RISK AVERSION: THE NEED FOR BEHAVIORAL EXPLANATION(S)

Jubo Yan [1]

April, 2017

[1]yanjubo@ntu.edu.sg. Division of Economics, Nanyang Technological University, Singapore 637332. I appreciate the invaluable suggestions and comments from William Schulze and Ted O'Donoghue. I am also grateful for the financial support provided by William Schulze. David Just, Richard Schuler, Brian Wansink, Dawn Schrader suggested revisions to the initial draft. Seminar participants at George Mason University, Maastricht University, Tsinghua University, Beijing Normal University, Xiamen University and audience at the 2014 international ESA meeting provided helpful comments. LEEDR lab manager Kevin Kniffin and staff Diana Viglucci, Kevin Krieger, Dan Connolly, and Lauren Schrader helped coordinating and conducting experiments. All errors remain the author's.

Abstract

Expected utility theory (EUT) is a parsimonious theory that explains behavior under risk and uncertainty. Previous research showed that EUT of *wealth* is not a satisfactory explanation of risk aversion. We use empirical data from a controlled laboratory experiment to show that EUT of *income* cannot explain risk aversion either. The experimental data suggests that the marginal utility of money would decrease at an absurdly high rate if the concavity of Bernoulli utility function is used to explain risk aversion. We demonstrate that loss aversion together with probability weighting explain the observed risk aversion well. Unlike many previous studies, we elicit valuations – Willingness to Pay (WTP) and Willingness to Accept (WTA) – for risky prospects to make the reference points more salient. Using an empirical model that features reference-dependent preference, we further show that our identification is robust to changes in reference points. The obtained parameters from estimating the structural model are consistent with literature.

Keywords: Risk Aversion, Loss Aversion, Probability Weighting, Reference Dependent, Lottery

1 Introduction

Economists favor parsimonious theories that can reasonably explain and predict behavior because of their portability and tractability. When explaining and predicting behavior under risk and uncertainty, expected utility theory (EUT) is considered a parsimonious theory. It has long been used to explain one of the most important empirical evidence in decision making – risk aversion – through a concave Bernoulli utility function[1]. Nonetheless, researchers found that EUT in fact does not provide a satisfactory explanation for risk aversion (Rabin and Thaler, 2001). The calibration practice that Rabin (2000) performed on EUT of wealth led to empirical implausibilities. In practice, especially empirical research, researchers often use the EUT of income model to explain and predict behavior because it is more feasible and convenient (Holt and Laury, 2002). More importantly, the method also circumvents the calibration problem that Rabin and Thaler (2001) raised. When the amount of income is small (e.g. payments in most laboratory experiments), is the EUT of income model indeed immune to the calibration problem? We answer this question empirically using a controlled laboratory experiment in this paper.

Prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) as an alternative to EUT is capable to make good explanations of many real world phenomena and its propositions have also been extensively tested in both the laboratory and the field. Most previous empirical studies have focused on using loss aversion to explain reference dependent behavior (Haigh and List, 2005; Kahneman et al., 1991; Thaler and Tversky, 1997; Tversky and Kahneman, 1991). In recent years, based on these empirical evidence, reference–dependent utility model has also been proposed and has received much attention (Köszegi and Rabin, 2006, 2007, 2009). In comparison, economic literature has less emphasized the empirical validity of probability weighting. The majority of empirical studies on probability weighting have been conducted by psychologists (Gonzalez and Wu, 1999; Kilka

[1] In the current paper, we use Bernoulli utility function to refer to the utility function $u(\cdot)$ on sure outcomes. The term Von–Neumann Morganstern (VNM) utility function is used to refer to the decision utility $U(\cdot)$ which is a weighted average of $u(\cdot)$s.

and Weber, 2000; Wu and Gonzalez, 1996) and relevant economic studies have been focusing on the theoretical aspect. These studies try to relax the linearity constraint on probability in expected utility theory to rationalize empirically observed anomalies (Abdellaoui, 2002; Machina, 2009; Quiggin, 1982, 1987). Fehr-Duda and Epper (2012) provides a comprehensive review of the studies on probability weighting and the literature recognize the importance of empirical testing. However, with field data, researchers usually do not observe the decision makers' perceived risks or probabilities which makes it very difficult or even impossible to distinguish probability distortion caused by probability weighting from those caused by systematic risk misperception (Barseghyan et al., 2013b)[2] . Laboratory experiments do not suffer from this issue because risk (objective probability) associated with a certain outcome can be made clearly known to subjects. Therefore, observed probability distortion in laboratory experiments can be safely attributed to probability weighting. In the current study, we *simultaneously* examine the roles of loss aversion and probability weighting in risk averse (loving) behavior using our experimental data.

We collect certainty equivalents (CE) of different risky prospects through a carefully designed laboratory experiment. Our experiment uses the iterative Multiple Price List (iMPL) elicitation technique (Andersen et al., 2006) to obtain Willingness to Pay (WTP) and Willingness to Accept (WTA) values of both pure monetary payoff ($5) and an actual commodity (coffee mug) in both the gain domain (represented by lottery) and the loss domain (represented by insurance). The valuation modes (i.e. gain vs. loss and WTA vs. WTP) are varied between subjects while probabilities of gain or loss are varied within subject. Using data from the experiment, we show that EUT of income model is not consistent with empirical evidence even when income is at small scale. Experimental results show a significant role of loss aversion under *all* probabilities.

Following Köszegi and Rabin (2006), we build an empirical model in which people receive

[2]Recently, some economists have started to investigate the role of probability distortion in explaining risk aversion (Barberis and Huang, 2008; Barseghyan et al., 2013a; Snowberg and Wolfers, 2010). The term probability distortion describes the behavior of *systematic* deviation from objective probabilities in decision making under EUT framework.

both consumption utility and gain–loss (reference dependent) utility to fit the experimental data. By assuming a *locally* linear utility function, we estimate the single loss aversion parameter as well as a non–parametric probability weighting function. The estimated loss aversion parameter is within a reasonable range as suggested by literature and the non-parametric probability weighting function is consistent with an inverse S–shaped curve. The model also fits the experimental data well and can be extended to other applications. Due to identification purpose, we employ a between–subject design so we have to assume a representative agent in our analyses. This method partially ignores the individual level heterogeneity that can be studied using within–subject design.

Our main contributions are twofold. On the one hand, we show that EUT of income model is empirically implausible and behavioral explanations are needed to explain the risk averse (loving) behavior observed in the laboratory. On the other hand, the design of the laboratory experiment allows us to *simultaneously* examine the roles of loss aversion and probability weighting in risk averse (loving) behavior. We are also the first to thoroughly test the reference-dependent utility model. By eliciting valuations for different risky prospects in both the gain and loss domains, the data allows for identification of an extended version of the reference-dependent utility model derived from the Koszegi–Rabin model(Köszegi and Rabin, 2006). Unlike direct elicitation of certainty equivalents, the reference-dependent model based this design is more robust to the changes in reference points.

The rest of the paper is organized as follows: the next section describes a controlled laboratory experiment which we use to elicit WTA and WTP values for either gains or losses of $5 or a coffee mug under risk. We also explain why and how this design allows us to examine loss aversion and probability weighting simultaneously. Section three presents the experimental results and conducts a simple reduced–form analysis. In section four, loss aversion and probability weighting are discussed in details before we present an empirical model to explain the experimental data. The structural estimation then shows the magnitudes of loss aversion and probability weighting we observed in the experiment. The last section concludes the study and discusses possible future research directions.

2 Experimental Design

The experiment is conducted with the risky prospects of \$5 (money experiment hereafter) or of a coffee mug with a university logo (mug experiment hereafter). In this section, we mostly refer to the money experiment when illustrating the design, but the same design applies to the mug experiment unless otherwise noted.

The experiment was designed to satisfy several criteria. First of all, we believe people's different attitudes towards gains and looses is the fundamental reason of risk aversion at small scales (Rabin, 2000), so all risky prospects in the experiment involve gains and/or losses depending on what the decision maker's reference point is[3]. We design our experiment to make the reference points salient to subjects so the researcher can better tell which direction (positive/gain vs. negative/loss) the decision maker is deviating from her reference point. This is critical in identifying the role of loss aversion in risk averse behaviors. Secondly, the influence of reference point on the estimation should be minimized. That is, our estimation is *empirically* robust to the changes in reference points. Although reference dependence has been generally accepted in decision science and behavioral economics, we know relatively little that how reference points are determined. In fact, the reference point can be context dependent and changes dynamically even within the scope of same experimental task (Baucells et al., 2011). It is *not* the purpose of the current paper to identify a particular reference point, so the plausibility of empirical model should not be compromised even the reference points change dynamically. Thirdly, enough data are collected so that one can conduct point estimates of a non–parametric probability weighting function. We

[3]There are many different ways of defining reference points including status quo and expectations. We shall discuss in more details later. With the conventional way of eliciting certainty equivalents, decision makers are usually asked to choose between a risky prospect and a fixed amount. It then may be incorrect to define her reference point as "having neither" because she knows she will receive at least one of the two options. It is also not plausible to assume her reference point is the lottery *or* the fixed amount. The best the researcher can tell is that the reference point is unclear in this case. Subjects' reference points may indeed very heterogeneous.

are interested in answering the question that whether probability weighting is necessary to explain risk averse (loving) behavior in addition to loss aversion. Hence, imposing minimal constraints on the probability weighting function is desirable. Lastly, although we employ a between–subject design, average valuations (WTAs and WTPs) elicited from the gain and loss domains can be treated as if they are elicited from the *same* representative agent. In addition, as in all experimental studies, subjects should face enough incentive to reveal their "true" valuations over different risky prospects.

2.1 Structure

A lottery ticket and an insurance policy are used to represent gain and loss respectively[4]. In all lottery (gain) sessions, subjects are told that they do not own the $5 but they have a chance to win (receive) it; in all insurance (loss) sessions, subjects are told that they own the $5 but they face a possibility to lose it. Considering gains and losses of the $5 and their certainty equivalents (WTA or WTP) we have the following four treatments.

1. Willingness To Pay in the Gain Domain (WTPG) is the maximum amount of money that one would be willing to give up for a lottery ticket with a known probability of obtaining $5.

2. Willingness To Accept in the Gain Domain (WTAG) is the minimum amount of money that one would be willing to accept to give up her lottery ticket with a known probability of obtaining $5.

3. Willingness To Pay in the Loss Domain (WTPL) is the maximum amount of money that one would be willing to give up to purchase an insurance policy to fully protect against the loss of the $5 with a known probability.

[4]Researchers used different methods in similar experiments to represent gains and losses (Bateman et al., 1997; Eisenberger and Weber, 1995). In the design of our experiment, we only use these terms to help subjects to understand the questions.

4. Willingness to Accept in the Loss Domain (WTAL) is the minimum amount of money that one would be willing to accept in return for giving up an insurance policy that fully protects against loss of the $5 with a known probability.

The framing of the four treatments is similar to Bateman et al. (1997) although following different naming rules [5]. Other related studies framed the questions in different ways. For example, Eisenberger and Weber (1995) reports elicited WTA and WTP values for a certain lottery. Specifically, their lottery experiment had four treatments including buying, selling, short buying, and short selling. Their experimental design is essentially the same as Bateman et al. (1997) and this in valuation methods. In this study, we use more concrete terms (lottery vs. insurance policy) to help subjects to understand the questions. The novelty of this experimental design is the introduction of systematic risks and thus the examination of risk preference. Also, we use the iMPL mechanism to directly elicit valuations while Bateman et al. (1997) conducted their experiments with transactions that exchange one good for another good, as in Knetsch (1989). The iMPL mechanism allows us to generate a richer dataset.

Before diving into the discussion of reference dependence, it is useful to lay out possible gains and losses that subjects face. Since the fixed participation fee is uniform across all treatments, we do not include it in the following discussion. In the WTPG sessions, participants were given $15 to purchase a lottery ticket and in the event of winning the lottery, $5 was added to their account balance. In the WTAG sessions, participants were given $10 *and* the lottery ticket which they could sell for cash. In the WTPL sessions, participants were given $20 and they faced a risk of losing $5 of it. They were given an opportunity to purchase an insurance policy that protects them from losing the $5 using the rest $15. In the WTAL sessions, participants were given $15 and the insurance policy that they could sell for cash. Let p represent the probability of winning the lottery or the probability of a loss. Changes in wealth in each treatment are listed below in lottery form. The first lottery in

[5] The WTAG and WTPL here correspond to their Equivalent Gain (EG) and Equivalent Loss (EL) values.

each treatment represents the risky option and the second (degenerated) lottery represents the safe option.

1. WTPG: $(\$20 - WTP_{Gain}, p; \$15 - WTP_{Gain}, 1-p) vs. (\$15, 1)$
2. WTAG: $(\$15, p; \$10, 1-p) vs. (\$10 + WTA_{Gain}, 1)$
3. WTPL: $(\$15, p; \$20, 1-p) vs. (\$20 - WTP_{Loss}, 1)$
4. WTAL: $(\$10 + WTA_{Loss}, p; \$15 + WTA_{Loss}, 1-p) vs. (\$15, 1)$

In the WTPG treatment, it is natural to take the safe option –($15,1)– as the reference point so does in the WTAL treatment. Therefore, the risky prospects in these two treatments involve both a possible gain and a possible loss. However, it is not as clear what the reference points are in the WTAG and WTPL treatments up to this point.

2.2 Details

All experimental sessions were conducted in a dedicated experimental economics laboratory at a major university in North America. All subjects were recruited from university undergraduate population. Eighteen sessions, including two pilot sessions[6], were conducted. On average, a subject earned a total amount between $15 and $25 (including the value of coffee mug in the case of mug experiment) upon completion of the experiment which took about 40 minutes. Each session consists of 9 parts, only one of which was actually implemented to determine the payment. In order to determine the part implemented, a bingo ball was drawn from a bingo cage having 9 bingo balls numbered 1–9. All subjects were incentivized to reveal their true valuations in all parts since they (and the experimenter) did not know which part would be implemented until the end of the experiment[7]. In all experimental sessions, the order of the 9 parts was randomized to mediate any possible order effect.

[6]Pilot sessions are excluded from analyses due to slight difference in design but including them does not alter results. Analyses with all data are available from the author upon request.

[7]To avoid possible confusion, subjects were told to treat each part as a separate experiment but their earnings are determined by only one of the separate experiments.

Once a subject was checked into the lab, he/she was randomly seated at a private desktop computer. All computers had a Visual Basic Application (VBA) program set up and subjects submitted their decisions through the program. Subjects were given sufficient time to read instructions before the experimenter read and explained the instruction with PowerPoint slides. Subjects were encouraged to ask any question either before or during the experiment but were not allowed to communicate with each other. Subjects then submitted their valuations by accepting or rejecting the prices shown on their computer screens for all 9 parts without further direction. After all subjects finished all parts of the experiment, attention was drawn to the experimenter and a bingo ball was drawn by a lab assistant to determine which single part, of the nine parts, was the implemented part for payment. To determine the buying/selling price, a poker chip was drawn with a price listed on it from a bag by a volunteer subject in the room. After the experimenter calculated earnings, subjects were directed to complete a survey before they were paid.

[Table 1 HERE]

Table 1 shows a summary of all the experimental sessions. In half of the sessions (lottery/gain treatments), subjects were shown the experimental good and were told that they did not yet own it but could have a chance to win it if they were to buy/keep the lottery ticket. In the other half of the sessions (insurance/loss treatments), subjects were given an identical experimental good and were told it was theirs to keep although they faced a risk of losing it during the experiment. An insurance policy can be bought/kept to protect them from losing the experimental good.

[Table 2 HERE]

A full version of experimental instructions can be found in the appendix. Table 2 shows the probabilities in each part. To elicit valuations in each part, we employ a iterative Multiple Price List (iMPL) mechanism(Harrison, 2006). For each price listed on subjects computer screen[8], subjects either accept or reject that price by selecting "Yes" or "No". After all parts were completed, a poker chip with price on it was randomly drawn by a subject to determine the price of the good. In the WTPG (WTAG) and WTPL (WTAL) sessions, if a subject said yes to the drawn price, he/she pays (receives) that amount and receives (gives up) the lottery ticket or insurance policy. If a subject says no to the drawn price, he/she does not buy (keep) the lottery ticket or insurance policy. To account for possible income effect, subjects in WTPG and WTPL sessions were given $15 at the start of the session as theirs to keep or spend. Subjects in WTAG and WTAL sessions were given $10. Finally, to implement the randomly determined binding part, two stacks of white and red chips are shown to the subjects to illustrate probabilities. Depending on the probability chosen for implementation, the appropriate number of white and red chips were put into a bag. For example, if part 3 was drawn, the experimenter put 95 white chips and 5 red chips into a bag. One of the subjects then drew a chip out of the bag to determine the outcome. In a gain (lottery) session, if a red chip was drawn, those who held a lottery ticket received the experimental good. In the loss (insurance) session, if a red chip was drawn, those who did *not* hold an insurance policy lost their experimental good.

3 Results

In this section, we report results from the experiment and fit EUT to the experimental data. Table 8 at the end summarizes means and medians of valuations in each treatment of the money experiment, while Table 9 summarizes the same statistics from the mug experiment. Each column represents valuations under a different probability. Under all probabilities, including certainty, tests show that both WTAG–WTPG and WTPL–WTAL disparities are statistically significant[9]. A closer examination of Table 8 and Table 9 reveals that the mean

[8]See screenshot in appendix.

[9]Under certainty, in the money experiment, Wilcoxon-Mann-Whitney test results in z=-4.21 with p=0.00 for WTAG-WTPG and z=-5.416 for WTAG-WTPL; in the mug experiment, Wilcoxon-Mann-Whitney test results in z=-3.24 with p=0.00 for WTAG-WTPG and z=-3.22 for WTAL–WTPL.

WTAG–WTPG and WTAL–WTPL disparities are approximately constant while the ratios increases when probabilities decrease. In fact, constant absolute disparities are found among the four different values under intermediate probabilities ($0.05 \leq p \leq 0.95$). In the money experiment, the ratio of (mean) WTAG/WTPG is 3.87 at $p = 0.05$, in contrast to a ratio of 1.08 under certainty, while these ratios are 3.62 versus 1.46 in the loss domain. In the mug experiment, the ratio of (mean) WTAG/WTPG is 8.63 at $p = 0.05$ versus 1.60 under certainty while the same ratios are 3.88 versus 1.61 in the loss domain. The ratios under small probabilities are especially high for a private good (Horowitz and McConnell, 2002). The ratios under all probabilities are also shown in Figure 1.

[Figure 1 HERE]

3.1 Fitting EUT of Wealth Model

Given the experimental design, if there is no income effect, EUT would predict identical valuations from all four different treatments. We explicitly controlled for income effect in the experiment by endowing subjects with different amount of money. We now show that it is not possible for *Expected Utility of Wealth* to explain the experimental data.

To start, let's assume a DARA (decreasing absolute risk aversion) utility function $u(\cdot)$ with $u^{'}(\cdot) > 0$ and $u^{''}(\cdot) \leq 0$. Subjects' decision rules are shown below under expected utility. Note, x^0 represents the subjects wealth before they start the experiment. x^0 is assumed to be (on average) the same across different treatments.

$$pu(x^0 + 20 - WTP_{Gain}) + (1-p)u(x^0 + 15 - WTP_{Gain}) = u(x^0 + 15) \tag{1a}$$

$$pu(x^0 + 15) + (1-p)u(x^0 + 10) = u(x^0 + 10 + WTA_{Gain}) \tag{1b}$$

$$pu(x^0 + 15) + (1-p)u(x^0 + 20) = u(x^0 + 20 - WTP_{Loss}) \tag{1c}$$

$$pu(x^0 + 10 + WTA_{Loss}) + (1-p)u(x^0 + 15 + WTA_{Loss}) = u(x^0 + 15) \tag{1d}$$

Following concavity of the utility function, $WTP_{Gain} \leq 5p, WTA_{Gain} \leq 5p$ and $WTP_{Loss} \geq$

$5p, WTA_{Loss} \geq 5p$. Using the DARA assumption on the Bernoulli utility function, it can be shown that expected utility on final wealth predicts $WTA_{Gain} \leq WTP_{Gain} \leq 5p \leq WTA_{Loss} \leq WTP_{Loss}$ (see appendix for details). This prediction is not consistent with the experimental result shown in Table 8.

3.2 Risk Aversion at Small

As a response to the "Rabin Critique", many have argued that expected utility should not be of final wealth but of income (Cox and Sadiraj, 2006; Holt and Laury, 2002)[10]. Moreover, the *expected utility of income* model has been a popular tool to analyze empirical data because initial wealth are usually not observed by researchers. This practice is even more common in analyzing laboratory data. Regardless the practical issues of using EUT of income model (Safra and Segal, 2008), we show that our empirical result does not allow such a model either.

If expected utility is defined on income, control of income effect becomes irrelevant and so does the initial wealth x^0. The decision rule then becomes Equation 2.

$$pu(5 - WTP_{Gain}) + (1-p)u(-WTP_{Gain}) = u(0) \tag{2a}$$

$$pu(5) + (1-p)u(0) = u(WTA_{Gain}) \tag{2b}$$

$$pu(0) + (1-p)u(5) = u(5 - WTP_{Loss}) \tag{2c}$$

$$pu(WTA_{Loss}) + (1-p)u(5 + WTA_{Loss}) = u(5) \tag{2d}$$

[10] Cox and Sadiraj (2006) argued that expected utility should be defined on both wealth and income. A line of research that follows that argument leads to the notion of "asset integration". While it is beyond the scope of this paper to test the "asset integration" model, our experimental data is also inconsistent with prediction from such a model

When valuations are small, we can take a first order approximation of Equation 2.

$$pu(5 - WTP_{Gain}) - pu(-WTP_{Gain}) \approx u'(-WTP_{Gain})WTP_{Gain} \tag{3a}$$

$$pu(5) + pu(0) \approx u'(0)WTA_{Gain} \tag{3b}$$

$$pu(0) + pu(5) \approx u'(5)(-WTP_{Loss}) \tag{3c}$$

$$pu(WTA_{Loss}) - pu(5 + WTA_{Loss}) \approx u'(5 + WTA_{Loss})(-WTA_{Loss}) \tag{3d}$$

Thus, $\frac{u'(0)}{u'(-WTP_{Gain})} = \frac{u(5)-u(0)}{u(5-WTP_{Gain})-u(-WTP_{Gain})}\frac{WTP_{Gain}}{WTA_{Gain}} \leq \frac{WTP_{Gain}}{WTA_{Gain}}$ because $u' > 0$ and $u'' \leq 0$. Similarly, we have $\frac{u'(5+WTA_{Loss})}{u'(5)} \leq \frac{WTP_{Loss}}{WTA_{Loss}}$. With these results, one immediately notices a kink between utility increment and decrement at income of \$0 and \$5 unless $WTA = WTP$. Because the amount of \$5 is arbitrarily chosen, this empirical result applies to any point on the utility function within a reasonable range[11]. Admittedly, people may incur irregular behaviors under small probabilities due to risk dismissal and other factors, but the marginal utility of money cannot decrease at such a high rate. Even income is not additive to wealth, experimentalists would have a very hard time to imagine that the seventh dollar they pay to their subjects is only $\frac{1}{4}$ effective as the fifth dollar they pay[12]. If the marginal utility of money decreases at a rate of 4 at \$1 interval, any payment beyond a very small amount becomes practically immaterial. This is apparently not the case in either the laboratory or the real world, hence an EUT of income prediction based on concavity of Bernoulli utility function leads to absurdity. This simple "empirical calibration" practice again emphasizes the importance of systematically examining behavior under different probabilities.

Note that the above argument applies to both EUT of wealth and EUT of income. In the EUT of wealth case, a slight risk aversion with small amount at stake leads to "Rabin Critique". In the EUT of income case, a much higher degree of loss aversion is required to show the implausibility. Our experimental result provides such evidence. Based on our

[11] Given the nature of laboratory experiment and incentives used, we do not extend this conclusion to larger amounts.

[12] In the loss domain, when $p = 0.05$, $\frac{u'(5+0.93)}{u'(5)} \leq \frac{1}{3.62}$

analyses and previous research, we can safely conclude that neither EUT of wealth nor EUT of income provides a satisfying explanation to the experimental data.

3.3 Other Stylized Facts

Given that both EUT of wealth and EUT of income do not provide a reasonable explanation of the observed risk aversion, we will turn to a behavioral model to explain the experimental data in the next section. Before doing so, we first report some stylized facts in the experimental data that help us to identify the key components in the model. Mean valuations elicited from all sessions are plotted in Figure 2. For medium probabilities, valuations appear to increase at similar rates (lines share the same slope). These parallel lines suggest valuations under risk responds to probability proportionally. Another noteworthy observation is that the WTAG and WTPL values from the money experiment appear to be the same under all probabilities.

[Figure 2 HERE]

The linear regression shown in Table 3 uses treatment dummy variables and change in probabilities to explain the valuations. For both the money and mug experiments, we first estimate a linear regression model with all of the collected data. We then estimate another linear regression which only uses data elicited under non extreme probabilities ($0.05 \leq p \leq 0.95$). All models include probability as an explanatory variable and its interactions with the treatment dummy variables. Probability is treated as a continuous variable although we elicited values under nine probabilities. F test on the three interaction terms (WTAG*Probability, WTPL*Probability, and WTAL*Probability) in all models suggests that we cannot reject the hypothesis that they are jointly equal to zero[13]. In other words, we cannot reject the hypothesis that the four lines share the same slope suggesting statistically same response to changes in risk.

[13]Model (1) results in F=0.59 with p=0.62. Model (2) results in F=0.80 with p=0.50. Model (3) results in F=1.45 with p=0.23. Model (4) results in F=0.76 with p=0.51.

[Table 3 HERE]

On the other hand, all treatment dummy variables are statistically different from each other at 1 percent level except for WTAG and WTPL in the money experiment[14]. This suggests that although the four lines in Figure 2 have statistically identical slopes, they do not overlap with each other except for WTAG and WTPL in the money experiment. This contradicts the prediction of EUT but is consistent with previous observed WTA–WTP disparities. A different set of linear regressions but with fixed effects are reported in the bottom panel of Table 3. The fixed effects are at individual level so all intercepts of the four lines are not separately identified since there is only between subject variation in terms of elicitation mode. The estimated slopes are identical to the ones reported in the top panel so they are suppressed from the table. This regression shows that individual level fixed effects can explain at least 50% of the remaining variance after accounting for responses to probabilities and thus indicates some heterogeneity in individual valuations.

4 Structural Analysis

The structural analysis illustrates that how loss aversion explains high degrees of risk aversion in the experiment. We also go beyond the explanation of risk aversion and ask what is the best *existing* model that can explain the experimental data in general and whether the estimated parameters are consistent with literature. The reduced–form analysis shows some degrees of loss aversion as well as a linear response of valuations to risk changes. Many studies after Rabin (2000) have shown that a nonlinear utility function of either wealth or income is not immune to the calibration problem (Barberis et al., 2006; Cox and Sadiraj, 2006; Rubinstein, 2006; Safra and Segal, 2008, 2009). Considering our empirical finding and previous studies, a linear utility function with a single loss aversion parameter is used in the structural analysis. One may then question the different degrees of loss aversion under

[14] A test of WTAG=WTPL gives F=2.92 (p=0.09) in model (1) and F=0.12 (p=0.73) in model (2).

different probabilities. We argue that probability weighting then comes into play. Following the KR model (Köszegi and Rabin, 2006, 2009), we use an empirical model that features both loss aversion and rank–dependent probability weighting to explain the experimental data. The model is flexible enough to accommodate both expected utility theory and prospect theory as special cases. It is then estimated with different imposed constraints to demonstrate the need of both loss aversion and probability weighting.

4.1 Loss Aversion and Possible Reference Points

Many empirical research suggest loss aversion (Barberis et al., 2001; Benartzi and Thaler, 1995; Coursey et al., 1987; Kahneman et al., 1991; Siegel and Thaler, 1997) and many theoretical models have formulated the concept. For example, Sugden (2003) incorporated reference dependence into subjective expected utility theory and showed that the theory implies WTA–WTP disparities in lottery valuation. Köszegi and Rabin (2006, 2007, 2009) developed a model (KR model hereafter) based on reference dependent preferences and use it to explain risk attitudes and consumption plans. KR model argues that people receive utility from *both* consumptions and changes of consumptions. Their original formulation is shown in the following equation.

$$\begin{aligned} U(F|r) &= \int u(c|r)dF(c)^{15} \\ u(c|r) &= m(c) + n(c|r) \end{aligned} \tag{4}$$

Where c is consumption bundle, r is reference points and F is the probability measure which c is drawn from. $m(c)$ represents the intrinsic consumption utility and $n(c|r)$ represents the gain–loss utility. They also proposed that rational expectations as reference points. Although, theoretically, endogenous reference point is appealing because it addes 0 degree of freedom to EUT[16], it may fail to describe real behavior. As Kahneman et al. (1991)

[16]Rational expectation as reference points are usually considered a major contribution of the KR model. To be more specific, they argued *"... a person's reference point is her probabilistic beliefs about the relevat consumption outcome held between the time she first focused on the decision determining the outcome and*

argued *"...the refernce state usually corresponds to the decision maker's current position, it can also be influenced by aspiration, expectations, norms and social comparisons..."*, the identification of the reference point is largely an empirical question and researchers have found different reference points in different settings (Crawford and Meng, 2011; Ericson and Fuster, 2011; Fehr and Tyran, 2008; Hart and Moore, 2008). We discuss several different possible reference points in our experiment including the rational expectation. We then relate them to the empirical results and show some of them are not consistent with our empirical evidence.

First consider rational expectation. We assume people are narrow bracketing when they determine their valuations. Specifically, when they accept or reject a certain amount of money as WTA or WTP, they *only* consider the outcome of that particular choice[17]. Under the narrow bracketing assumption, rational expectations are shown in the last two columns in Table 4. When a decision maker making herself a plan to answer "no" or "yes" to a certain amount, she knows how she will feel after her decision is implemented. If she rejects an amount, she will keep what she is given; and if she says yes, she will receive or give up the good. If she chooses the safe option (without lottery/with insurance), there will be no gain–loss utility associated with that choice because she already knows exactly what will happen then the outcome immediately becomes the reference point. If she chooses the risky option, either the lottery or risky situation without insurance becomes the reference point. Since there is risk associated with the realized option, she will receive gain–loss utility based on her expectation of the final outcome.

---

shortly before consumption occurs". This means once a decision maker figures out a consumption plan, the reference point is also determined so the model requires no extra parameter compared with EUT.

[17]When determining their binding decisions, a random draw from a uniform distribution is used as described in section 2. However, under narrow bracketing, when subjects answer "yes" and "no" to each different amount, they treat their answer *as if it is chosen*. In this sense, most if not all laboratory experiment assumes narrow bracketing at a certain level by ignoring background risks. See Köszegi and Rabin (2006); Rabin and Weizsäcker (2009) for discussions on narrow bracketing.

Now, consider a KR style model under risk.

$$U(x, x^0) = \sum_{i=1}^{N} p_i[m(x_i) + n(x_i - x^0)] \tag{5}$$

p_i is the probability of state i. x_i is the consumption level of x in each state and x^0 is the reference point. Let $m(x_i) = x_i$ and assume a two–part linear form for the gain–loss utility function $n(x_i - x^0) = \begin{cases} \eta(x_i - x^0), & if x_i \geq x^0 \\ \lambda\eta(x_i - x^0), & if x_i \leq x^0 \end{cases}$. Subjects decision rules under KR model are shown below. Let $\Delta x = 5$ be the possible gain or loss.

$$p[(\Delta x - WTP_{Gain}) + \eta(1-p)\Delta x] + (1-p)[(-WTP_{Gain}) + \eta p(-\lambda x)] = 0 \tag{6a}$$

$$p[\Delta x + \eta(1-p)\Delta x] + (1-p)[\eta p(-\lambda \Delta x)] = WTA_{Gain} \tag{6b}$$

$$p[-\Delta x + \eta(1-p)(-\lambda \Delta x)] + (1-p)[\eta p \Delta x] = \Delta x - WTP_{Loss} \tag{6c}$$

$$p[(-\Delta x + WTA_{Loss}) + \eta(1-p)(-\lambda \Delta x)] + (1-p)[WTA_{Loss} + \eta p \Delta x] = 0 \tag{6d}$$

Through some manipulation, the predictions are shown below.

$$WTP_{Gain} = p\Delta x + \eta p(1-p)(1-\lambda)\Delta x \tag{7a}$$

$$WTA_{Gain} = p\Delta x + \eta p(1-p)(1-\lambda)\Delta x \tag{7b}$$

$$WTP_{Loss} = p\Delta x - \eta p(1-p)(1-\lambda)\Delta x \tag{7c}$$

$$WTA_{Loss} = p\Delta x - \eta p(1-p)(1-\lambda)\Delta x \tag{7d}$$

One immediately notices that the *fully* rational expectations as reference points under narrow bracketing is inconsistent with our empirical results. This observation is consistent with the experimental results reported by Heffetz and List (2014). Köszegi and Rabin (2006) also recognized the limitation of this practice and it could be the case that subjects in our experiment do not have enough time to form a fully rational expectation. Table 4 also lists two *exogenously* determined reference points. Given the relatively short period in the laboratory experiment, exogenously determined reference points are more realistic. The first column is endowment which has been widely used in literature. The second reference point is the status-quo[18].

In the WTPG and WTAL treatments the endowment and status-quo are the same. In the WTAG and WTPL treatments, endowment means one faces a possible gain or possible loss. It is, again, quite arbitrary to assume that subjects fully perceive their endowment given the short time frame in the experiment. For this reason, we argue that the endowment and status quo are the two boundaries of subjects' reference points. To approximate this, we assume the reference point is a linear combination (in probability) of endowment and status-quo. That is, in the money experiment, the reference point in the WTAG treatment is $(x^0 + \Delta x, ap; x^0, 1 - ap)$ with $0 \leq a \leq 1$ (relative weighting on endowment hereafter) and the reference point in WTPL treatment is $(x^0, ap; x^0 + \Delta x, 1 - ap)$. Reference points in the mug experiment are defined in a similar manner.

[Table 4 HERE]

4.2 Rank Dependent Probability Weighting

To resolve the inconsistent degrees of loss aversion under different probabilities, we employ another important feature in many alternative theories to EUT including prospect theory and rank–dependent expected utility theory. Probability weighting argues that people do not behave according to the given objective probabilities. Instead, they act as if they overweight some probabilities and underweight other probabilities. The behavioral economic literature has focused less on probability weighting partly because of the difficulty in empirical identification. Our laboratory experiment systematically varies probabilities and made them clearly known to subjects. Therefore, any observed probability distorting behavior should

[18] We are using this name with risk of confusion. "Status quo" here refers to the status where people think they are. People usually do not feel they own the prize even when they hold a lottery ticket; people usually do not feel a loss even when they do not hold an insurance policy (a good example is driving a car without comprehensive insurance).

be explained by probability weighting instead of systematic risk misperception.

Empirical research shows behavior deviates systematically from objective probabilities (Camerer and Ho, 1994; Gonzalez and Wu, 1999; McClelland et al., 1993). To account for anomalies such as Allais Paradox within the framework of expected utility, researchers raised the idea of probability weighting even before prospect theory (Handa, 1977; Karmarkar, 1978). These studies, including the original paper on prospect theory, all seek to transform single outcome probabilities into weighted or edited decision weights. However, this approach was criticized for violating a fundamental rule–stochastic dominance (Bawa, 1975; Hadar and Russell, 1969). Quiggin (1982) proved that there does not exist any non–identity probability weighting function that transforms single probabilities without violating stochastic dominance. He then proposed to transform the cumulative probabilities. That is, decision weights are derived from the entire probability distribution instead of single probabilities. This approach preserves stochastic dominance while still accounting for probability distorting behaviors. Since Quiggin's work, many researchers have contributed to the axiomatization of rank dependence theory (Abdellaoui, 2002; Machina, 1987, 2009; Schmeidler, 1989). To account for probability distorting behavior, we replace the probability measure in Equation 5 with decision weights in the model shown in Equation 8,

$$U(x, x^0) = \sum_{i=1}^{N} \pi_i[m(x_i) + n(x_i - x^0)] \tag{8}$$

π_i is the decision weight on state i. x_i is the consumption level of x in each state and x^0 is reference point. Note, we have yet defined or imposed any constraints on π_i.

To accommodate possible probability weighting, while not violating stochastic dominance, we impose rank dependence on the probability weighting function $\pi_i(\cdot)$. Following Tversky and Kahneman (1992), S is a finite set of states of nature. A capacity W is a function that assigns to each $A \subset S$ a number $W(A)$ satisfying $W(\emptyset) = 0$, $W(S) = 1$ and $W(A) \geq W(B)$ whenever $A \supset B$. The decision weight π_i, associated with a positive outcome, is the difference between the capacities of the two events "the outcome is at least as good as x_i," and "the outcome is strictly better than x_i". The decision weight π_i, associated with a negative outcome, is the difference between the capacities of the two events, "the outcome is at least as bad as x_i" and "the outcome is strictly worse than x_i".

In a binary-outcome lottery setting, this is easy to implement. For example, in the WTPG treatment, the decision weight on the winning state is W("Win \$5")=$\pi(p)$ while the decision weight on the non-winning state is W("Win \$0")–W("Win \$5")=$1-\pi(p)$[19]. Decision weights in other treatments can be defined in the same way.

4.3 Empirical Model

To explain our empirical data, we use a behavioral model similar to KR model but adding rank–dependent probability weighting. Also, exogenously determined reference points are used because the endogenous reference point is not consistent with our experimental results; Equation 9 shows the decision equations in the money experiment using the model presented in Equation 8. The left hand side is the utility from "default" or endowment and the right hand side is the utility from buying or selling the endowment. A subject should be indifferent between the left hand side and the right hand side at the amount where she switches from "yes" to "no" or vice verser. Note that we replace Δx with $\bar{u}$ to represent utility increment of receiving the \$5. The utility of winning the lottery can be different from receiving \$5 so

[19] W is the capacity and the worst outcome for subjects who purchase a lottery ticket is to win \$0 so its capacity is 1

we also take it into account.

$$\begin{aligned} x^0 =&\pi(p)[x^0 - WTP_{Gain} + \bar{u} + \eta(-WTP_{Gain} + \bar{u})] \\ &+[1-\pi(p)][x^0 - WTP_{Gain} + \lambda\eta(-WTP_{Gain})] \end{aligned} \tag{9a}$$

$$\begin{aligned} &\pi(p)[x^0 + \bar{u} + (1 - a\pi(p))\eta\bar{u}] + [1-\pi(p)][x^0 + a\pi(p)\eta(-\lambda\bar{u})] \\ &=x^0 + WTA_{Gain} + \eta[a\pi(p)\lambda(WTA_{Gain} - \bar{u}) \\ &\quad +(1 - a\pi(p))(WTA_{Gain})] \end{aligned} \tag{9b}$$

$$\begin{aligned} &\pi(p)[x^0 + (1 - a\pi(p))\eta(-\lambda\bar{u})] + [1-\pi(p)][x^0 + \bar{u} + a\pi(p)\eta\bar{u}] \\ &=x^0 + \bar{u} - WTP_{Loss} + \eta[a\pi(p)(\bar{u} - WTP_{Loss}) \\ &\quad +(1 - a\pi(p))(-\lambda WTP_{Loss})] \end{aligned} \tag{9c}$$

$$\begin{aligned} x^0 + \bar{u} =&\pi(p)[x^0 + WTA_{Loss} + \eta\lambda(WTA_{Loss} - \bar{u})] \\ &+[1-\pi(p)][x^0 + \bar{u} + WTA_{Loss} + \eta(WTA_{Loss})] \end{aligned} \tag{9d}$$

The relative weight η plays a critical role in the model. When $\eta = 0$, the model degenerates to (rank dependent) EUT. In other words, the weight of gain–loss utility becomes zero. When $\eta \to \infty, \beta = \lambda$, agents only receive gain–loss utility. Therefore, these two special cases are EUT and prospect theory respectively. Predicted valuations are shown in Equation 10. The derivation of the valuations in the mug experiment requires considering a consumption good and money simultaneously. The derivation is included in the appendix.

$$\begin{aligned} WTP_{Gain} =& \frac{(1+\eta)\pi(p)\bar{u}}{1+\lambda\eta+(1-\lambda)\eta\pi(p)} = \frac{\bar{u}}{1+\frac{1+\lambda\eta}{1+\eta}(\frac{1}{\pi(p)}-1)} \\ =& \frac{\bar{u}}{1+\beta(\frac{1}{\pi(p)}-1)} \end{aligned} \tag{10a}$$

$$WTA_{Gain} = \pi(p)\bar{u} \tag{10b}$$

$$WTP_{Loss} = \pi(p)\bar{u} \tag{10c}$$

$$\begin{aligned} WTA_{Loss} =& \frac{(1+\lambda\eta)\pi(p)\bar{u}}{1+\eta+(\lambda-1)\eta\pi(p)} = \frac{\bar{u}}{1+\frac{1+\eta}{1+\lambda\eta}(\frac{1}{\pi(p)}-1)} \\ =& \frac{\bar{u}}{1+\frac{1}{\beta}(\frac{1}{\pi(p)}-1)} \end{aligned} \tag{10d}$$

In Equation 10, λ and η cannot be identified separately but $\beta = \frac{1+\lambda\eta}{1+\eta}$ can be estimated as a single parameter. Note that a is irrelevant in predicting the valuations in the money experiment. On the one hand, this model with exogenous reference point is robust to the specification of reference point when only money is involved. On the other hand, a is not identified with data from the money experiment. In the two dimension case, however, different values of a can lead to different predictions. Equation 11 shows valuations derived for the mug experiment.

$$WTP_{Gain} = \frac{\pi(p)(1+\eta)\bar{m}}{1+\lambda\eta} = \pi(p)\bar{m}\frac{1+\eta}{1+\lambda\eta} = \frac{\pi(p)\bar{m}}{\beta} \tag{11a}$$

$$\begin{aligned} WTA_{Gain} =& \pi(p)\bar{m} + \frac{a\pi(p)^2(\lambda-1)\eta\bar{m}}{1+\eta} = \pi(p)\bar{m} + a\pi(p)^2\bar{m}(\frac{1+\lambda\eta}{1+\eta}-1) \\ =& \pi(p)\bar{m}[\beta a\pi(p) - a\pi(p) + 1] \end{aligned} \tag{11b}$$

$$\begin{aligned} WTP_{Loss} =& \pi(p)\bar{m} + \frac{a\pi(p)^2(1-\lambda)\eta\bar{m}}{1+\lambda\eta} = \pi(p)\bar{m} + a\pi(p)^2\bar{m}(\frac{1+\eta}{1+\lambda\eta}-1) \\ =& \pi(p)\bar{m}[\frac{1}{\beta}a\pi(p) - a\pi(p) + 1] \end{aligned} \tag{11c}$$

$$WTA_{Loss} = \frac{\pi(p)(1+\lambda\eta)\bar{m}}{1+\eta} = \pi(p)\bar{m}\frac{1+\lambda\eta}{1+\eta} = \beta\pi(p)\bar{m} \tag{11d}$$

4.4 Estimation

We conduct two groups of estimations with the theoretical predictions. In the first group, we rely on the predictions in Equation 10 and use the data collected from the money experiment. This group of estimation allows us to examine the value of β and the shape of the nonparametric probability function. In the second group, we conduct a joint estimation using predictions from both Equations 10 *and* 11. Data collected from both the money and mug experiments are used. We can examine the property of the exogenous reference point with this second group of estimation.

To form an econometric model, we assume the stochastic component in valuations is introduced at the final stage of decision making for two reasons. Firstly, unlike many of previous studies that rely on the random utility model (e.g. Holt and Laury (2002)), we can observe the empirical distribution of the valuations so we are more confident in our distribution assumption of the error term. Secondly, the assumption is that the behavioral model *fully* determines the valuations and there is no unobservable that *systematically* alters valuations. The only source of deviation from theoretical prediction is represented by an additive error term to the valuations. Other individual characteristics may also affect the valuations but we assume, on average, they play no role.

$$\text{Switching Price}_j = \text{Valuation} + \nu_j, \nu_j \sim N(0, \sigma) \tag{12}$$

Where j denotes individuals and "Valuation" is the theoretical prediction from the behavioral model. ν_j is an additive error term that follows a normal distribution. The non-linear (in parameters) nature of the model requires global optimization when using maximum likelihood to estimate the parameters. Since the model has been specified in a way that it can take EUT as a special case, EUT is thus nested in the model. Hence, we are able to test whether introducing loss aversion and probability weighting is necessary at the cost of parsimony.

Table 5 shows the first group of estimates. (1) estimates a model without incurring probability weighting by imposing the restriction $\pi(p) = p$. (2) estimates the unrestricted model with both loss aversion and probability weighting. Given the large number of parameters in the unrestricted model, we employed two different estimating techniques to ensure the estimates are at or close enough to optimality. (2) uses the maximum likelihood method while (3) uses the Markov Chain Monte Carlo (MCMC) method. The consistency between (2) and (3) provides confidence in these estimates. The comparison between (1) and (2) reveals the need of probability weighting to explain the experimental data. Likelihood ratio test reports a score of 239.6 and thus the null hypothesis "no probability weighting" is rejected with a p value of 0.00. Also note that the estimated parameter $\beta = \frac{1+\lambda\eta}{1+\eta}$ is always statistically greater than 1. This again verifies that $\eta > 0$ and confirms loss aversion in preference.

Most previous works assumed a functional form for the probability weighting function, we treat the decision weight for each probability as a separate parameter so the estimated probability weighting function is non–parametric. Estimation suggests the marginal utility is about 7.8 for a \$5 *prize*. The parameter $\beta = \frac{1+\lambda\eta}{1+\eta}$ is 2.6. Figure 3 shows the predicted valuations and the non-parametric probability weighting function from the unrestricted model. Valuations predicted by the behavioral model mimics the experimental data well and the probability weighting function replicates the inverse S-shaped curve found in literature.

[Table 5 HERE]

The parameters for the joint estimation are estimated in the same ways. We restrict the probability weighting function $\pi(p)$ to be the same across different commodities but allow other parameters to vary. Table 6 reports the estimates. The estimated loss aversion parameter is about 1.5 for the mug data. With this joint estimate, we are also able to pin down a which is the parameter in reference point specification. a is estimated to be between 0.5 and 0.6. This estimate is consistent with our conjecture that subjects may not fully take endowment as the reference point given the short period they are allowed to posses the lottery or insurance in the laboratory. This also alerts us that using endowment or status quo as reference points in empirical work may be problematic.

[Table 6 HERE]

[Figure 3 HERE]

[Fiture 4 HERE]

4.5 Validation

Increasing the number of parameters in a model leads to better fit but at the cost of parsimony. We validate the estimated model by performing a calibration using part of the sample while using the calibrated model to predict the rest of the sample. We conduct this validation process for the money and mug experiments separately. Table 7 compares four different models including expected value with σ (the error term) as its only parameter, EUT with an extra parameter of $\bar{u}$ or $\bar{m}$[20], reference dependent utility model without probability

[20]Assuming linear form of $u(\cdot)$.

weighting[21] and the proposed model in this paper. Log-likelihood function values (LL) are reported for both the calibration and the validation subsamples. The calibration of models also reports the Akaike information criteria (AIC)[22] and Bayesian information criteria (BIC)[23]. The validation of the models reports mean squared error (MSE)[24] and a pseudo R^2 value[25]. The calibration uses 70% of the sample and the validation uses the rest 30% of the sample. To do this, 70% of the subjects within *each* treatment are randomly drawn to form a calibration sample. The rest of the sample in each treatment are combined together to form the validation sample. This way of drawing observations guarantees balanced observations among different treatments in both the calibration and validation samples.

On the one hand, the calibration practice shows that the proposed reference–dependent utility model with rank–dependent probability weighting performs the best. Note that probability weighting functions that have fewer parameters are likely to perform better. For example, both Prelec (1998) and Tversky and Kahneman (1992) suggested inverse S-shaped probability weighting functions that have only one parameter. However, the purpose of this practice is to show that probability weighting is *necessary* to explain risk averse behavior under risk so we do not discuss specific functional forms of the weighting function. On the other hand, the validation practice confirms the explanatory power of the behavioral model. The predictions made by the model are the best among the four models after correcting for degrees of freedom.

[Table 7 HERE]

[21] The model used is the one in Equation 5 with linear and two-part linear assumption on $m(\cdot)$ and $n(\cdot)$.

[22] $AIC = -2LL + 2k$ where k is the number of parameters.

[23] $BIC = -2LL + k \cdot ln(LL)$.

[24] $MSE = \frac{(Bid-Prediction)^2}{No.of Observations}$

[25] $Pseudo R^2 = 1 - \frac{LL_{Model} - k}{LL_{EV}}$. This is the McFaddens definition of pseudo R^2 which is an indication of improvement of the full model from the intercept model. Here the intercept model is the expected value.

4.6 Robustness and Discussion

The estimates from the unrestricted model fall into reasonable ranges including a loss aversion parameter that is greater than 1 and a probability weighting function that overweights small probabilities and underweights large probabilities. The predictions from the estimated model also recover several important features of the raw data. First of all, it predicts the identical valuations for the WTAG and WTPL treatments in the money experiment as well as the result that these two values are different in the mug experiment. Secondly, as shown in Figure 3 the four lines from the money experiment are mostly parallel with each other. Lastly, although the four lines in Figure 4 are not quite parallel, the two WTP lines and two WTA lines are parallel respectively. This echoes the results of the reduced form analysis in Table 3. In addition to the plausibility of the model, there are two noteworthy observations.

1. The estimated parameter $\beta = \frac{1+\lambda\eta}{1+\eta}$ itself can be taken as a "revealed" loss aversion parameter because β is increasing in both λ and η. Note that β is larger in the one dimension (money) case than in the two dimension (mug) case. If the behavioral model is correctly specified, this implies a larger weight (η) on gain-loss utility in the one dimension case.

2. The decision weight under certainty $\pi(1)$ is statistically different from 1 from the money data.

Observation 1 suggests that subjects in the money experiment generally care more about the gain–loss utility while subjects in the coffee mug experiment also care about the intrinsic consumption utility. Many studies on rank dependent theory imposed the restrictions that $\pi(0) = 0$ and $\pi(1) = 1$ (Quiggin, 1982) while there are also other rank dependent models that impose no such restrictions (Tversky and Kahneman, 1992). Kahneman and Tversky (1979) argued that "*Although* $\pi(p) > p$ *for low probabilities, there is evidence to suggest that, for all* $0 < p < 1$, $\pi(p) + \pi(1-p) < 1$. *We label this property subcertainty (p. 281)*". They also noted that people tend to ignore very small probabilities, so we should also expect the weighting behavior under very small probabilities to be less predictable in empirical works. Note that imposing $\pi(1) = 1$ is essentially restricting the valuations in the money experiment to be identical under certainty. This can be seen directly from Equation 9. Our experimental results can be seen as a contradiction to this restriction (given our model is correctly specified). There is, however, another possible explanation to this observation – heterogeneity. The significant individual level fixed effects in the reduced form analysis is an indication.[26]

To account for possible heterogeneity, we estimate a different set of parameters for the money experiment which is shown in the fourth column of Table 5. This specification assumes individuals possess different intrinsic consumption values for the \$5 prize and their consumption values are normally distributed. Combining the MCMC method with Gauss quadrature[27], (4) in Table 5 shows these estimates. Considering heterogeneity in consumption value ($\bar{u}$) reduces variance of econometric error term. It also increases the decision weight under certainty –$\pi(1)$– from 0.77 to 0.90. In addition, Figure 5 plots the model predictions and probability weighting function for this set of parameters. This result shows that, when considering heterogeneity in $\bar{u}$, the estimated parameters are more consistent with the ones reported in literature, especially the weighting function[28].

[Figure 5 HERE]

Formation of reference points is largely an empirical question. It may be natural and convenient to take the endowment or status quo as the reference point in many cases but this strategy is not a panacea. We form our behavior model is a way to avoid the identification of particular reference points and hope our lacking of knowledge on the reference points does

[26]Another indication is that, under certainty, the distribution of valuation from money experiment has fat tails. This is also demonstrated by the divergence of mean and median values. We do not exclude any observation from the experimental dataset by imposing "rational" constraints. There is no reason for us to believe "irrational" subjects (e.g. those who bid more than \$5) are not part of the population. Also remember the subjects are undergraduate students at a research university, if there is any representation bias, these subjects tend to be more "rational" than the general population.

[27]20 points used to approximate the value of integral.

[28]There are possible heterogeneity in other parameters too such as λ, η, and the probability weighting function.

not hinder us from testing the model. From the joint estimation shown in Table 6, it is likely that neither the status quo nor the endowment is the actual reference point. It is also possible that individuals are heterogeneous in terms of their reference points. This issue definitely calls for further investigation.

Before ending the current section, we also mention recent studies that use mixture model to explain choices under risk based on the argument that considerable heterogeneity exists both within and between individuals (Conte et al., 2011; Harrison and Rutström, 2009). Identification strategies in these studies rely heavily on the chosen functional form of the behavioral models. For example, Harrison and Rutström (2009) used a CRRA (constant relative risk aversion) utility function to represent EUT in their estimation and indicated that the risk aversion parameter is critical for the identification. Although this practice is quite useful in recognizing individual level heterogeneity, such an approach is sensitive to errors in selecting the functional forms[29]. Instead, we assume a uniform theory that explains behaviors of all individuals in the population and make as few assumptions as possible regarding functional forms. Admittedly this approach ignores heterogeneity at the individual level due to identification requirements, it however tests if a single theoretical model can reasonably explain risk averse (loving) behaviors.

5 Conclusion

Risk aversion is one of the most important empirical evidence that we have observed in decision making under risk and uncertainty. Many studies have suggested that a concave Bernoulli utility function is not a satisfactory explanation to this prevalent phenomenon. In this paper, we use a laboratory experiment to study risk averse behavior and seek an empirical model to explain the experimental results. By eliciting valuations of risky prospects, we show that both EUT of wealth and EUT of income fail to explain risk aversion even when

[29] In this particular example, it is implausible to assume a universal risk aversion parameter and fit it to laboratory experimental data as we showed in the result section.

income is only at small scale. In addition, the experimental result also suggests reference dependence within the full range of probabilities and the reference dependent behavior is more prominent under small probabilities. Hence, we employed loss aversion and probability weighting – both from prospect theory – to explain the observed patterns. Although it has been more than three decades since prospect theory was firstly proposed by Kahneman and Tversky(Kahneman and Tversky, 1979), we have yet to reach a consensus on taking loss aversion and probability weighting as the reasons behind risk averse behavior. We hope the current study provides more empirical evidence to support this view.

Importantly, assuming a linear utility function, our experimental data allows one to estimate this model without imposing any additional constraint, in particular on the probability weighting function. In the structural analysis, the role of probability weighting has been emphasized in addition to loss aversion suggested by Rabin (2000). Estimation shows that subjects' behavior is consistent with the empirical model and key parameters are within the range found in the literature. Results suggest people show both loss aversion and probability weighting when making decisions under risk and the magnitude of both is empirically significant. The success of this empirical model in explaining the experimental data again demonstrates the role of loss aversion and probability weighting in explaining risk averse behavior and decisions under risk and uncertainty in general.

This study, being the first one that uses endowment effect under a set of different probabilities, complements the literature studying risk aversion in the laboratory. We provide an empirical component to the "Rabin Critique" and other related theoretical studies. The experimental design created an environment that allows one to thoroughly examine the roles loss aversion and probability weighting in an empirical setting.

References

Abdellaoui, M. (2002). A Genuine RankDependent Generalization of the Von Neumann-Morgenstern Expected Utility Theorem. *Econometrica*, 70(2):717–736.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9(4):383–405.

Barberis, N. and M. Huang. (2008). Stocks as Lotteries: The Implications of Probability Weighting for Security Prices. *American Economic Review*, 98(5):2066–2100.

Barberis, N., M. Huang, and T. Santos. (2001). Prospect theory and asset prices. *Quarterly Journal of Economics*, 116(1):1–54.

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

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum. (2013)a. The Nature of Risk Preferences: Evidence from Insurance Choices. *American Economic Review*, 103(6): 2499–2529.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum. (2013)b. Distinguishing Probability Weighting from Risk Misperceptions in Field Data. *American Economic Review*, 103(3):580–585.

Bateman, I., A. Munro, B. Rhodes, C. Starmer, and R. Sugden. (1997). A test of the theory of reference-dependent preferences. *Quarterly Journal of Economics*, 112(2):479–505.

Baucells, M., M. Weber, and F. Welfens. (2011). Reference-Point Formation and Updating. *Management Science*, 57(3):506–519.

Bawa, V. S. (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2(1):95–121.

Benartzi, S. and R. R. H. Thaler. (1995). Myopic loss aversion and the equity premium puzzle. *quarterly journal of Economics*, 110(1):73–92.

Camerer, C. F. and T.-H. Ho. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8(1):167–196.

Conte, A., J. D. Hey, and P. G. Moffatt. (2011). Mixture models of choice under risk. *Journal of Econometrics*, 162(1):79–88.

Coursey, D. L., J. L. Hovis, and W. D. Schulze. (1987). The disparity between willingness to accept and willingness to pay measures of value. *Quarterly Journal of Economics*, 102 (3):679–690.

Cox, J. C. and V. Sadiraj. (2006). Small- and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games and Economic Behavior*, 56(1):45–60.

Crawford, V. P. and J. Meng. (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rationalexpectations targets for hours and income. *American Economic Review*, 101(5):1912–1932.

Eisenberger, R. and M. Weber. (1995). Willingness-to-pay and willingness-to-accept for risky and ambiguous lotteries. *Journal of Risk and Uncertainty*, 10(3):223–233.

Ericson, K. M. M. and A. Fuster. (2011). Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments. *Quarterly Journal of Economics*, 126(4):1879–1907.

Fehr, E. and J.-R. Tyran. (2008). Limited rationality and strategic interaction: the impact of the strategic environment on nominal inertia. *Econometrica*, 76(2):353–394.

Fehr-Duda, H. and T. Epper. (2012). Probability and Risk: Foundations and Economic Implications of Probability-Dependent Risk Preferences. *Annual Review of Economics*, 4 (1):567–593.

Gonzalez, R. and G. Wu. (1999). On the shape of the probability weighting function. *Cognitive psychology*, 38(1):129–166.

Hadar, J. and W. R. Russell. (1969). Rules for ordering uncertain prospects. *American Economic Review*, 151(3712):867–8.

Haigh, M. S. and J. A. List. (2005). Do professional traders exhibit myopic loss aversion? An experimental analysis. *The Journal of Finance*, 60(1):523–534.

Handa, J. (1977). Risk, probabilities, and a new theory of cardinal utility. *Journal of Political Economy*, 85(1):97–122.

Harrison, G. W. (2006). Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics*, 34(1):125–162.

Harrison, G. W. and E. E. Rutström. (2009). Expected utility theory and prospect theory: one wedding and a decent funeral. *Experimental Economics*, 12(2):133–158.

Hart, O. and J. Moore. (2008). Contracts as reference points. *Quarterly Journal of Economics*, 123(1):1–48.

Heffetz, O. and J. A. List. (2014). Is the endowment effect an expectations effect? *Journal of the European Economic Association*, 12(5):1396–1422.

Holt, C. A. and S. K. Laury. (2002). Risk aversion and incentive effects. *The American Economic Review*, 92(5):1644–1655.

Horowitz, J. K. and K. E. McConnell. (2002). A Review of WTA/WTP Studies. *Journal of Environmental Economics and Management*, 44(3):426–447.

Kahneman, D. and A. Tversky. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kahneman, D., J. L. Knetsch, and R. H. Thaler. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of economic perspectives*, 5(4):193–206.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance*, 21(1):61–72.

Kilka, M. and M. Weber. (2000). What determines the shape of the probability weighting function under uncertainty? *Management science*, 47(12):1712–1726.

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *American Economic Review*, 79(5):1277–1284.

Köszegi, B. and M. Rabin. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4):1133–1165.

Köszegi, B. and M. Rabin. (2007). Reference-dependent risk attitudes. *American Economic Review*, 97(4):1047–1073.

Köszegi, B. and M. Rabin. (2009). Reference-Dependent Consumption Plans. *American Economic Review*, 99(3):909–936.

Machina, M. J. (1987). Decision-making in the presence of risk. *Science*, 236(4801):537–543.

Machina, M. J. (2009). Risk, ambiguity, and the rank-dependence axioms. *American Economic Review*, 99(1):385–392.

McClelland, G. H., W. D. Schulze, and D. L. Coursey. (1993). Insurance for low-probability hazards: A bimodal response to unlikely events. *Journal of Risk and Uncertainty*, 7(1): 95–116.

Prelec, D. (1998). The probability weighting function. *Econometrica*, 66(3):497–527.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Quiggin, J. (1987). Decision weights in anticipated utility theory: Response to Segal. *Journal of Economic Behavior & Organization*, 8(4):641–645.

Rabin, M. (2000). Risk aversion and expectedutility theory: A calibration theorem. *Econometrica*, 68(5):1281–1292.

Rabin, M. and R. Thaler. (2001). Anomalies: risk aversion. *Journal of Economic Perspectives*, 15(1):219–232.

Rabin, M. and G. Weizsäcker. (2009). Narrow Bracketing and Dominated Choices. *American Economic Review*, 99(4):1508–1543.

Rubinstein, A. (2006). Dilemmas of An Economic Theorist. *Econometrica*, 74(4):865–883.

Safra, Z. and U. Segal. (2008). Calibration Results for Non-Expected Utility Theories. *Econometrica*, 76(5):1143–1166.

Safra, Z. and U. Segal. (2009). Risk aversion in the small and in the large: Calibration results for betweenness functionals. *Journal of Risk and Uncertainty*, 38(1):27–37.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587.

Siegel, J. J. and R. H. Thaler. (1997). Anomalies: the equity premium puzzle. *Journal of Economic Perspectives*, 11(1):191–200.

Snowberg, E. and J. Wolfers. (2010). Explaining the favorite-longshot bias: Is it risk-love or misperceptions? *Journal of Political Economy*, 118(4):723–746.

Sugden, R. (2003). Reference-dependent subjective expected utility. *Journal of Economic Theory*, 111(2):172–191.

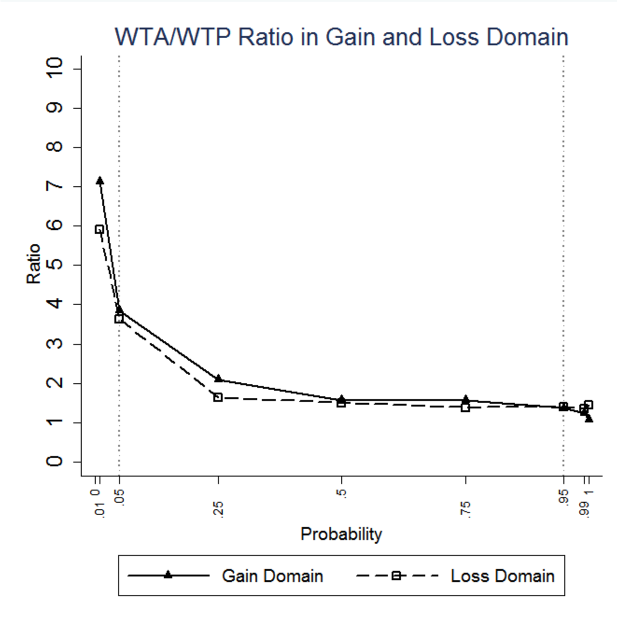
Thaler, R. H. and A. Tversky. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *The Quarterly Journal of ...*, 112(2):647–661.

Tversky, A. and D. Kahneman. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, 106(4):1039–1061.

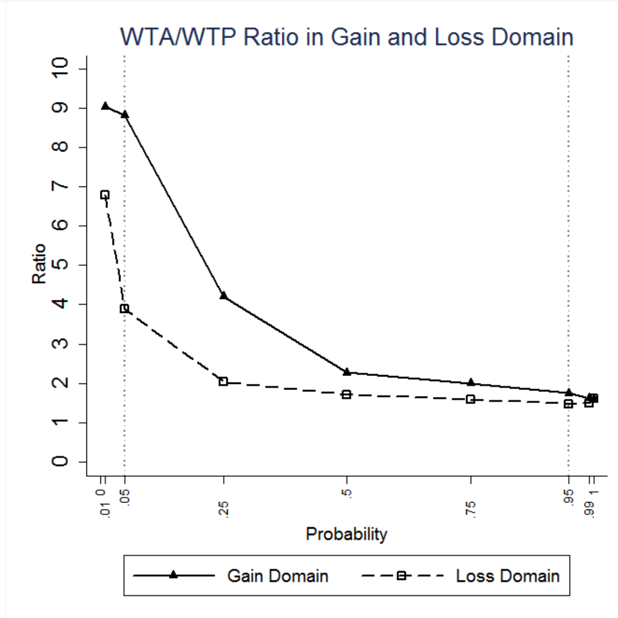
Tversky, A. and D. Kahneman. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Wu, G. and R. Gonzalez. (1996). Curvature of the probability weighting function. *Management science*, 42(12):1676–1690.

Figure 1: WTAG/WTPG Ratio (of means) under Different Probabilities

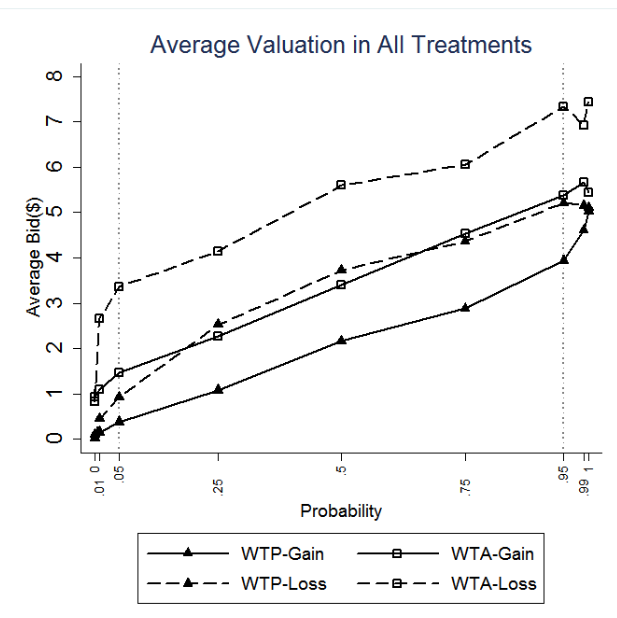


(a) Money

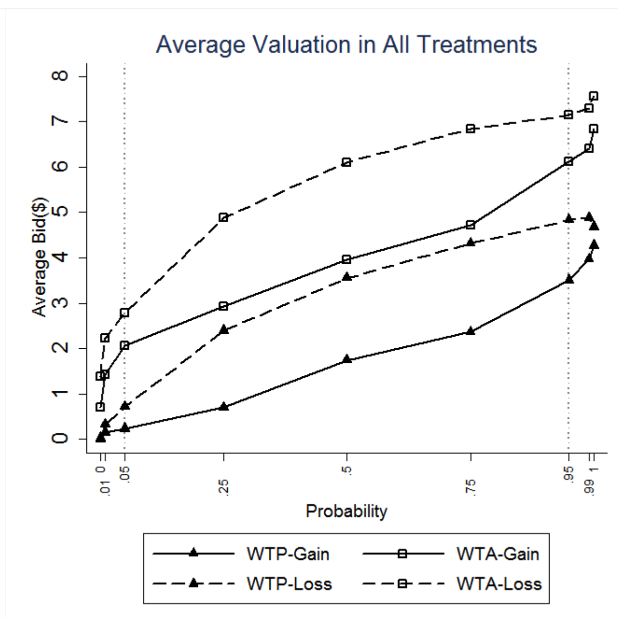


(b) Mug

Figure 2: Mean Valuations under Different Probabilities

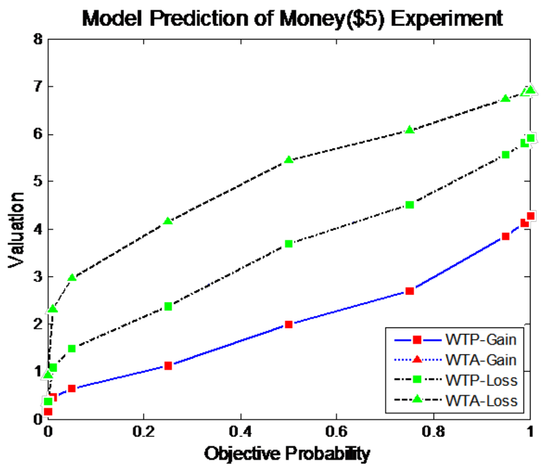


(a) Money

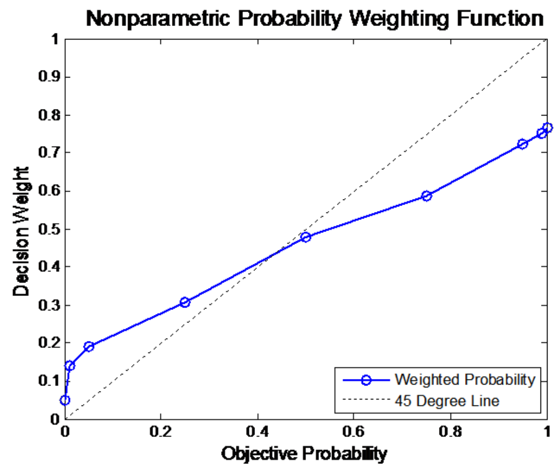


(b) Mug

Figure 3: Model Prediction and Probability Weighting Function from Estimation Using Money Data

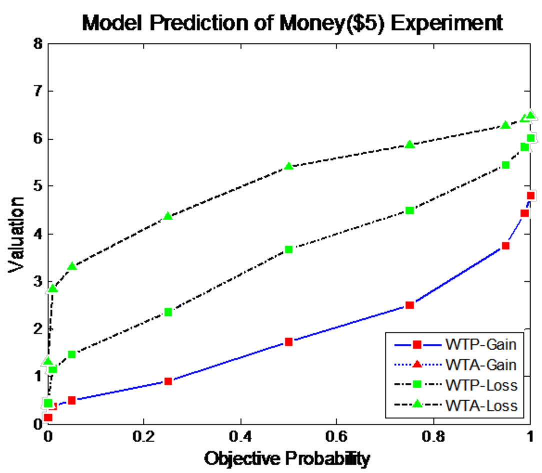


(a) Model Prediction

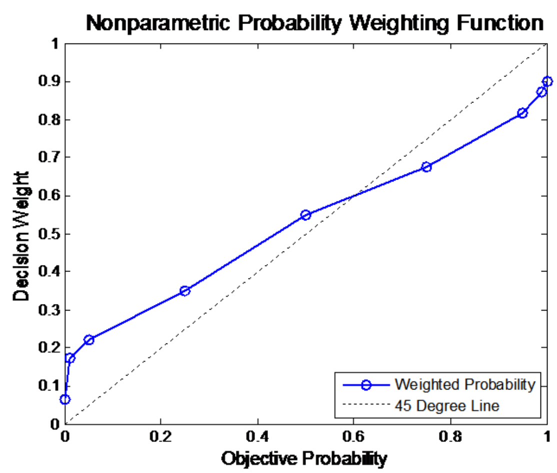


(b) Weighting Function

Figure 5: Model Prediction and Probability Weighting Function for Money Experiment with Heterogeneity in Consumption Utility of Money

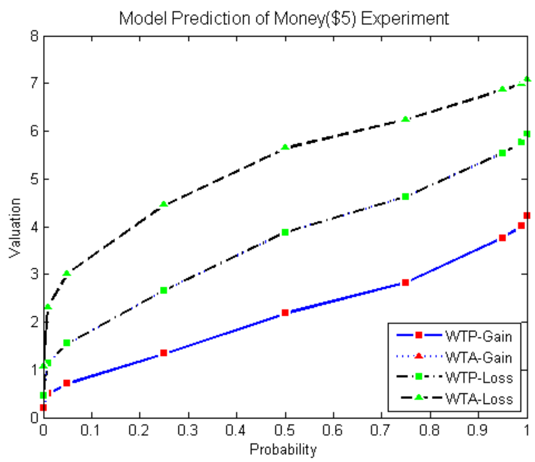


(a) Model Prediction

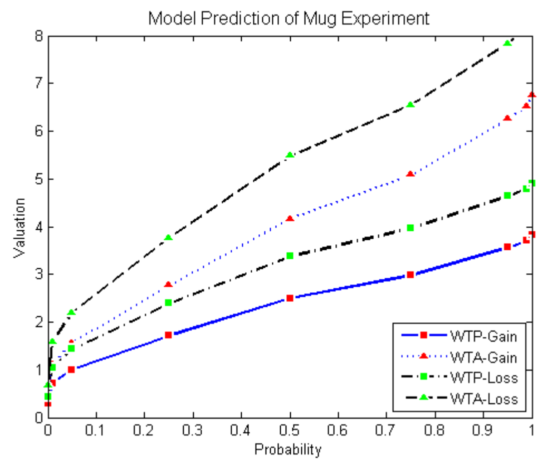


(b) Weighting Function

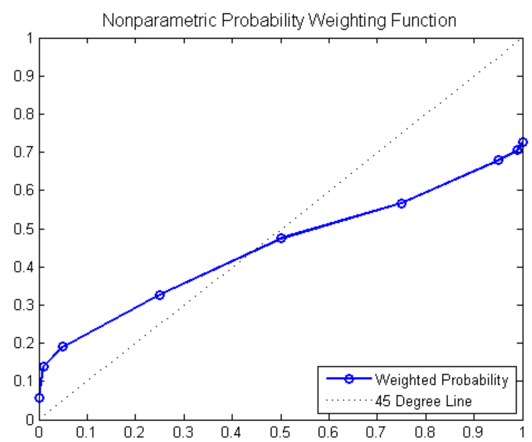
Figure 4: Model Prediction of Money Experiment (top), Model Prediction of Mug Experiment (middle) and Probability Weighting Function (bottom) from Joint Estimation



(a) Model Prediction for Money



(b) Model Prediction for Mug



(c) Weighting Function

Table 1: Summary of Treatments in Laboratory Experiment

	Treatment	Gain or Loss	Mode	# of Subjects
Money	Lottery-WTP	Gain	Buy	48
	Lottery-WTA	Gain	Sell	48
	Insurance-WTP	Loss	Buy	45
	Insurance-WTA	Loss	Sell	45
Mug	Lottery-WTP	Gain	Buy	41
	Lottery-WTA	Gain	Sell	48
	Insurance-WTP	Loss	Buy	46
	Insurance-WTA	Loss	Sell	47

Table 2: Parameters in Each Part of Lab Experiment

Part Number	Probability of Receiving/Losing $5 (Mug)	# of Chips
1	0%	100 white chips and 0 red chip
2	1%	99 white chips and 1 red chip
3	5%	95 white chips and 5 red chips
4	25%	75 white chips and 25 red chips
5	50%	50 white chips and 50 red chips
6	75%	25 white chips and 75 red chips
7	95%	5 white chips and 95 red chips
8	99%	1 white chip and 99 red chips
9	100%	0 white chip and 100 red chips

All treatments have the same parameters in each part. In gain/lottery sessions, drawing a red chip means a gain. In loss/insurance sessions, drawing a red chip means a loss.

Table 3: Linear Regression Analysis

	Money ($5)		Coffee Mug	
	Full Data (1)	Moderate Ps (2)	Full Data (3)	Moderate Ps (4)
Probability	4.43***	3.87***	3.89***	3.57***
	(0.38)	(0.34)	(0.48)	(0.48)
WTAG*Probability	0.08	0.51	1.26*	0.71
	(0.50)	(0.48)	(0.72)	(0.74)
WTPL*Probability	0.36	0.63	0.68	0.84
	(0.44)	(0.41)	(0.72)	(0.74)
WTAL*Probability	0.64	0.39	1.42*	1.05
	(0.56)	(0.54)	(0.82)	(0.79)
WTPG	0.05	0.16	(0.06)	(0.07)
	(0.09)	(0.10)	(0.06)	(0.09)
WTAG	1.10***	1.22***	1.34***	1.83***
	(0.24)	(0.29)	(0.28)	(0.34)
WTPL	0.66***	1.10***	0.57***	0.96***
	(0.10)	(0.16)	(0.12)	(0.20)
WTAL	2.40***	3.17***	2.48***	3.24***
	(0.36)	(0.38)	(0.38)	(0.39)
		Fixed Effects		
Fixed Effects	1.65	1.89	2.09	2.3
Other Factors	1.59	1.21	2.08	1.72
No. of Individuals	186	186	182	182
No. of Obs.	1674	930	1638	910

Top panel shows coefficient estimates. WTPG, WTAG, WTPL, WTAL are dummy variable for WTP-Gain, WTA-Gain, WTP-Loss and WTA-Loss treatments. Probability is treated as continuous variable. WTAG*Probability, WTPL*Probability and WTAL*Probability are interaction terms. Clustered (by subject) standard errors are reported in parentheses. *indicates significance at 10% level; **indicates significance at 5% level while ***indicates significance at 1% level.

(1) uses all data in the money experiment and (3) uses all data in the mug experiment. (2) and (4) only include observations with $0.05 \leq p \leq 0.95$ which are considered non extreme probabilities.

Bottom panel reports result from fixed effects estimation. All coefficients and significance levels are suppressed since they are the same as the ones shown in top panel except all intercepts are not identified.

Table 4: Possible Reference Points in Each Treatment

		Money – one dimension		
	Endowment	Status Quo	Rational Expectation	
			Answer “No”	Answer “Yes”
WTPG	x^0	x^0	x^0	$(x^0 + \Delta x - WTP, p; x^0 - WTP, 1-p)$
WTAG	$(x^0 + \Delta x, p; x^0, 1-p)$	x^0	$(x^0 + \Delta x, p; x^0, 1-p)$	$x^0 + WTA$
WTPL	$(x^0, p; x^0 + \Delta x, 1-p)$	$x^0 + \Delta x$	$(x^0, p; x^0 + \Delta x, 1-p)$	$x^0 + \Delta x - WTP$
WTAL	$x^0 + \Delta x$	$x^0 + \Delta x$	$x^0 + \Delta x$	$(x^0 + WTA, p; x^0 + \Delta x + WTA, 1-p)$
		Coffee Mug – two dimension		
	Endowment	Status Quo	Rational Expectation	
			Answer “No”	Answer “Yes”
WTPG	$(0, x^0)$	$(0, x^0)$	$(0, x^0)$	$((\bar{m}, x^0 - WTP), p; (0, x^0 - WTP), 1-p)$
WTAG	$((\bar{m}, x^0), p; (0, x^0), 1-p)$	$(0, x^0)$	$((\bar{m}, x^0), p; (0, x^0), 1-p)$	$(0, x^0 + WTA)$
WTPL	$((0, x^0), p; (\bar{m}, x^0), 1-p)$	$(\bar{m}, x^0)$	$((0, x^0), p; (\bar{m}, x^0), 1-p)$	$(\bar{m}, x^0 - WTP)$
WTAL	$(\bar{m}, x^0)$	$(\bar{m}, x^0)$	$(\bar{m}, x^0)$	$((0, x^0 + WTA), p; (\bar{m}, x^0 + WTA), 1-p)$

Endowment is what one has. That is, whether one has a lottery ticket/insurance policy or not. Status quo is defined empirically. That is, people do not *feel* they own the prize even they possess a lottery ticket and people do not feel they have suffered from a loss even they face risks. This reference point is in the spirit of the original prospect theory. Rational expectation assumes that decisions makers can perfectly predict what they have in the future in a rational way and use that as their reference points. This definition was used in the KR model (Köszegi and Rabin, 2006).
x^0 is current wealth; $\Delta x = \$5$ is the prize of lottery. $\bar{m}$ is intrinsic consumption value of the coffee mug. p is probability of winning the lottery or experiencing the loss. In the money experiment, subjects consider only on dimension while in the mug experiment, subjects consider two dimensions consumption of coffee mug and money.

Table 5: Maximum Likelihood (ML) and Markov Chain Monte Carlo (MCMC) Estimates of Parameters in Money Experiment

	MLE - Restricted	MLE	MCMC	MCMC W. Heterogeneity
	(1)	(2)	(3)	(4)
$\bar{u}(\mu_{\bar{u}})$	6.013	7.830	7.715	6.677
	(0.168)	(1.075)	(0.385)	(0.267)
$\sigma_{\bar{u}}$	–	–	–	1.721
				(0.100)
β	11.045	2.647	2.627	3.483
	(4.889)	(0.357)	(0.229)	(0.301)
$\pi(0)$	–	0.050	0.049	0.065
		(0.017)	(0.015)	(0.014)
$\pi(0.01)$	–	0.139	0.140	0.174
		(0.024)	(0.019)	(0.016)
$\pi(0.05)$	–	0.191	0.192	0.220
		(0.025)	(0.019)	(0.019)
$\pi(0.25)$	–	0.310	0.307	0.351
		(0.039)	(0.023)	(0.026)
$\pi(0.50)$	–	0.477	0.478	0.549
		(0.059)	(0.034)	(0.035)
$\pi(0.75)$	–	0.587	0.586	0.674
		(0.073)	(0.034)	(0.035)
$\pi(0.95)$	–	0.716	0.722	0.816
		(0.085)	(0.037)	(0.037)
$\pi(0.99)$	–	0.747	0.752	0.874
		(0.092)	(0.042)	(0.034)
$\pi(1)$	–	0.767	0.766	0.899
		(0.094)	(0.042)	(0.031)
σ	2.157	2.021	2.048	1.487
	(0.134)	(0.126)	(0.385)	(0.058)
MSE	4.794	4.159	4.155	4.197
No. of Obs.	1,674.000	1,674.000	1,674.000	1,674.000

$\bar{u}$ is marginal utility of \$5. $\beta = \frac{1+\lambda\eta}{1+\eta}$ is loss aversion parameter in the model. $\pi(\cdot)$ is non-parametric probability weighting function. σ is estimated standard deviation of additive econometric error term. Standard errors reported in parentheses. Bootstrap standard errors for ML estimates (N=100). Standard errors for MCMC estimates are calculated by burning the first 25% of the chain and then thinning the chain by every 200 iteration (Total Chain Length = 10^6).

Table 6: Maximum Likelihood (ML) and Markov Chain Monte Carlo (MCMC) Estimates of Parameters in Both Money and Mug Experiments

	MLE	MCMC
	(1)	(2)
$\bar{u}$	8.165	7.933
	(1.265)	(0.684)
$\bar{m}$	7.783	7.540
	(1.265)	(0.705)
β_{Money}	2.478	2.467
	(0.339)	(0.252)
β_{Mug}	1.483	1.492
	(0.086)	(0.039)
a	0.555	0.514
	(0.264)	(0.109)
$\pi(0)$	0.057	0.059
	(0.016)	(0.013)
$\pi(0.01)$	0.138	0.141
	(0.022)	(0.016)
$\pi(0.05)$	0.190	0.196
	(0.025)	(0.020)
$\pi(0.25)$	0.326	0.333
	(0.050)	(0.031)
$\pi(0.50)$	0.475	0.484
	(0.071)	(0.044)
$\pi(0.75)$	0.567	0.578
	(0.085)	(0.052)
$\pi(0.95)$	0.679	0.690
	(0.100)	(0.058)
$\pi(0.99)$	0.705	0.718
	(0.108)	(0.062)
$\pi(1)$	0.726	0.739
	(0.111)	(0.063)
σ_{Money}	2.011	2.047
	(0.143)	(0.036)
σ_{Mug}	2.705	2.722
	(0.110)	(0.047)

MSE	5.759	5.752
No. of Obs.	3,312	3,312

$\bar{u}$ and $\bar{m}$ are intrinsic consumption utility of the \$5 prize and coffee mug respectively. $\beta = \frac{1+\lambda\eta}{1+\eta}$ is loss aversion parameter in the model. $\pi(\cdot)$ is non-parametric probability weighting function. a is the linear multiplier in the reference point specification. σ_Money and σ_Mug are estimated standard deviations of additive econometric error terms in the money and mug experiments respectively.
Standard errors reported in parentheses. Bootstrap standard errors for ML estimates (N=100). Standard errors for MCMC estimates are calculated by burning the first 25% of the chain and then thinning the chain by every 200 iteration (Total Chain Length = 10^6).

Table 7: Model Calibration and Validation

	Model	No. of Parameters	Calibration (70% of sample)			Validation (30% of sample)		
			LL	AIC	BIC	LL	MSE	Pseudo R2
Money	EV	1	-2660.83	5323.66	5328.71	-1204.24	5.9065	0.0000
	EUT	2	-2616.02	5236.04	5246.14	-1197.00	5.7394	0.0052
	RDU	3	-2524.52	5055.04	5070.19	-1163.55	5.0404	0.0321
	RDU with PW	12	-2444.53	4913.05	4973.64	-1124.55	4.3438	0.057
Mug	EV	1	-2949.19	5990.38	5905.41	-1321.83	10.0865	0.0000
	EUT	2	-2860.67	5725.34	5735.39	-1289.4	8.9103	0.0238
	RDU	3	-2803.2	5612.4	5627.48	-1261.43	7.9861	0.0442
	RDU with PW	12	-2718.04	5460.08	5520.39	-1240.87	7.3877	0.0529

EV Expected value; EUT Expected utility theory; RDU Reference-dependent utility model; RDU with PW Reference-dependent utility model with rank-dependent probability weighting; LL Log-likelihood function value; AIC Akaike information criterion; BIC Bayesian information criterion; MSE Mean squared errors.

Table 8: Valuations in Each Money Treatment

			Probability								
Treatment	N		0.00	0.01	0.05	0.25	0.50	0.75	0.95	0.99	1.00
Expected Value			0.00	0.05	0.25	1.25	2.50	3.75	4.75	4.95	5.00
Willingness to Pay in Gain Domain (WTPG)	48	Mean	0.11	0.15	0.38	1.08	2.16	2.88	3.94	4.61	5.02
		S.E.	(0.10)	(0.06)	(0.11)	(0.10)	(0.17)	(0.25)	(0.31)	(0.41)	(0.43)
		Median	0.00	0.05	0.20	1.18	2.40	3.08	4.25	4.63	4.95
Willingness to Accept in Gain Domain (WTAG)	48	Mean	0.94	1.10	1.47	2.27	3.39	4.53	5.38	5.68	5.43
		S.E.	(0.31)	(0.28)	(0.29)	(0.25)	(0.21)	(0.19)	(0.27)	(0.26)	(0.26)
		Median	0.05	0.48	0.78	1.93	2.93	4.05	4.95	5.00	5.05
Disparity	–	Mean	0.83	0.95	1.02	1.19	1.23	1.65	1.44	1.07	0.41
		Median	0.05^{**}	0.43^{**}	0.58^{**}	0.75^{**}	0.53^{**}	0.97^{**}	0.70^{**}	0.37^{**}	0.10^{**}
Ratio	–	Mean	8.55	7.73	3.87	2.10	1.57	1.57	1.37	1.23	1.08
		Median	∞	9.60	3.90	1.64	1.22	1.31	1.16	1.08	1.02
Willingness to Pay in Loss Domain (WTPL)	45	Mean	0.02	0.45^{++}	0.93^{++}	2.53^{++}	3.73^{++}	4.36^{++}	5.21^{++}	5.16^{++}	5.11
		S.E.	(0.02)	(0.11)	(0.16)	(0.20)	(0.28)	(0.23)	(0.25)	(0.25)	(0.28)
		Median	0.00	0.25	0.60	2.25	3.25	4.10	4.95	4.95	5.00
Willingness to Accept in Loss Domain (WTAL)	45	Mean	0.83	2.67^{++}	3.37^{++}	4.15^{++}	5.60^{++}	6.05^{++}	7.34^{++}	6.93^{+}	7.44^{++}
		S.E.	(0.41)	(0.49)	(0.42)	(0.34)	(0.37)	(0.42)	(0.45)	(0.46)	(0.47)
		Median	0.05	1.30	2.75	3.75	5.00	5.00	5.05	5.05	5.50
Disparity	–	Mean	0.81	2.22	2.44	1.62	1.87	1.69	2.13	1.77	2.33
		Median	0.05^{**}	1.05^{**}	2.15^{**}	1.50^{**}	1.75^{**}	0.90^{**}	0.10^{**}	0.10^{**}	0.50^{**}
Ratio		Mean	41.5	5.93	3.62	1.64	1.50	1.39	1.41	1.34	1.46
		Median	∞	5.20	4.58	1.67	1.54	1.22	1.02	1.02	1.10

Table 9: Valuations in Each Coffee Mug Treatment

			Probability								
Treatment	N		0.00	0.01	0.05	0.25	0.50	0.75	0.95	0.99	1.00
Willingness to Pay in Gain Domain (WTPG)	41	Mean	0.04	0.16	0.24	0.70	1.75	2.37	3.51	3.98	4.27
		S.E.	(0.03)	(0.06)	(0.08)	(0.12)	(0.23)	(0.31)	(0.47)	(0.50)	(0.50)
		Median	0.00	0.00	0.00	0.50	1.25	1.65	2.50	3.25	3.75
Willingness to Accept in Gain Domain (WTAG)	48	Mean	0.71	1.42	2.07	2.94	3.96	4.72	6.12	6.42	6.84
		S.E.	(0.30)	(0.34)	(0.38)	(0.27)	(0.28)	(0.38)	(0.48)	(0.52)	(0.53)
		Median	0.05	0.53	1.23	2.95	4.45	4.65	5.75	5.95	6.68
Disparity	–	Mean	0.67	1.26	1.83	2.24	2.21	2.35	2.61	2.44	2.57
		Median	0.05^{**}	0.53^{**}	1.23^{**}	2.45^{**}	3.20^{**}	3.00^{**}	2.25^{**}	2.70	2.93
Ratio	–	Mean	17.75	8.88	8.63	4.20	2.26	1.99	1.74	1.61	1.60
		Median	∞	∞	∞	5.90	3.56	2.82	2.30	1.83	1.78
Willingness to Pay in Loss Domain (WTPL)	46	Mean	0.01	0.33	0.72^{++}	2.40^{++}	3.55^{++}	4.32^{++}	4.84	4.89	4.68
		S.E.	(0.01)	(0.14)	(0.16)	(0.28)	(0.34)	(0.41)	(0.53)	(0.55)	(0.57)
		Median	0.00	0.00	0.50	2.20	3.85	4.38	4.88	5.35	4.98
Willingness to Accept in Loss Domain (WTAL)	47	Mean	1.38	2.23	2.79	4.88^{++}	6.11^{++}	6.84^{++}	7.14	7.30	7.55
		S.E.	(0.44)	(0.46)	(0.44)	(0.37)	(0.40)	(0.44)	(0.48)	(0.50)	(0.61)
		Median	0.05	0.95	1.90	4.95	5.75	6.90	7.15	7.50	8.00
Disparity	–	Mean	1.37	1.90	2.07	2.48	2.56	2.52	2.30	2.41	2.87
		Median	0.05^{**}	0.95^{**}	1.40^{**}	2.75^{**}	1.90^{**}	2.52^{**}	2.27^{**}	2.15^{*}	3.02^{**}
Ratio		Mean	138	6.76	3.88	2.03	1.72	1.58	1.48	1.49	1.61
		Median	∞	∞	3.80	2.25	1.49	1.58	1.47	1.40	1.61

Ratios are calculated as $\frac{Mean(WTA)}{Mean(WTP)}$ and $\frac{Median(WTA)}{Median(WTP)}$.

Wilcoxon-Mann-Whitney test conducted to test hypothesis that WTA=WTP in the same domain (gain or loss). Null hypothesis is rejected at 1% level for all probabilities.

Pearson χ^2 test (using medians) conducted to test hypothesis that WTP=WTA in the same domain (gain or loss). *indicates null hypothesis being rejected at 5% level. **indicates hypothesis being rejected at 1% level.

Wilcoxon-Mann-Whitney test conducted to test hypothesis that WTP-Gain =WTP -Loss and WTA -Gain =WTA -Loss separately. $^+$ indicates hypothesis being rejected at 5% level. $^{++}$ indicates hypothesis being rejected at 1