  vector[G] kr_utils_mug;
  vector[G] kr_utils_money_pen;
  vector[G] p;
  utils_mug ~ normal(utils_mean, sd_utils);

  log_sig_lambda ~ normal( -0.85 , 0.15 );
  lambda_bar ~ normal( 0.75 , 0.25 );
  lambda ~ lognormal(lambda_bar , exp(log_sig_lambda));
  for ( i in 1:G ) {
    if(permission[i] == 1) {
    // Permission
    // CPE Version of utility given endowed mug
      if(endowment[i] == 1) {
         // U(Mug | Mug)
        kr_utils_mug[i]= utils_mug[experiment_id[i]];
         // U(Attempt Pen | Attempt Pen)
         kr_utils_money_pen[i]= (q[i] * utils_known[i] +
         (1 - q[i])*utils_mug[experiment_id[i]] +
        q[i] * (1 - q[i]) * (1 - lambda[i]) * (utils_mug[experiment_id[i]] + utils_known[i]));
        // Probability of Exchange
        p[i] = inv_logit((kr_utils_money_pen[i] - kr_utils_mug[i]));
     }
    // CPE Version of utility given endowed Pen
      if(endowment[i] == 2) {
        // U(Attempt Mug | Attempt Mug)
        kr_utils_mug[i]= (q[i] * utils_mug[experiment_id[i]] +
        (1 - q[i]) * utils_known[i] +
          q[i] * (1 - q[i]) * (1 - lambda[i]) * (utils_mug[experiment_id[i]] + utils_known[i]));
        // U(Pen | Pen)
        kr_utils_money_pen[i]= utils_known[i];
        // Probability of Exchange
        p[i] = inv_logit((kr_utils_mug[i] - kr_utils_money_pen[i]));
      }
    }

    if(permission[i] == 0) {
      // CPE Version of utility given endowed mug
      if(endowment[i] == 1) {
         // U(Mug | Mug)
        kr_utils_mug[i]= (q[i] * utils_known[i] +
        (1 - q[i])*utils_mug[experiment_id[i]] +
        q[i] * (1 - q[i]) * (1 - lambda[i]) * (utils_mug[experiment_id[i]] + utils_known[i]));
        // U(Attempt Pen | Attempt Pen)
        kr_utils_money_pen[i]= utils_known[i];
        // Probability of Exchange
        p[i] = inv_logit((kr_utils_money_pen[i] - kr_utils_mug[i]));
      }
      // CPE Version of utility given endowed Pen
      if(endowment[i] == 2) {
       // U(Attempt Mug | Attempt Mug)
        kr_utils_mug[i]= utils_mug[experiment_id[i]];
        // U(Pen | Pen)
        kr_utils_money_pen[i]= (q[i] * utils_mug[experiment_id[i]] +
        (1 - q[i]) * utils_known[i] +
         q[i] * (1 - q[i]) * (1 - lambda[i]) * (utils_mug[experiment_id[i]] + utils_known[i]));
        // Probability of Exchange
        p[i] = inv_logit((kr_utils_mug[i] - kr_utils_money_pen[i]));
    }
  }

  }
  exchange ~ binomial(group_size, p);
}
```

Figure B.2: **Stan Code, *Low-GL* Model**

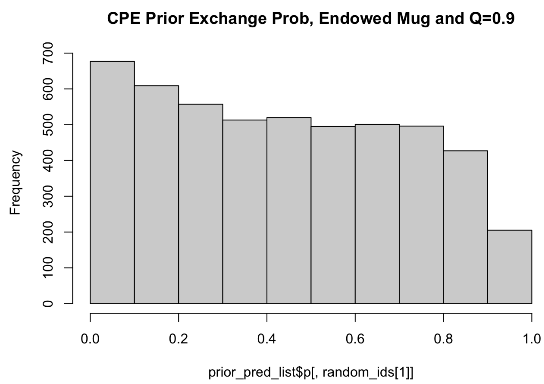
B.2.1 Prior Predictive Simulations

Priors can have important regularizing properties for Bayesian estimators; for our model, the implications of our priors are particularly hard to understand because they are latent variables that are subsequently fed into the KR CPE machinery. In order to understand how our priors on the distribution of λ affect the outcome – the probability that an individual voluntarily exchanges – we draw what are known as prior predictive samples. As implied by the name, we simply provide our estimator with our priors (and no data) and draw samples of our relevant variables. We proceed by conducting this exercise on both the *high-GL* and *low-GL* models.

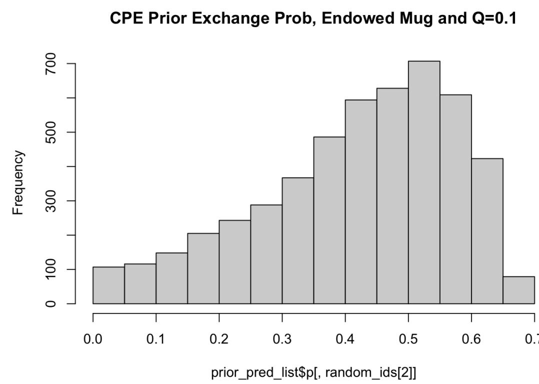
High GL Prior

Implied exchange probabilities are displayed in Figure B.3. Under the condition with high exchange permission ($q = 0.9$), where exchange is predominantly dictated by the intrinsic utils, we see the model predicts nearly uniform exchange probabilities from 0 to 1. When considering the low exchange permission condition of $q = 0.1$, gain-loss preferences play a stronger role in the CPE framework, and exchange is skewed left; thus, our prior suggests it's unlikely to see extreme exchange probabilities in this condition, and extremely unlikely to see exchange probabilities of 1.

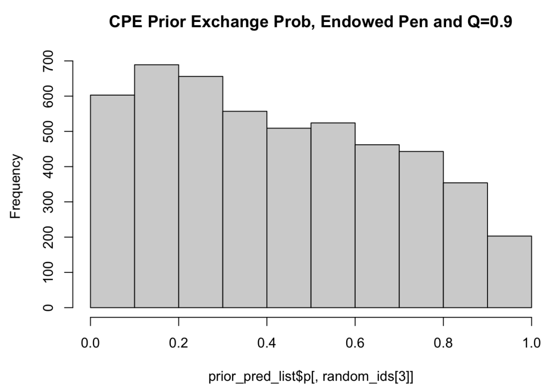
Since our data consists of one choice per individual, observed exchange probabilities per person will definitionally be 0% or 100%. However, our implied prior rules out exchange probabilities of 100%, due to the mechanics of KR CPE and the underlying latent variables. This will help regularize the latent parameters away from extreme values that would lead to 100% exchange (e.g., extreme draws of relative utility, so that one good is always preferred, or extreme λ so that exchange is always favored regardless of intrinsic utilities). We view this as a feature of our model – we don't over-extrapolate from a single coin flip landing heads and update towards a 100% chance of heads, but still allow the model to update above 50%.



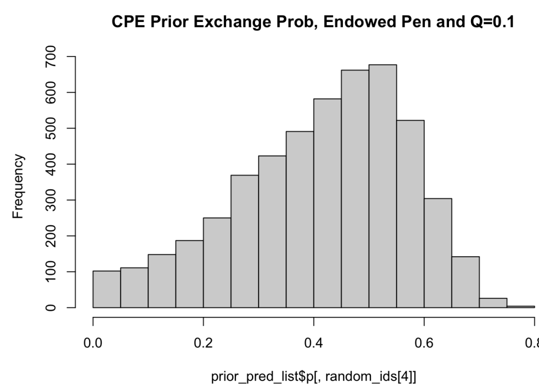
(a) Endowed Mug, 90% Permission



(b) Endowed Mug, 10% Permission

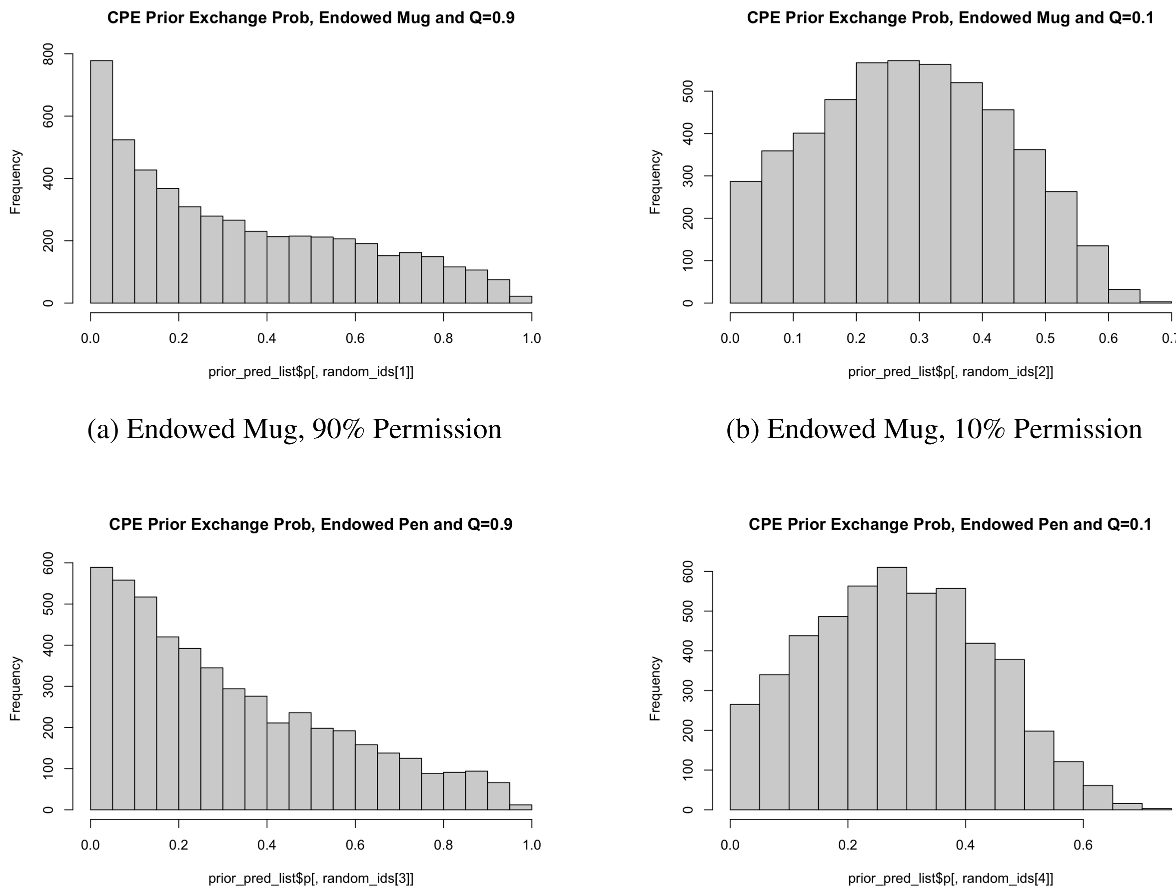


(c) Endowed Pen, 90% Permission



(d) Endowed Pen, 10% Permission

Figure B.3: **Prior Exchange Probabilities by Endowment and Treatment, *High-GL* Model**



(a) Endowed Mug, 90% Permission (b) Endowed Mug, 10% Permission

(c) Endowed Pen, 90% Permission (d) Endowed Pen, 10% Permission

Figure B.4: **Prior Exchange Probabilities by Endowment and Treatment, *Low-GL* Model**

Low GL Prior

The increase in value for the underlying λ_i under our *low-GL* prior has the consequence of shifting the implied exchange priors towards 0; this follows intuitively from KR CPE, since loss averse agents in exchange environments (where permission bites) tend to dislike relinquishing their good. Nonetheless, the analogous exercise conducted on these implied priors suggests this model makes sensible predictions.

B.2.2 Assumption on Intrinsic Utils

Because of the constraints within the existing data, we discuss our assumption for identifying the ratio of intrinsic utilities. Specifically, we opt to fix one of the object's intrinsic utility to 5. To motivate this assumption as relatively innocuous, we refer back to our structural set up derived from the discrete choice literature. There, an agent in the permission design attempts to trade if

$$U_i(q(0,Y)+(1-q)(X,0)|q(0,Y)+(1-q)(X,0))+\varepsilon_1 > U_i(X,0|X,0)+\varepsilon_2,$$

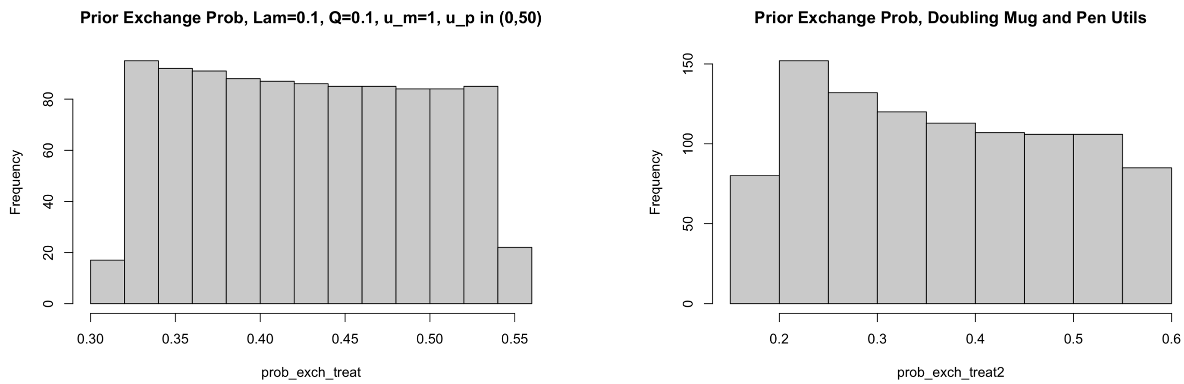
with ε_i distributed as Type I Extreme Value. Thus, the probability of exchange is

$$p = logit^{-1}[qm_Y+(1-q)m_X+q(1-q)\eta(1-\lambda)(m_X+m_Y)-m_X].$$

Implicit in this discrete choice formulation is an assumption on the scale of our random variables (ie $\nu \sim \text{Logistic}(0,1)$, which arises from the location and scale assumed on the Type I EV ε_i). However, under the KR specification, these assumptions on the noise terms interact considerably with the consumption utility values.

To see this, consider Figure B.5 below, which shows the admissible exchange probabilities under our model supposing the standard Logistic noise and fixing the utility of one object to 1. Here, $p_i = logit^{-1}(q[m_X-m_Y+(1-q)(1-\lambda)(m_X+m_Y)])$ is bounded above at around 0.55 (and below near 0.3), which suggests our model rules out exchange probabilities above 0.55 (below 0.3) in this particular condition ($q=0.1$, endowed Pen) – a rather strict belief. Of course, these bounds are a function of the values of m_X, m_Y, and the scale of our logistic noise.

Recall that the probability of exchange is just the CDF of the logistic ($F(x,\mu,s) = \frac{1}{1+\exp(-(x-\mu)/s)}$) observed at some difference in KR utilities. If our Logistic distribution has a smaller variance (s) but the same utility (and thus the same argument, x), the value of the CDF



(a) Implied Exchange, Mug Utils = 1 (b) Implied Exchange, Rescale Utils

Figure B.5: **Prior Exchange Probabilities Identification**

would be larger, implying larger exchange probabilities; alternatively, if we fix the variance but increased m_X and m_Y by a constant, $R > 1$, we would similarly increase the range of admissible exchange probabilities. This is shown in the second panel, where we factor out a 2 from doubling the utilities of each good.

Since intrinsic utility values are not experimentally observed for some of the data, and we can only identify the relative utils (the ratio of the consumption utilities), the symmetry noted above allows us some flexibility. Specifically, we can choose to manipulate the priors by imposing restrictions on the scale of the Logistic noise, or simply multiply the identified ratio of intrinsic utility by a constant – both of which helps us tune our priors on p_i in the same way. Thus, fixing the intrinsic utility of a good to 5 is mathematically equivalent to taking the scale of the Logistic to be $1/5$ the standard Logistic and identifying the ratio of utilities. We opted for the value 5 because the implied priors seemed the most reasonable to us.

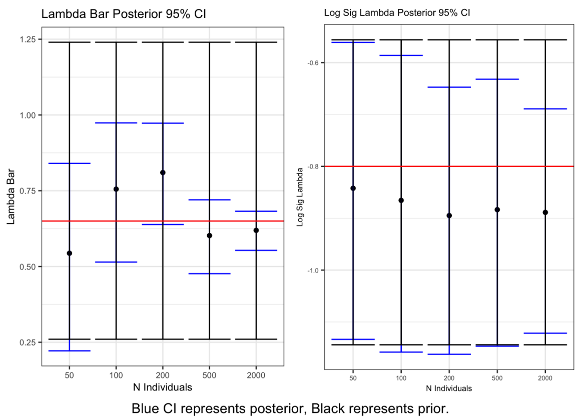
B.2.3 Data Generation Process

A key component of verifying our model's performance is a functional data generating process which is able to replicate the structure of the data we will be analyzing. Following our discrete choice model presented in section 2.2.2, we devise a a series of functions to generate individual level data analogous to Ericson and Fuster (2011) and Heffetz and List (2014). We assume that λ_i is drawn from a LogNormal population distribution (with location and scale as inputs), and that the two goods under consideration (Mugs, Pens) have intrinsic utility (5, 4.4) for all individuals. Endowments and treatments are assigned with 50% chance respectively (independent). Given these values, we compute the KR utilities of the two actions (attempt to exchange, keep), generate a Logistic noise term per individual and decision, and determine whether the synthetic individual attempts to exchange based on the relevant comparison of KR CPE utils and the Logistic noise.[2]

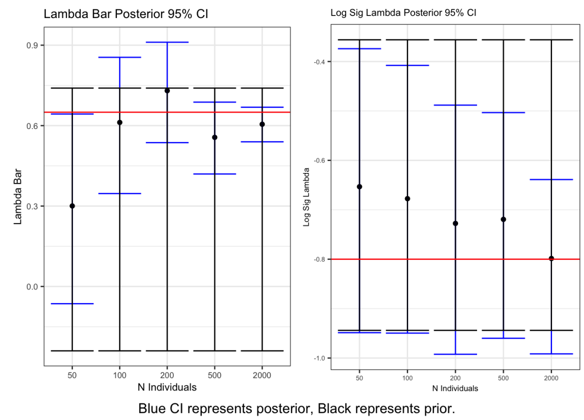
The DGP for Cerulli-Harms et al. (2019) is quite similar, with a few specialized differences. In particular, the goods under consideration are money and mugs, and the probabilistic forced exchange varies between $t \in \{0, 0.25, 0.5, 0.75\}$. Fixing mug utils – which the model will estimate – and a single price offer (e.g. \$6), we compute the KR utils of each possible decision under CPE, and map these to a noisy decision about whether to attempt to keep the endowment or exchange.

For each paradigm, we generate datasets under our permission-to-exchange DGP fixing $N = \{50, 100, 200, 500, 2000\}$. This provides us with a straightforward way to identify the effects of sample size (when each individual makes just one choice) on the identification of the latent λ distribution. We generate two such DGPs – one to create a dataset where the true distribution of gain-loss preferences follows $\lambda \sim LogNormal(0.17, \exp(-0.62)^2)$ – which yields a large fraction of gain-lovers – and a second where $\lambda \sim LogNormal(0.65, \exp(-0.8)^2)$, with few gain-lovers.

[2] The functions are created to allow for multiple decisions per individual. That is, holding the same endowment and treatment, suppose we observed the individual making the decision D times – we would be interested in how many times they decided to exchange. In this set up, a new draw of Logistic noise is taken for each of the individuals in their D decisions. To match Ericson and Fuster (2011) and Heffetz and List (2014), we set $D = 1$ here.



(a) Low-GL Prior



(b) High-GL Prior

Figure B.6: **Posterior and Prior 95% Credible Intervals, Low-GL DGP**

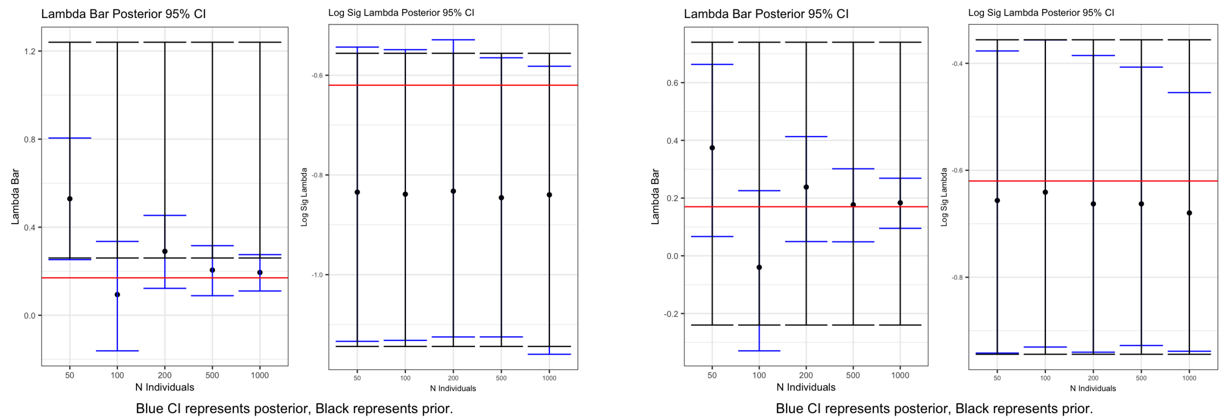
B.2.4 Synthetic Recovery under Permission Design: Low-GL DGP

We augment our synthetic recovery exercise in the main text by ensuring our results are not specifc to the *high-GL* DGP. Using the *low-GL* DGP, we supply our model with a prior quite close to the true DGP as well as the more gain-loving *high-GL* prior to verify that our estimator performs robustly across DGPs. The convergence plots suggest that our estimator is able to recover both hyper-parameters quite well, with posterior uncertainty over $\bar{\lambda}$ decreasing as N increases.

B.2.5 Synthetic Recovery: Probabilistic Forced Exchange Design

The main text focused on a synthetic recovery exercise using simulated data under the permission-to-exchange paradigm. In this section, we instead simulate data from the probabilistic forced exchange paradigm, basing our DGP on Cerulli-Harms et al. (2019). In particular, we fix the mug utilities to a homogenous number across the population, and examine the exchange/keep decision for a single price offering under the logit choice specification.

Note that we allow an additional degree of freedom in this exercise compared to the permission paradigm – namely, we allow estimates of the intrinsic utils of the mug to differ between



(a) Low-GL Prior (b) High-GL Prior

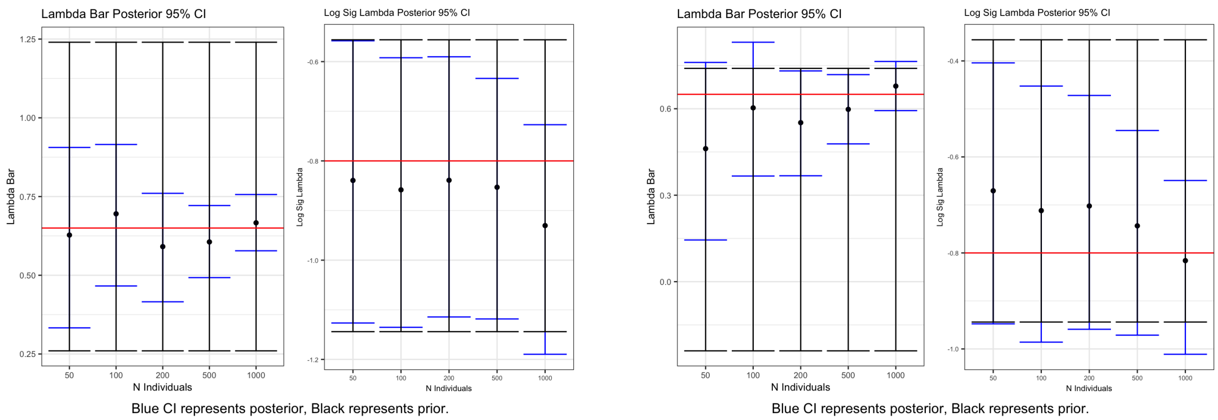
Figure B.7: **Posterior and Prior 95% Credible Intervals, High-GL DGP (PFE)**

buyers and sellers. Our results in the synthetic recovery exercise are robust to this assumption. We include it here because the assumption is required to overcome a large endowment effect in the Cerulli-Harms et al. (2019) data, which cannot be rationalized under KR CPE without a population full of gain-lovers. By introducing this degree of freedom, we identify the population distribution of gain-loss preferences in Cerulli-Harms et al. (2019) by relying on the relative proportion of exchange as the forced exchange probability varies.

Figures B.7 and B.8 demonstrate that our synthetic recovery results under the probabilistic forced exchange paradigm are analogous to those under the permission-to-exchange design; importantly, our estimator is once again able to overcome misspecified priors under either of the two DGPs.

B.3 Results by Session

In the main text, we presented two key figures for our results: the estimated posterior proportion of gain-lovers per experiment, and the sequential posteriors as we add in newer data. In this section, we discuss each experiment in more detail, highlighting the session by session results. Because Ericson and Fuster (2011) only contains one session format, it is omitted from



(a) Low-GL Prior　　(b) High-GL Prior

Figure B.8: **Posterior and Prior 95% Credible Intervals, Low-GL DGP (PFE)**

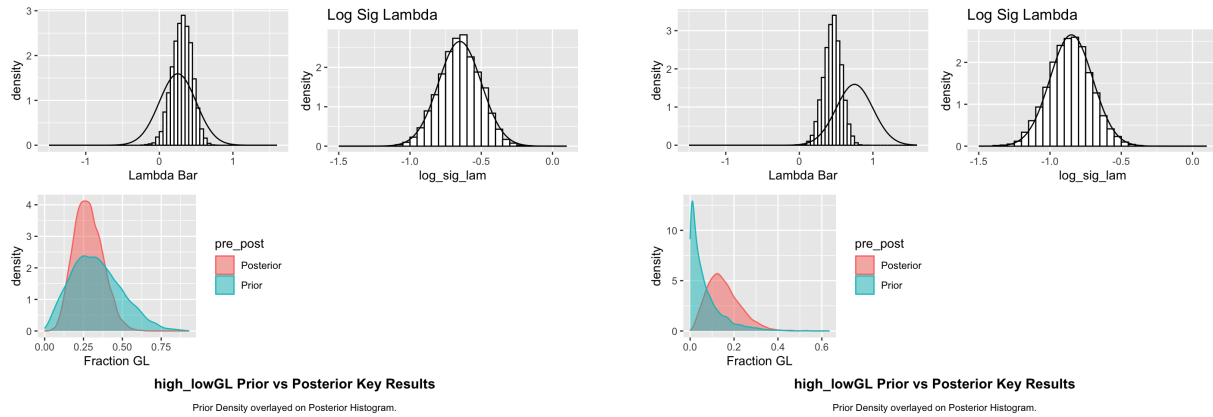
this section.

B.3.1 HL Data

HL 1

The first session we analyze is data from the **More Endowment** session, consisting of 117 participants. The key parameter of this study (besides endowment and permission probabilities) is stronger language in the instructions (resembling those in Ericson and Fuster (2011)). For instance, there is a greater emphasis on possession of the endowment – “The item you own is yours to keep. You own it for real, not just for the purpose of the study”, and the instructions elicit “keep or trade” decisions as opposed to “mug or pen”. Among the 117 participants, there is a treatment effect of roughly 13%, providing suggestive evidence in favor of a more loss averse population.

Regardless of the prior, the posterior distribution of $\bar{\lambda}$ in this study hovers close to a mean of 0.4, roughly splitting the difference between our *low-GL* and *high-GL* priors. Once again, the posterior of $\log(\sigma_\lambda)$ remains essentially the same as the prior. The overall effect of the data on these hyper-parameters leads to different interpretations based on the priors – likely a result of a



(a) High-GL Model

(b) Low-GL Model

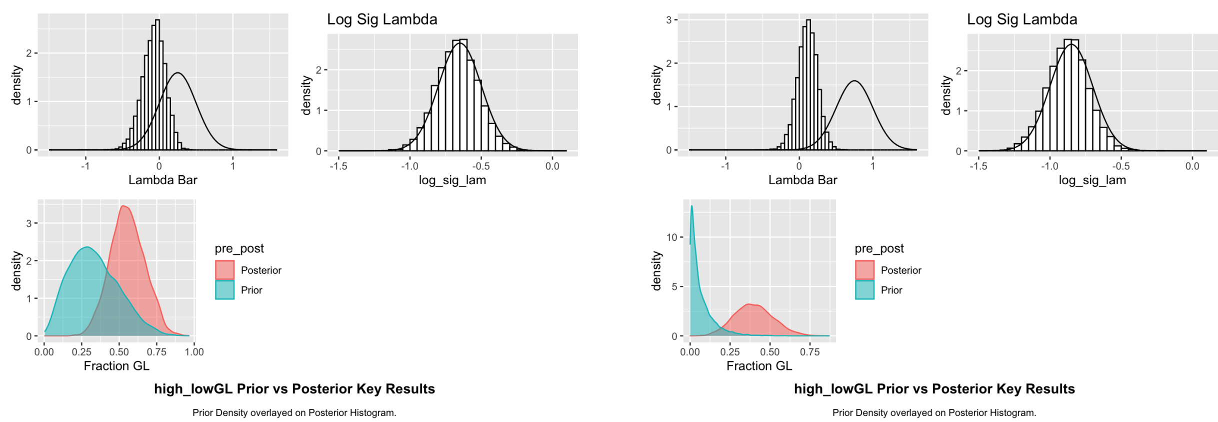
Figure B.9: **Posterior Results, HL 1 Data**

relatively small sample. Under the *high-GL* prior, the estimator suggests that the data reduces our uncertainty, ruling out extreme fractions of gain-lovers on each side but maintaining the same peak at 25%. The data combined with the *low-GL* prior, however, suggests that the posterior population is relatively more gain-loving than assumed; hovering around 7.5% to 22.5%, this posterior fraction of gain-lovers is still smaller than under the *high-GL* posterior. Thus, the data from this session suggests a relatively heterogeneous distribution of gain-loss preferences.

HL 2

For the second session analyzed, the authors shy away from the strong ownership language and focus on what they call the **Less Endowment** experiment. Once again, this predominantly consists of language choices throughout the instructions, avoiding words that evoke ownership of the endowed object. This experiment contains a sample of 116 participants, with an average treatment effect of -8% – suggestive of a sizable fraction of gain-lovers.

Regardless of our prior, the data seems to move our posterior distribution on $\bar{\lambda}$ to a mean below 0.2 – smaller than our *high-GL* prior. Under the *high-GL* prior, our estimator updates so that the estimated percentage of gain-lovers lies between 37.5% and 75%, with a tight peak around 55%. Assuming our *low-GL* prior instead yields a fairly flat posterior fraction, with a



(a) High-GL Model

(b) Low-GL Model

Figure B.10: **Posterior Results, HL 2 Data**

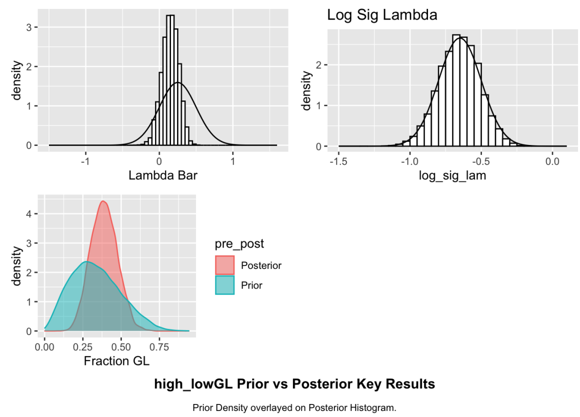
peak around 35% but substantial mass between 20% and 50% gain lovers.

HL 3

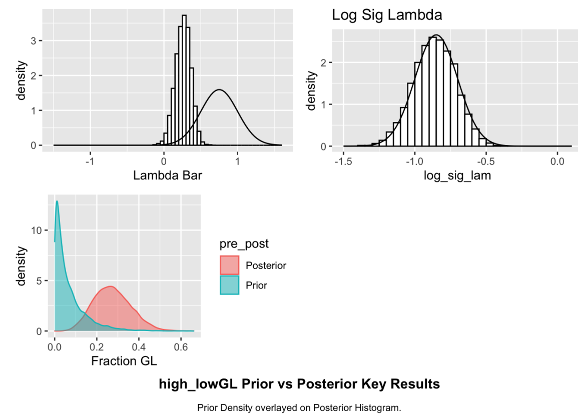
The third study differs from the first two in terms of the randomization procedure. Experimental sessions 1 and 2 had participants flip a coin to determine their assigned good, but had pre-randomized the permission probabilities within the instructions. In this session, the authors provide transparency in both processes, allowing a coin flip for both the endowment and the permission probability. The instructions are identical to the **More Endowment** experiment otherwise, excepting a number of conditional statements (one set of instructions for each of the possible permission to exchange outcomes). This study is substantially larger than the first two, with a total of 225 participants, and has an average treatment effect of about 4%.

Our posterior estimates with this experimental data are very similar to that of the **More Endowment** data. However, with the additional 100 subjects, we see a decrease in reliance on the prior for $\bar{\lambda}$. While the histogram of $\bar{\lambda}$ is slightly rightward shifted under the *low-GL* prior, the two are very similar. Once more, though, the posterior of $\log(\sigma_\lambda)$ changes very little relative to the respective priors.

With regards to the posterior fraction of gain-lovers, we do see variation by prior at-



(a) High-GL Model



(b) Low-GL Model

Figure B.11: **Posterior Results, HL 3 Data**

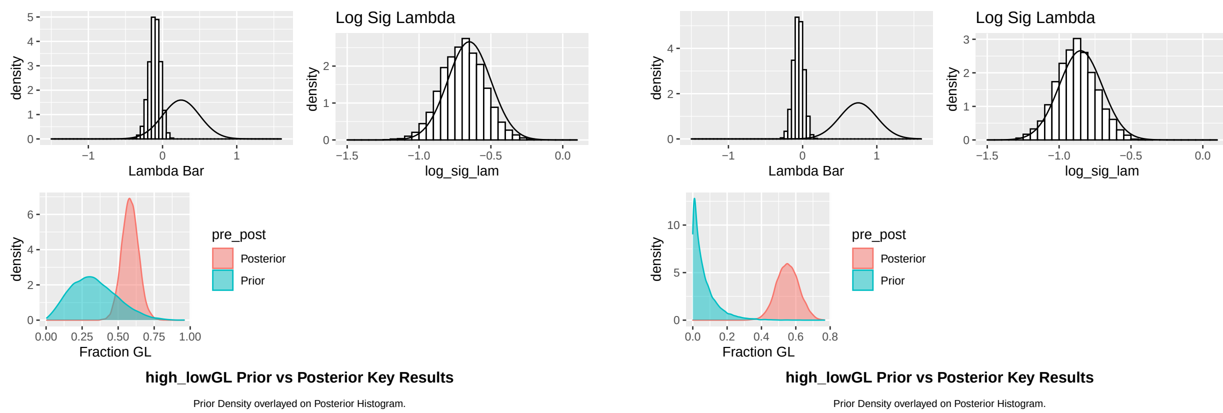
tributable to the differences in $\log(\sigma_\lambda)$. Nevertheless, the data suggests that it is very likely more than 15% of the sample are gain-lovers, and we shouldn't rule out a fraction as high as 50%.

B.3.2 CGS

CGS 1

For the main experiment, subjects were physically split into buyers and sellers – each placed on separate sides of a lecture hall. The students then began to read the instructions while the sellers were subsequently endowed with mugs and buyers with CHF 10. In the instructions, the participants were informed about the market mechanisms – namely that the market price will be determined by the intersection of supply and demand of all buyers and sellers – as well as the forced exchange mechanism. Then, each individual made choices by deciding whether they would prefer a mug or the market price for each price from CHF 0 to CHF 10 (increments of CHF 0.5).

In this session, the median price for buyers (sellers) in the 50% forced exchange condition was CHF 4 (CHF 7). This large endowment effect persisted through each of the forced exchange conditions in the study, and the changes in WTP/WTA across these conditions is often consistent



(a) High-GL Model (b) Low-GL Model

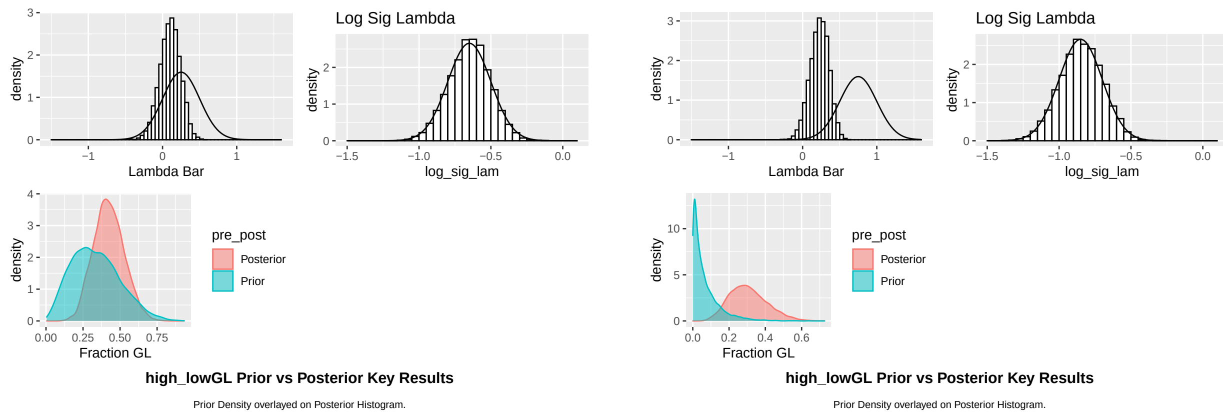
Figure B.12: **Posterior Results, CGS 1 Data**

with theoretical results for gain-lovers. As such, we expect this data to result in a highly gain-loving posterior distribution, which we indeed find in Figure B.12. Regardless of prior, the posterior distribution of gain-loss preferences skews quite gain-loving – with a mode of around 55-60% gain-lovers.

CGS 2

In the data from the first robustness check, the order of endowment and learning about the forced exchange mechanism is reversed; now, subjects are informed about the forced exchange, then endowed with their good, and then reveal their WTP/WTA. This version of the experiment was conducted to address (KR)'s first-focus concept of CPE – perhaps subjects endowed with mugs (who later learned about forced exchange) held the mug as the referent rather than the stochastic distribution induced by forced exchange.

The results from this data were broadly consistent with a more loss-averse population than session 1; while there is still an endowment effect in the median prices – buyers WTP at CHF 4 and sellers WTA at CHF 5 – these numbers converge towards one another. Figure B.13 presents the estimated posterior distributions, which for this smaller sample is a bit more dependent on the prior. Under the *high-GL* prior, the data pushes the posterior to a slightly more



(a) High-GL Model (b) Low-GL Model

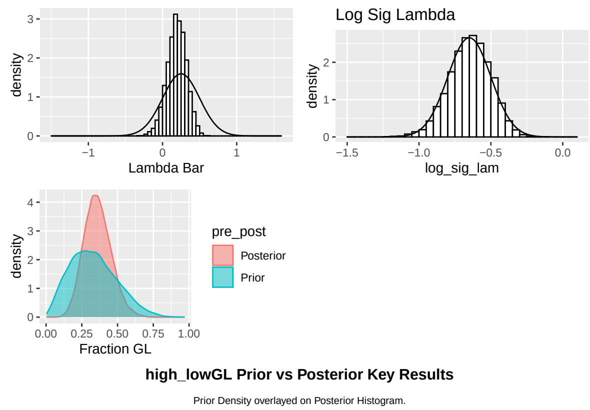
Figure B.13: **Posterior Results, CGS 2 Data**

gain-loving distribution, essentially ruling out less than 20% of gain-lovers and peaking around 40%. Meanwhile, under the *low-GL* prior, the data once again push the posterior towards a larger fraction of gain-lovers, but it contains most of its mass between 20% and 40%.

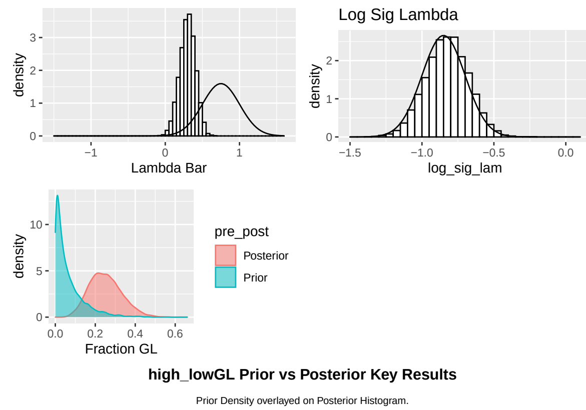
CGS 3

In the data for the final robustness check we consider, the authors keep the ”Mechanism First” design (as in Experiment 2) and introduce a Random Price determination mechanism, as opposed to the Market Price in the main experiment. In these sessions, the price was randomly set as one of the rows of the MPL for WTA/WTP, and the choice at that price was enacted. The motivation was to overcome any possible problems of variation in price expectations for buyers versus sellers, as well as to ensure that individuals were not mistakenly attempting to exert market power.

The data from this session are more mixed; median prices for buyers (sellers) are around CHF 2.5 (5) at the 50% forced exchange, and the treatment effects as the forced exchange probability shifts vary from positive to negative, consistent with a heterogeneous composition of gain-loss attitudes. This is reflected in the posteriors shown in Figure B.14, which are again fairly wide and indicate that anywhere between 20% to 60% of the data are estimated as gain-lovers



(a) High-GL Model



(b) Low-GL Model

Figure B.14: **Posterior Results, CGS 3 Data**

(depending on the prior).

Appendix C

C.1 KR Comparative Statics

In the main text, we derive comparative static predictions of effort under the KR CPE framework when the probability of the high fixed amount is varied. We design our experiment so that all effort levels fall into the case presented into the main text ($L < we < H$ because $L = 0, H = 20, w = 0.2$ and $e \in [0, 100]$. However, for completeness, we discuss the other two cases in this appendix.

C.1.1 Case 1: $we < L < H$

Assume first that $we < L < H$, so that the considered level of effort falls below the low fixed fee.The first order condition yielding optimal effort is

$$0.5w\,[1 + (p+q)\eta(\lambda - 1)] = c'(e),$$

and because $c'(e)$ is continuous and differentiable, $c'^{-1}(e)$ exists and the optimal e^* is

$$e^* = c'^{-1}\left(0.5w\,[1 + (p+q)\eta(\lambda - 1)]\right).$$

Turning back to $\frac{\partial e^*}{\partial p}|_{1-p-q=0.5}$, let $p+q=\bar{P}=0.5$ – since changes in p must leave $p+q$ constant, we have that $\frac{\partial e^*}{\partial p}|_{1-p-q=0.5}=0$ in this case.[1]

C.1.2 Case 3: $L<H<we$

Lastly, we can consider the case when $L<H<we$, so that the considered effort is above the high fixed fee. Again, we examine the first order condition given by

$$0.5w\left[1-(p+q)\eta(\lambda-1)\right]=c'(e),$$

and

$$e^*=c'^{-1}(0.5w[1-(p+q)\eta(\lambda-1)]),$$

yielding $\frac{\partial e^*}{\partial p}|_{1-p-q=0.5}=0$.

C.2 Robustness Results

[1] For a more concrete example, consider the cost function used in Augenblick and Rabin (2018): $c_i(e_i)=\frac{1}{\alpha\gamma_i}(e_i+10)^{\gamma_i}$. In this case, we can solve for

$$e^*=(\alpha 0.5w\left[1+(p+q)\eta(\lambda-1)\right])^{\frac{1}{\gamma-1}}.$$

The ratio of effort under two different treatment conditions, P' and Q' is then

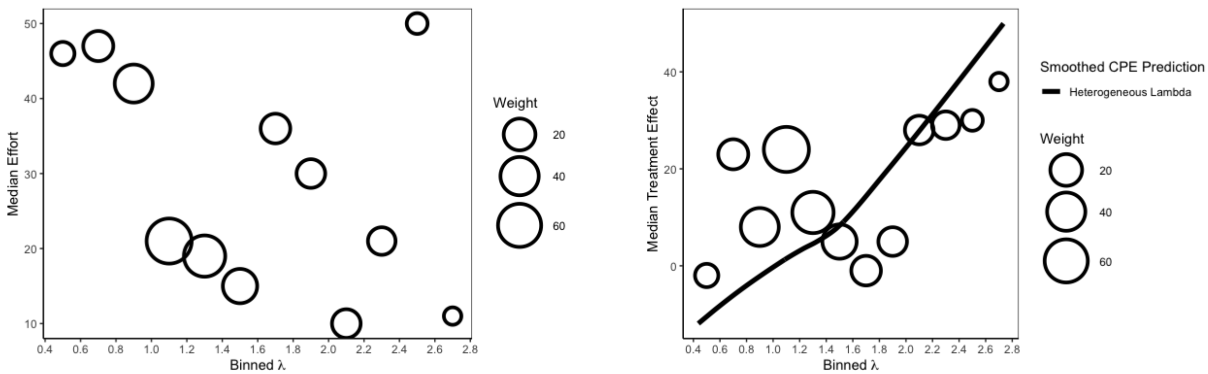
$$\frac{e_1^*+10}{e_2^*+10}=\frac{(\alpha 0.5w)^{\frac{1}{\gamma-1}}\,(1+(P')\eta(\lambda-1))^{\frac{1}{\gamma-1}}}{(\alpha 0.5w)^{\frac{1}{\gamma-1}}\,(1+(Q')\eta(\lambda-1))^{\frac{1}{\gamma-1}}},$$

so that the α terms disappear.

Table C.1: **Aggregate Parameter Estimates**

	(1)	(2)
	Estimate	(SD)
	$c(e) = \frac{1}{\alpha\gamma}(e+10)^{\gamma}$	
Gain-Loss Parameter:		
$\hat{\lambda}$	1.31	(0.573)
Cost of Effort:		
$\hat{\alpha}$	735.25	(4.930)
$\hat{\gamma}$	2.28	(0.256)

Notes: Posterior estimates of mean and standard deviation for each parameter using the full sample of 265. These are computed by simply taking the average (standard deviation) of our posterior draws. $c_i(e)$ refers to the cost-of-effort assumption made.



(a) Baseline Effort

(b) Treatment Effects

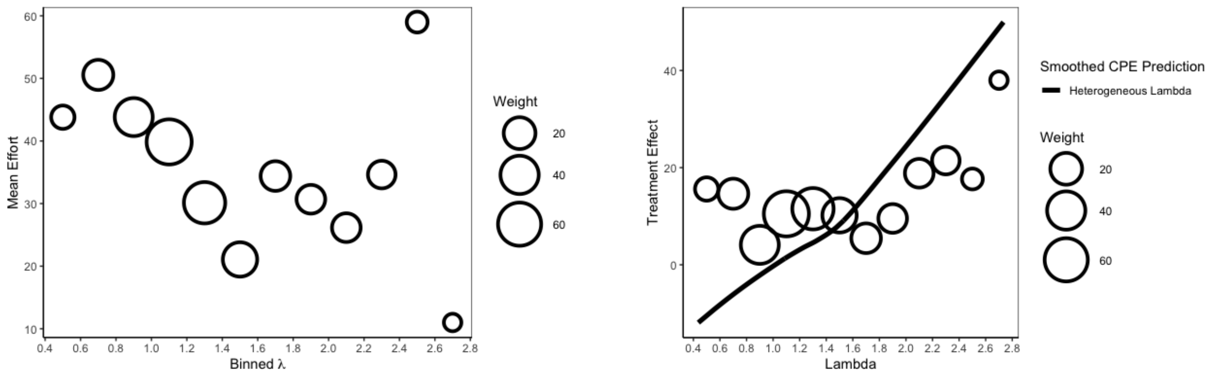
Figure C.1: **Stage 2 Results, Full Sample**

Notes: Smoothed CPE predictions are based on posterior λ_i means in our sample, while our Binned λ_i takes the median individual effort (in panel A) and treatment effect (panel B) for those with $\lambda_i \in [0.4, 0.6], [0.6, 0.8] \ldots$ and plots them based on the midpoint. Full sample, with no trimming based on posterior uncertainty (N = 265).

Table C.2: **Treatment Effects over Gain-Loss Type, First Round**

	Effort Choice	Effort Choice
Constant	35.00	40.00
	(3.94)	(7.20)
$\mathbf{1}(\lambda > 1)$		-10.00
		(8.74)
Treatment	15.00	10.00
	(4.38)	(8.16)
$\mathbf{1}(\lambda > 1) \times$ Treatment		5.00
		(10.52)
Observations	212	212

Notes: Quantile ($\tau = 0.5$) regression with bootstrapped standard errors in parentheses, trimmed sample, first round.



(a) Baseline Effort (b) Treatment Effects

Figure C.2: **Stage 2 Results, Full Sample, Average**

Notes: Smoothed CPE predictions are based on posterior λ_i means in our sample, while our Binned λ_i takes the average individual effort (in panel A) and treatment effect (panel B) for those with $\lambda_i \in [0.4, 0.6], [0.6, 0.8] \ldots$ and plots them based on the midpoint. Full sample, with no trimming based on posterior uncertainty (N = 265).

Table C.3: **Treatment Effects over Gain-Loss Preference**

	Effort Choice	Effort Choice
Constant	30.00	51.34
	(4.26)	(8.16)
λ		-16.81
		(5.41)
Treatment	14.00	6.51
	(5.52)	(11.75)
$\lambda \times$ Treatment		5.79
		(8.75)
Observations	265	265

Notes: Quantile ($\tau = 0.5$) regression with bootstrapped standard errors in parentheses, full sample. Aggregate represents the regression of effort on a constant and a treatment indicator (over all gain-loss types). Thus, the constant in column 3 represents the average number of tasks chosen when facing the low condition, and the treatment coefficient represents the aggregate treatment effect – how much the average number of tasks chosen changes when the fixed amount rises.

Table C.4: **Treatment Effects over Gain-Loss Preference, Mean**

	Effort Choice	Effort Choice
Constant	35.69	50.28
	(1.93)	(5.86)
λ		-11.13
		(4.16)
Treatment	10.89	3.53
	(2.83)	(8.60)
$\lambda\times$ Treatment		5.62
		(6.22)
Observations	265	265

Notes: OLS regression with robust standard errors in parentheses, full sample. Aggregate represents the regression of effort on a constant and a treatment indicator (over all gain-loss types). Thus, the constant in column 3 represents the average number of tasks chosen when facing the low condition, and the treatment coefficient represents the aggregate treatment effect – how much the average number of tasks chosen changes when the fixed amount rises.

Table C.5: **Treatment Effects over Gain-Loss Type, Mean**

	Effort Choice	Effort Choice
Constant	35.69	45.62
	(1.93)	(3.28)
$\mathbf{1}(\lambda > 1)$		-12.83
		(3.99)
Treatment	10.89	7.87
	(2.83)	(4.89)
$\mathbf{1}(\lambda > 1)\times$ Treatment		3.91
		(5.92)
Observations	265	265

Notes: OLS regression with robust standard errors in parentheses, full sample. Aggregate represents the regression of effort on a constant and a treatment indicator (over all gain-loss types). Thus, the constant in column 3 represents the average number of tasks chosen when facing the low condition, and the treatment coefficient represents the aggregate treatment effect – how much the average number of tasks chosen changes when the fixed amount rises.

Table C.6: **Treatment Effects over Gain-Loss Type, Mean, First Round**

	Effort Choice	Effort Choice
Constant	34.84	41.61
	(2.69)	(5.42)
$\mathbf{1}(\lambda > 1)$		-8.66
		(6.23)
Treatment	12.64	13.80
	(3.90)	(6.67)
$\mathbf{1}(\lambda > 1)\times$ Treatment		-1.68
		(8.12)
Observations	265	265

Notes: OLS regression with robust standard errors in parentheses, full sample. Aggregate represents the regression of effort on a constant and a treatment indicator (over all gain-loss types). Thus, the constant in column 3 represents the average number of tasks chosen when facing the low condition, and the treatment coefficient represents the aggregate treatment effect – how much the average number of tasks chosen changes when the fixed amount rises.

Table C.7: **Between Subject Comparative Static Test**

	(1)	(2)	(3)	(4)
	Dependent Variable: e			
	Full Sample	Loss Averse	Loss Neutral	Gain Loving
Structural Bounds Taxonomy				
Treatment	10.89	12.43	7.85	12.68
	(2.83)	(3.75)	(5.38)	(5.72)
Baseline (Constant)	35.69	29.20	41.63	44.71
	(1.93)	(2.42)	(3.86)	(3.81)
R-Squared	0.027	0.039	0.011	0.062
# Observations	265	136	91	38

Notes: Loss averse is defined as $\lambda_i > 1.2$, gain loving as $\lambda_i < 0.9$, and loss neutral is in between.

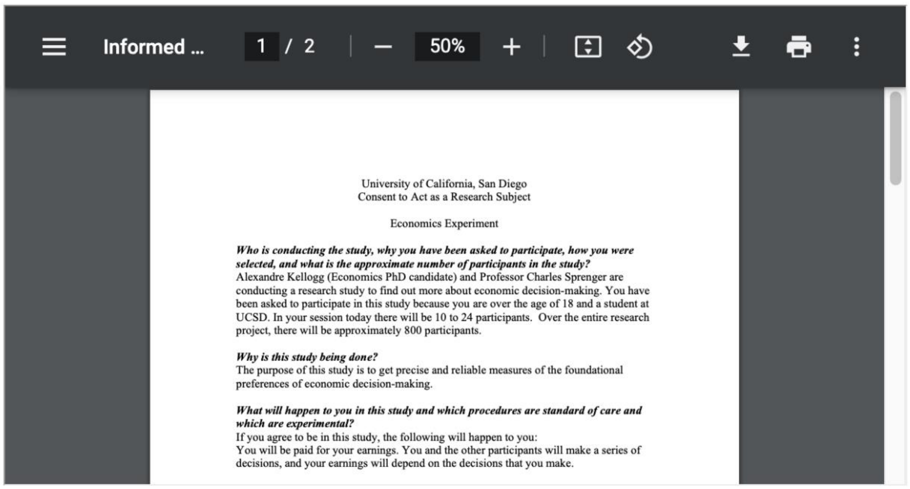
C.3 Instructions and Material Presented to Participants

The following set of screenshots demonstrates a demo version of our experiment, designed on oTree (Chen et al., 2016).

Welcome

Hello and thank you for taking the time to participate in this study. This experiment consists of several parts. Each part will include self-contained instructions about the relevant choices. If you're not sure about something at any time throughout the study, please feel free to contact your host for this session.

This study has been approved by the IRB at UCSD, and a copy of the consent form is attached below. By clicking next, you agree to take part in the experiment.



University of California, San Diego
Consent to Act as a Research Subject

Economics Experiment

Who is conducting the study, why you have been asked to participate, how you were selected, and what is the approximate number of participants in the study?
Alexandre Kellogg (Economics PhD candidate) and Professor Charles Sprenger are conducting a research study to find out more about economic decision-making. You have been asked to participate in this study because you are over the age of 18 and a student at UCSD. In your session today there will be 10 to 24 participants. Over the entire research project, there will be approximately 800 participants.

Why is this study being done?
The purpose of this study is to get precise and reliable measures of the foundational preferences of economic decision-making.

What will happen to you in this study and which procedures are standard of care and which are experimental?
If you agree to be in this study, the following will happen to you:
You will be paid for your earnings. You and the other participants will make a series of decisions, and your earnings will depend on the decisions that you make.

☐ I consent to be a participant in this study.

Next

Receiving Payment and Your Identity

All payments will be sent to you via Venmo/Zelle. In order to receive payment, we will need to collect an email address linked to this form of payment. This information will only be seen by the PIs in this study. As soon as your payments are made, the link between the choices you made and your payment will be destroyed, and the record with your email address will be deleted. Your identity will not be a part of the subsequent data analysis.

Please enter the email associated with your Zelle account:

Next

Experiment Overview

In the main part of this experiment, you will have the possibility to earn money by completing a number of tasks. Each task consists of transcribing a line of blurry Greek letters from a Greek text. The experiment is divided into five parts, which we will explain in turn.

Part 1

In **Part 1**, you will make 32 decisions. In each decision, you will be offered a rate for each task that you complete, and you will be asked to decide how many tasks you want to complete at that rate. For example, in one of the decisions you could be offered $0.20 for every task that you complete, and you will have to decide how many tasks (from 0 to 100) you want to complete at that rate.

Before you make any decisions about the number of tasks you wish to complete for each rate, you will be asked to complete 2 practice tasks. This will allow you to become acquainted with the task and give you a sense of how long a task takes you to complete.

Part 1 takes approximately 20 minutes.

Part 2

In **Part 2**, you will make 42 decisions. In each decision, you will be presented with two options: Option A will be a lottery, paying (for example) $10 with 20% chance and $0 with 80% chance and Option B will be a sure amount of $5. For each choice, the probabilities in Option A will vary, and you will be asked to indicate which option you prefer.

Part 2 takes approximately 10 minutes.

Part 3

After you have made your decisions in **Part 1** and **Part 2**, a computer will randomly determine which decision from **Part 1** or **Part 2** is chosen to be the *decision-that-counts*. The *decision-that-counts* will determine which of your prior choices will be implemented, and will thus determine part of your earnings for this study. Each of the choices you make are equally likely to be selected.

Next, everyone will be required to complete 10 tasks. Completing these mandatory tasks is required in order to earn your completion fee of $7.00 as well as the earnings determined from the *decision-that-counts*.

Part 3 can take anywhere from 5 minutes to 20 minutes, depending on how long it takes you to complete a task. On average, it takes about 42 seconds to complete one task.

Part 4

Once you have completed the 10 tasks, you may be asked to complete additional tasks as determined by your prior choices from the *decision-that-counts*.

Part 4 may take anywhere from 1 minute to 2 hours depending on which decision was selected as the *decision-that-counts*, the number of tasks that you selected, and the time it takes you to complete each task.

Part 5

Once you have finished with **Part 4**, you will then be asked to solve a few puzzles and answer a few demographic questions. There are a total of 5 puzzles to solve within 10 minutes. For each correct answer you submit, you will receive an additional $1.

After filling out your responses, you will receive your final payment via the account information you provided in the previous page. Your final payment will consist of a $7.00 completion fee + your earnings from completing each of the parts. If you do not complete all of the tasks you had previously chosen during **Part 4**, you will not receive the completion fee nor payment for any of the tasks, and will instead receive a show-up fee of $5.00.

Summary

To review, this experiment consists of 5 distinct parts. In **Part 1** and **Part 2**, you will make a series of choices, each of which is equally likely to determine your payment. In **Part 3**, the *decision-that-counts* will be revealed and you will be asked to complete the required 10 tasks. In **Part 4**, you will be asked to complete any additional tasks determined by your prior choices (if relevant). Finally, **Part 5** has a brief set of puzzles and survey questions prior to concluding the experiment.

More detailed instructions will be presented prior to each part. If you have questions or want clarification, remember that you can always contact your host for this session. On the next page, you will learn more about the Greek transcription tasks you will be asked to complete.

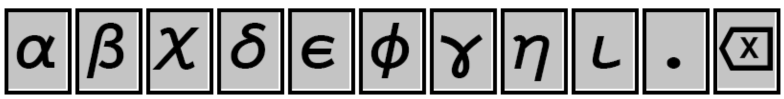
Next

The Task

To complete a task, you will have to transcribe a line of blurry letters from a Greek text. For each task, Greek text will appear on your screen. You will be asked to transcribe these letters by finding and clicking on the corresponding letter, which will insert that letter into the completion box. If you would like to delete the most recently added character, please click on the backspace image. One task is one row of Greek text. For the task to be complete, your transcription must be 80% accurate or better. The following is an example of a row of blurry Greek text and the row of letters you will be asked to select from, along with its solution:

δηιφα.βαιδφι.ηβεχφηφιεδ.γηειφφγφαγ.

Please select from the following characters to enter your transcription.



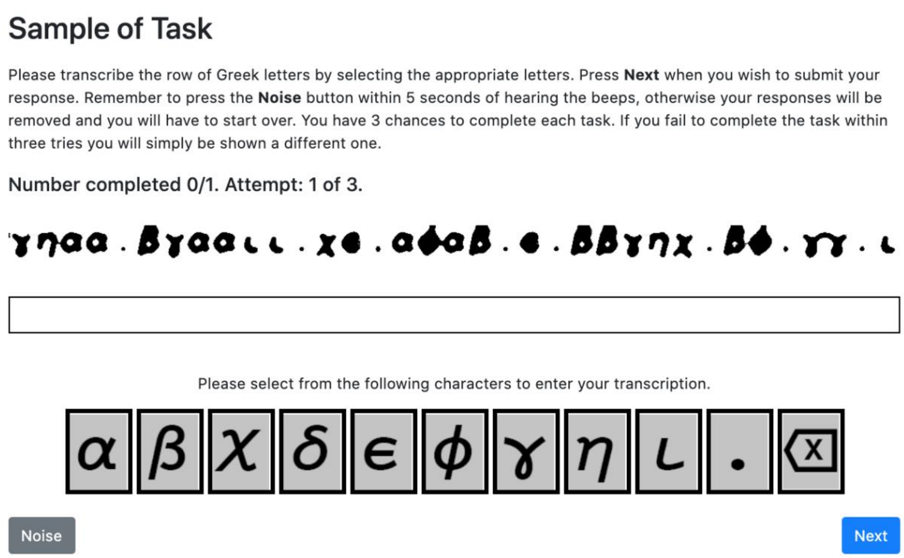
The correct transcription for the example task is provided above; recall, however, that you do not need to be 100% accurate. If you submit a transcription that is 80% accurate or better (defined as requiring 7 or fewer insertions, deletions, or substitutions to achieve the perfect transcription), you will have completed the task. If your answer is incorrect (less than 80% accurate), you will have 2 more tries (for a total of 3) to submit a correct answer, after which you will be presented with a new task and the incorrect submissions will not be counted as a completed task.

As part of the task, several auditory "beeps" will sound randomly while you are completing the tasks. Please put on your headphones and/or turn your volume up so that you can hear the beeping noises. After each time you hear this beeping noise, you must press the "Noise" button at the bottom left of the screen. If you do not press the "Noise" button within five seconds of hearing the beeping noise, your transcription will be reset. If you press the noise button erroneously (when there was no beeping sound), your transcription will be reset. Note that resetting the transcription does not count as a try; your current progress will simply be deleted and you will have to re-enter the transcription.

The time it takes to complete a task will vary from person to person. **On average, however, each task takes about 42 seconds to complete.**

Before you make your decisions, we will present you with 1 tasks so that you better understand what it means to complete a task as you make your choices. This will also provide you a chance to ensure that you can hear the beeping noises correctly. If you have any trouble with the task or any questions about it, please contact your host!

Next


Sample of Task
Please transcribe the row of Greek letters by selecting the appropriate letters. Press Next when you wish to submit your response. Remember to press the Noise button within 5 seconds of hearing the beeps, otherwise your responses will be removed and you will have to start over. You have 3 chances to complete each task. If you fail to complete the task within three tries you will simply be shown a different one.
Number completed 0/1. Attempt: 1 of 3.
Please select from the following characters to enter your transcription.
α β χ δ ϵ ϕ γ η ι .
Noise
Next

Part 1

Throughout the following screens, you will make a total of 32 decisions. We will begin by explaining the kind of decisions that you will make for the first 30 decisions of the study. After you make these 30 decisions, you will receive a new set of instructions regarding the last 2 decisions.

Decisions 1 to 30

In the next six screens, you will have to decide how many tasks you are willing to complete for a given rate. *As a reminder, one task means a correct transcription of a blurry line of Greek text*. The rates will be presented in lists of 5 at a time, and all rates within a list will either be *deterministic*, for example $0.15/task, or *stochastic*, for example a 50% chance of $0.10/task and a 50% chance of $0.20/task. The rates per task will range from $0.00 to $0.60 per task.

An example of your choice environment is provided below.

Low Wage (50%)	High Wage(50%)		Chosen Tasks	
$0.0/task	$0.1/task	(50% Chance of $0.00 50% Chance of $2.20)	22 tasks (~16mins)	

Each rate will have a corresponding slider where you can choose, for that rate, how many tasks you are willing to complete. As you move the slider, you will see a subtotal next to the rate, as well as the estimated time to complete the number of Greek tasks indicated. This time is estimated based on the average of 42 seconds per task at the bottom of the page, but you may enter your own estimated time given what you learned in the practice tasks.

Recall that each of the decisions that you will make throughout **Part 1** and **Part 2** of this study is equally likely to be the decision that counts. **Thus, it is very important that you think carefully about each decision you make, as it could be the one selected for payment.** If one of these 30 decisions is randomly selected to be the *decision-that-counts* , you will be asked to complete the number of Greek tasks that you indicated and you will be compensated at the rate specified by the *decision-that-counts*. Recall that everyone will be required to complete their 10 tasks before continuing on to the number of tasks that you indicated in the *decision-that-counts*.

If the rate for the *decision-that-counts* is deterministic, say $0.15/task, and you said you would work 50 tasks at that rate, you will be paid a $7.00 completion fee + $7.50 for your tasks for a total of $14.50 (plus an additional $1 for each puzzle you correctly solve in **Part 5**).

If the *decision-that-counts* involves a stochastic rate, say $0.10/task with 50% chance and $0.20/task with 50% chance, and you chose to work 50 tasks, then you will be asked to complete the 50 tasks after the mandatory 10. After you have completed all of these tasks, the computer will reveal which of these two rates applies by flipping a coin. Once the rate is determined, say the computer selects $0.20/task ($0.10/task), you will be paid a total of $17.00 ($12.00), $7.00 for the completion fee + $10.00 ($5.00) for your 50 tasks (plus an additional $1 for each puzzle you correctly solve in **Part 5**).

Over the next six pages, you will be presented with a series of 5 wages per page and asked to indicate the amount of tasks you wish to complete at the given rates.

Next

Effort Choices

Please indicate the number of tasks you are willing to do at each of the following wages by adjusting the slider below. Recall that, in addition to the number of tasks you select here, everyone will be required to complete 1 tasks paid for by the completion fee. There is an equal chance that each of these wages will be selected as the decision-that-counts, so please make your decisions carefully. **Make sure to adjust all the sliders before continuing, otherwise you will be asked to re-enter your choices.**

Low Wage (50%)	High Wage(50%)		Chosen Tasks
\$0.0/task	\$0.1/task	(50% Chance of \$0.00 50% Chance of \$2.20)	22 tasks (~16mins)
\$0.0/task	\$0.2/task	(50% Chance of \$0.00 50% Chance of \$13.40)	67 tasks (~47mins)
\$0.025/task	\$0.225/task	(50% Chance of \$0.48 50% Chance of \$4.28)	19 tasks (~14mins)
\$0.05/task	\$0.25/task	(50% Chance of \$3.65 50% Chance of \$18.25)	73 tasks (~52mins)
\$0.075/task	\$0.275/task	(50% Chance of \$1.95 50% Chance of \$7.15)	26 tasks (~19mins)

Hourly wage and time computed using task time of (sec): 42

☐ I confirm my final choices for all 5 sliders.

Next

Effort Choices

Please indicate the number of tasks you are willing to do at each of the following wages by adjusting the slider below. Recall that, in addition to the number of tasks you select here, everyone will be required to complete 1 tasks paid for by the completion fee. There is an equal chance that each of these wages will be selected as the decision-that-counts, so please make your decisions carefully. **Make sure to adjust all the sliders before continuing, otherwise you will be asked to re-enter your choices.**

Wage	Chosen Tasks
$0.2/task	
$0.225/task	
$0.25/task	
$0.275/task	
$0.3/task	

Hourly wage and time computed using task time of (sec): 42

☐ I confirm my final choices for all 5 sliders.

Next

Effort Choices

Please indicate the number of tasks you are willing to do at each of the following wages by adjusting the slider below. Recall that, in addition to the number of tasks you select here, everyone will be required to complete 1 tasks paid for by the completion fee. There is an equal chance that each of these wages will be selected as the decision-that-counts, so please make your decisions carefully. **Make sure to adjust all the sliders before continuing, otherwise you will be asked to re-enter your choices.**

Wage	Chosen Tasks
$0.05/task	
$0.1/task	
$0.125/task	
$0.15/task	
$0.175/task	

Hourly wage and time computed using task time of (sec): 42

☐ I confirm my final choices for all 5 sliders.

Next

Effort Choices

Please indicate the number of tasks you are willing to do at each of the following wages by adjusting the slider below. Recall that, in addition to the number of tasks you select here, everyone will be required to complete 1 tasks paid for by the completion fee. There is an equal chance that each of these wages will be selected as the decision-that-counts, so please make your decisions carefully. **Make sure to adjust all the sliders before continuing, otherwise you will be asked to re-enter your choices.**

Low Wage (50%)	High Wage(50%)	Chosen Tasks	
$0.1/task	$0.3/task		
$0.125/task	$0.325/task		
$0.15/task	$0.35/task		
$0.175/task	$0.375/task		
$0.2/task	$0.4/task		

Hourly wage and time computed using task time of (sec): 42

☐ I confirm my final choices for all 5 sliders.

Next

Effort Choices

Please indicate the number of tasks you are willing to do at each of the following wages by adjusting the slider below. Recall that, in addition to the number of tasks you select here, everyone will be required to complete 1 tasks paid for by the completion fee. There is an equal chance that each of these wages will be selected as the decision-that-counts, so please make your decisions carefully. **Make sure to adjust all the sliders before continuing, otherwise you will be asked to re-enter your choices.**

Low Wage (50%)	High Wage(50%)	Chosen Tasks
\$0.075/task	\$0.375/task	
\$0.05/task	\$0.45/task	
\$0.0/task	\$0.5/task	
\$0.1/task	\$0.5/task	
\$0.0/task	\$0.6/task	

Hourly wage and time computed using task time of (sec): 42

☐ I confirm my final choices for all 5 sliders.

Next

Part 1 Continued

Decisions 31 and 32

Next, you will be asked to make your final two decisions for **Part 1** of this study. Each of the two decisions will be presented on their own page, so **please make sure you carefully review the rates for each decision**. As in the previous decisions, you will have to decide how many tasks to complete at different rates. The only difference in these two decisions is the structure of the rates: with 50.0% chance, you will get $0.20/task, with 5.0% chance you will get a fixed payment of $X *regardless of the number of tasks that you decided to do*, and with 45.0% chance you will get a fixed payment of $Y *regardless of the number of tasks that you decided to do*. For example, if you select to complete 30 tasks, then after you complete the 30 tasks you will either be paid $0.20/task, $X, or $Y.

Recall that at the end of the study, one of the 32 decisions you've made in **Part 1**, including these final two, may be randomly selected for payment. This means that each decision is equally likely to be the decision-that-counts. **Thus, it is very important that you think carefully about each decision you make, as it could be the one selected for payment.**

Next

Effort Choices

Please indicate the number of tasks you are willing to do at each of the following wages by adjusting the slider below. Recall that there is an equal chance that each of these wages will be selected as the decision-that-counts. Make sure to adjust all the sliders before continuing, otherwise you will be asked to re-enter your choices.

Fixed (L) (5.0%)	**Fixed (H) (45.0%)**	**Wage (50.0%)**		**Chosen Tasks**	
$0	$20	$0.2/task	(50% Chance of $3.20 5% Chance of $0.00 45% Chance of $20.00)	16 tasks (~12mins)	

Hourly wage and time computed using task time of (sec): 42

Next

Effort Choices

Please indicate the number of tasks you are willing to do at each of the following wages by adjusting the slider below. Recall that there is an equal chance that each of these wages will be selected as the decision-that-counts. Make sure to adjust all the sliders before continuing, otherwise you will be asked to re-enter your choices.

Fixed (L) (45.0%)	**Fixed (H) (5.0%)**	**Wage (50.0%)**	**Chosen Tasks**	
$0	$20	$0.2/task		

Hourly wage and time computed using task time of (sec): 42

Next

Instructions for Part 2

On the following pages, you will be asked to make 21 choices per page. In each choice, you will be presented with two options -- "Option A" and "Option B" -- and asked to indicate which of the two you prefer.

"Option A" will be a lottery that pays either $10.00 ($3.00) with probability varying from 0.0% to 100.0%, or $0.00 (-$3.50) otherwise; "Option B" yields a payoff of $5.00 ($0.00) for sure, i.e. with a probability of 100%.

On each page, the first and last choice will be selected by default to help demonstrate that Option B is initially the preferred option, but Option A grows more desirable in each row; by the last choice, Option A should clearly be preferred. You will not be able to change these choices.

For the remaining choices, please select your preference between Option A and Option B. Once you have switched from Option B to Option A, all subsequent choices will be automatically switched to Option A. This is intended to help maintain consistency due to the ordering of the choices: if you prefer Option A to Option B in choice number 10 (for example), then you should prefer Option A to Option B in choice 11, 12, and so on. An example of this for a set of potential choices is shown below. Someone who prefers a 10% chance of $10 (and 90% chance of $0) to $5 for sure should also prefer a 15% chance of $10 (and 85% chance of $0) to $5 for sure, because a 15% chance is strictly better than a 10% chance and the $5 for sure never changes.

Option A		Option B
$10.00 with a probability of 0.0%,$0.00 otherwise	○ Option A ◉ Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 5.0%,$0.00 otherwise	○ Option A ◉ Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 10.0%,$0.00 otherwise	◉ Option A ○ Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 15.0%,$0.00 otherwise	◉ Option A ○ Option B	$5.00 with a probability of 100.0%

After you have made all of your choices, please review the page prior to submitting these decisions. Recall that one of these decisions may be randomly chosen for your payment.

If you indicated that you prefer Option A (the lottery) for the relevant decision, a random number between 1 and 100 will be generated to determine the outcome of the lottery. For instance, if the **decision-that-counts** is $10.00 with 20% chance and $0.00 with 80% chance, a random number between 1-80 will result in payment of $0.00, but a random number between 81-100 will result in payment of $10.00.

If you indicated that you prefer Option B for the relevant decision, you would receive $5.00 in this example.

Recall that, along with your choices from **Part 1**, each of the choices you make in **Part 2** are equally likely to be the *decision-that-counts*. Please carefully consider each choice as they are all equally likely to determine your final payment.

Next

Part 2 Decisions

Option A		Option B
$3.00 with a probability of 0.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 5.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 10.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 15.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 20.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 25.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 30.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 35.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 40.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 45.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 50.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 55.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 60.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 65.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 70.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 75.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 80.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 85.0%,-$3.50 otherwise	○ Option A ◉ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 90.0%,-$3.50 otherwise	◉ Option A ○ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 95.0%,-$3.50 otherwise	◉ Option A ○ Option B	$0.00 with a probability of 100.0%
$3.00 with a probability of 100.0%,-$3.50 otherwise	◉ Option A ○ Option B	$0.00 with a probability of 100.0%

Next

Part 2 Decisions

Option A		Option B
$10.00 with a probability of 0.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 5.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 10.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 15.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 20.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 25.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 30.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 35.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 40.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 45.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 50.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 55.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 60.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 65.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 70.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 75.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 80.0%,$0.00 otherwise	○ Option A ● Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 85.0%,$0.00 otherwise	● Option A ○ Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 90.0%,$0.00 otherwise	● Option A ○ Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 95.0%,$0.00 otherwise	● Option A ○ Option B	$5.00 with a probability of 100.0%
$10.00 with a probability of 100.0%,$0.00 otherwise	● Option A ○ Option B	$5.00 with a probability of 100.0%

Next

Results

The following decision was randomly chosen for your payment:

Option A		Option B
$10.00 with a probability of 85.0%, $0.00 otherwise	◉ ○	$5.00 with a probability of 100.0% (sure payoff)

As shown above, you indicated that you prefer Option A in this decision.

For the lottery, one of the two possible outcomes has been randomly realized based on the corresponding probabilities.

Your payoff in this task equals **$10.00**.

Next

The Task

Recall that for each task, you will have to transcribe a line of blurry Greek letters from a Greek text. One task is one row of Greek text. For the task to be complete, your transcription must be 80% accurate or better. The following is an example of a row of blurry Greek text and the dictionary:

δηιφα.βαιδφι.ηβϵχφηφιϵδ.γηϵιφφγφαγ.

δηιφα.βαιδφι.ηβϵχφηφιϵδ.γηϵιφφγφαγ.

Please select from the following characters to enter your transcription.

α	β	χ	δ	ϵ	φ	γ	η	ι	.	⌫

The correct transcription for the example task is provided; recall, however, that you do not need to be 100% accurate. If you submit a transcription that is 80% accurate or better (defined as requiring 7 or fewer insertions, deletions, or substitutions to achieve the perfect transcription), you will have completed the task. If your answer is incorrect (less than 80% accurate), you will have 2 more tries (for a total of 3) to submit a correct answer, after which you will be presented with a new task and the incorrect submissions will not be counted as a completed task.

As part of the task, several auditory "beeps" will sound randomly throughout the transcription process. Please put on your headphones and/or adjust the volume so that you can hear the beeping noises. After each time you hear this beeping noise, you must press the "Noise" button at the bottom left of the screen. If you do not press the "Noise" button within five seconds of hearing the beeping noise, your transcription will be reset. If you press the noise button erroneously (when there was no beeping sound), your transcription will be reset. Note that resetting the transcription does not count as a try; your current progress will simply be deleted and you will have to re-enter the transcription.

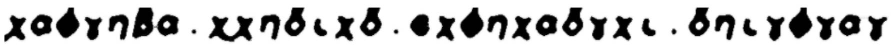
Once you click the next button, you will be presented with the 1 required tasks. After you complete these, you will be continue onto **Part 5** and attempt to solve several puzzles ($1 per correct submission) and answer a few demographic questions. Then, you will receive compensation based on your *decision-that-counts*. Recall that you were randomly selected to be paid $10.00 from the lottery task.
Once you finish all of these tasks, you will receive a completion fee of $7.00 in addition to your lottery payout, for a total of $17.00 (plus $1 per correct puzzle entry).

Next

Mandatory Tasks

Please transcribe the row Greek letters by selecting the appropriate letters. Press **Next** when you wish to submit your response. Remember to press the **Noise** button within 5 seconds of hearing the beep, otherwise your responses will be removed and you will have to start over. You will have 3 chances to complete each task. If you fail to complete the task within three tries you will simply be shown a different one.

Number completed 0/1. Attempt: 1 of 3.



Please select from the following characters to enter your transcription.

α β χ δ ϵ ϕ γ η ι . ⌫

Noise Next

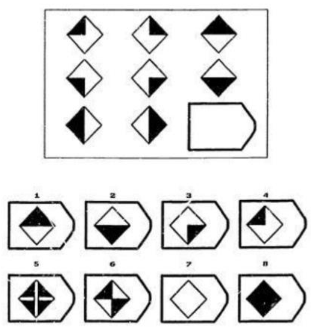
Part III Solving puzzles

Instructions

You will be presented with 5 problems, each showing a pattern with a bit cut out of it. Look at the pattern, think what piece is needed to complete the pattern correctly both along the rows and down the columns, BUT NOT THE DIAGONALS.

You will be paid $1 for each correct problem you solve. You have 10 minutes to complete all problems.

For example, for the following matrix, the correct pattern is 8.



Click next to start solving the problems.

Next

Make your choice

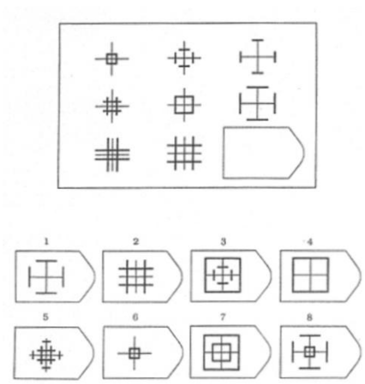
Time left to complete this section: **9:52**

Question 1 of 5

Instructions

You will be presented with 5 problems, each showing a pattern with a bit cut out of it. Look at the pattern, think what piece is needed to complete the pattern correctly both along the rows and down the columns, BUT NOT THE DIAGONALS.

You will be paid $1 for each correct problem you solve. You have 10 minutes to complete all problems.



Please choose an item that best fits the pattern:

Next

Make your choice

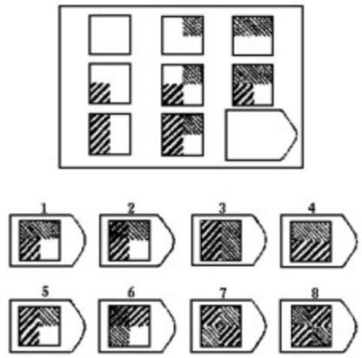
Time left to complete this section: **9:37**

Question 2 of 5

Instructions

You will be presented with 5 problems, each showing a pattern with a bit cut out of it. Look at the pattern, think what piece is needed to complete the pattern correctly both along the rows and down the columns, BUT NOT THE DIAGONALS.

You will be paid $1 for each correct problem you solve. You have 10 minutes to complete all problems.



Please choose an item that best fits the pattern:

Next

Make your choice

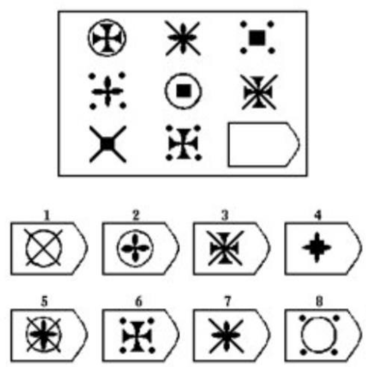
Time left to complete this section: **9:21**

Question 3 of 5

Instructions

You will be presented with 5 problems, each showing a pattern with a bit cut out of it. Look at the pattern, think what piece is needed to complete the pattern correctly both along the rows and down the columns, BUT NOT THE DIAGONALS.

You will be paid $1 for each correct problem you solve. You have 10 minutes to complete all problems.



Please choose an item that best fits the pattern:

4

Next

Make your choice

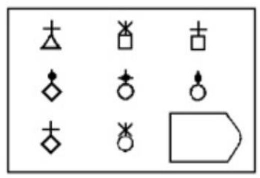
Time left to complete this section: **9:06**

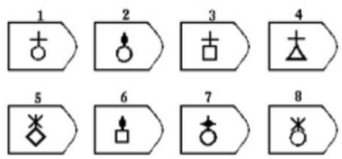
Question 4 of 5

Instructions

You will be presented with 5 problems, each showing a pattern with a bit cut out of it. Look at the pattern, think what piece is needed to complete the pattern correctly both along the rows and down the columns, BUT NOT THE DIAGONALS.

You will be paid $1 for each correct problem you solve. You have 10 minutes to complete all problems.





Please choose an item that best fits the pattern:

Next

Make your choice

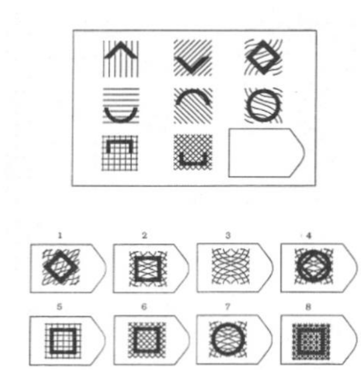
Time left to complete this section: **8:43**

Question 5 of 5

Instructions

You will be presented with 5 problems, each showing a pattern with a bit cut out of it. Look at the pattern, think what piece is needed to complete the pattern correctly both along the rows and down the columns, BUT NOT THE DIAGONALS.

You will be paid $1 for each correct problem you solve. You have 10 minutes to complete all problems.



Please choose an item that best fits the pattern:

Next

Results

You have completed all problems.

You have correctly solved 0 problems.

Your total payment for this part is $0.

Next

Thank You for Participating

Before we finalize your earnings, please answer the following short survey.

What year of your undergraduate education are you in?

○ First ○ Second ○ Third ○ Fourth ○ Other

What is your major or intended major?

What is your gender?

○ Male ○ Female ○ Other ○ Decline to Answer

Which of the following income brackets do your parents fall into?

○ Below 50k ○ 50k to 100k ○ Above 100k ○ Decline to Answer

How do you evaluate yourself: Are you in general a more risk-taking (risk-prone) person (10) or do you try to avoid risks (0, risk-averse)?

○ 0 (Risk averse) ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 (Fully prepared to takes risks)

Next

Thank You for Participating

As a reminder, your lottery payoff was randomly determined to be $10.00.
Your total earnings, including the completion fee of $7.00 and the earnings from the Raven Matrices of $0.00, are $17.00.

Next

Bibliography

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

Augenblick, Ned and Matthew Rabin (Mar. 2018). “An Experiment on Time Preference and Misprediction in Unpleasant Tasks”.

Backus, Matthew, Thomas Blake, Dimitriy V. Masterov, and Steven Tadelis (Jan. 2017). “Expectation, Disappointment, and Exit: Reference Point Formation in a Marketplace”.

Bell, David E. (1985). “Disappointment in Decision Making under Uncertainty”. In: *Operations Research* 33.1, pp. 1–27.

Bradley, Margaret and Peter Lang (1994). “Measuring emotion: the self-assessment manikin and the semantic differential”. In: *Journal of behavior therapy and experimental psychiatry* 25.1, pp. 49–59.

Burks, Stephan, Jeffrey Carpenter, Lorenz Goette, and Aldo Rustichini (2009). “Cognitive skills affect economic preferences, strategic behavior, and job attachment”. In: *Proceedings of the National Academy of Science* 106.19, pp. 7745–7750.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler (1997). “Labor supply of New York City cabdrivers: One day at a time”. In: *The Quarterly Journal of Economics*, pp. 407–441.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu (2016). “Evaluating replicability of laboratory experiments in economics”. In: *Science* 351.6280, pp. 1433–1436.

Cantillo, Vıctor, Johanna Amaya, and Juan de Dios Ortúzar (2010). “Thresholds and Indifference in Stated Choice Surveys”. In: *Transportation Research Part B* 44, pp. 753–763.

Cerulli-Harms, Annette, Lorenz Goette, and Charles Sprenger (Feb. 2019). “Randomizing Endowments: An Experimental Study of Rational Expectations and Reference-Dependent Preferences”. In: *American Economic Journal - Microeconomics* 11.1, pp. 185–207.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer (2017). *Willingness to Pay and Willingness to Accept are Probably Less Correlated Than You Think*. Tech. rep. National Bureau of Economic Research.

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer (2018). *Loss Attitudes in the US Population: Evidence from Dynamically Optimized Sequential Experimentation (DOSE)*. Tech. rep. National Bureau of Economic Research.

Chen, Daniel L., Martin Schonger, and Chris Wickens (Mar. 2016). "oTree - An open-source platform for laboratory, online, and field experiments". In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Crawford, Vincent P and Juanjuan Meng (2011). "New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets for hours and income". In: *The American Economic Review* 101.5, pp. 1912–1932.

Dean, Mark and Pietro Ortoleva (2015). "Is It All Connected". In: *Working Paper*.

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F Schmieder (2017). "Reference-dependent job search: evidence from Hungary". In: *The Quarterly Journal of Economics*.

Erev, Ido, Eyal Ert, and Eldad Yechiam (2008). "Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions". In: *Journal of Behavioral Decision Making* 21.5, pp. 575–597.

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.

Farber, Henry (2005). "Is Tomorrow Another Day? The Labor Supply of New York City Cab Drivers". In: *Journal of Political Economy* 113, pp. 46–82.

Fehr, Ernst and Lorenz Goette (2007). "Do Workers Work More When Wages Are High? Evidence from a Ranomized Field Experiment". In: *The American Economic Review* 97.1, pp. 298–317.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental economics* 10.2, pp. 171–178.

Frederick, Shane (2005). "Cognitive Reflection and Decision Making". In: *Journal of Economic Perspectives* 19.4, pp. 25–42.

Gachter, Simon, Eric J. Johnson, and Andreas Herrmann (2007). "Individual-Level Loss Aversion in Riskless and Risky Choices". In: *Working Paper*.

Gill, David and Victoria Prowse (2012). "A structural analysis of disappointment aversion ina real effort competition". In: *The American Economic Review* 102.1, pp. 469–503.

Gneezy, Uri, Lorenz Goette, Charles Sprenger, and Florian Zimmermann (2017). “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 Gain-Loss Attitudes and Expectations-Based Reference Points”.

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

Harinck, Fieke, Eric Van Dijk, Ilja Van Beest, and Paul Mersmann (2007). “When gains loom larger than losses reversed loss aversion for small amounts of money”. In: *Psychological Science* 18.12, pp. 1099–1105.

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.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1990). “Experimental Tests of the Endowment Effect and the Coase Theorem”. In: *Journal of Political Economy* 98.6, pp. 1325–1348.

Kahneman, Daniel and Amos Tversky (1979). “Prospect Theory: An Analysis of Decision under Risk”. In: *Econometrica* 47.2, pp. 263–292.

Kellogg, Alexandre (n.d.). “Gain-Loss Distributions in KR Exchange Paradigms: Bayesian Analysis of Experimental Literature”.

Knetsch, Jack and Wei-Kang Wong (2009). “The endowment effect and the reference state: Evidence and manipulations”. In: *Journal of Economic Behavior & Organization* 71.2, pp. 407–413.

Kőszegi, Botond and Matthew Rabin (2006). “A Model of Reference-Dependent Preferences”. In: *The Quarterly Journal of Economics* 121.4, pp. 1133–1165.

— (2007). “Reference-Dependent Risk Attitudes”. In: *The American Economic Review* 97.4, pp. 1047–1073.

Loomes, Graham and Robert Sugden (1986). “Disappointment and Dynamic Consistency in Choice under Uncertainty”. In: *Review of Economic Studies* 53.2, pp. 271–82.

Masatlioglu, Yusufcan and Collin Raymond (2016). “A Behavioral Analysis of Stochastic Reference Dependence”. In: *American Economic Review* 106.9, pp. 2760–82.

McElreath, Richard (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.

McFadden, Daniel (1974). “Conditional Logit Analysis of Qualitative Choice Behavior”. In: *Frontiers in Econometrics*. Ed. by Paul Zarembka. New York: Academic Press. Chap. 4.

Nicolau, Juan L (2012). "Asymmetric tourist response to price: loss aversion segmentation". In: *Journal of Travel Research* 51.5, pp. 568–676.

Odean, Terrance (1998). "Are Investors Reluctant to Realize Their Losses?" In: *The Journal of Finance* 53.5, pp. 177–1798.

Plott, Charles R. and Kathryn Zeiler (2005). "The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions and Experimental Procedures for Eliciting Valuations". In: *The American Economic Review* 95.3, pp. 530–545.

— (2007). "Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory?" In: *The American Economic Review* 97.4.

Rabin, Matthew (2000). "Risk Aversion and Expected Utility Theory: A Calibration Theorem". In: *Econometrica* 68.5, pp. 1281–1292.

Raven, John and Jean Raven (2003). "Raven Progressive Matrices". In: *Handbook of Nonverbal Assessment*. Ed. by R. Steve McCallum. Boston, MA: Springer US, pp. 223–237.

Schad, Daniel J., Michael Betancourt, and Shravan Vasishth (2021). "Toward a principled Bayesian workflow in cognitive science". In: *Psychological Methods* 26.1, pp. 103–126.

Smith, Alex (2019). "Lagged Beliefs and Reference-Dependent Utility". In: *Journal of Economic Behavior & Organization* 167, pp. 331–340.

Sokol-Hessner, Peter, Ming Hsu, Nina G Curley, Mauricio R Delgado, Colin F Camerer, and Elizabeth A Phelps (2009). "Thinking like a trader selectively reduces individuals' loss aversion". In: *Proceedings of the National Academy of Sciences* 106.13, pp. 5035–5040.

Sprenger, Charles (2015). "An endowment effect for risk: Experimental tests of stochastic reference points". In: *Journal of Political Economy* 123.6, pp. 1456–1499.

Stan Development Team (2020). *RStan: the R interface to Stan*. R package version 2.21.2. URL: http://mc-stan.org/.

Thakral, Neil and Lihn T. To (Forthcoming). "Daily Labor Supply and Adaptive Reference Points". In: *American Economic Review*.

Train, Kenneth E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Tversky, Amos and Daniel Kahneman (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty". In: *Journal of Risk and Uncertainty* 5.4, pp. 297–323.