Expectations as Endowments:

Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments*

Keith M. Marzilli Ericson
and
Andreas Fuster **(Job Market Paper)**†

November 9, 2010

Abstract

Substantial evidence suggests that people are loss averse: they weigh losses more heavily than equivalent gains. Yet little is known about the reference point relative to which gains or losses are defined. The most common assumption is that the reference point is given by a person's current endowment (the status quo). We conduct two experiments that instead provide evidence that a person's expectations about outcomes determine her reference points. In the first experiment, we endow subjects with an item and randomize the probability they will be allowed to trade it for an alternative. Subjects that have a lower exogenous probability of being able to trade their item (and who therefore expect to keep it) are less likely to choose to trade when such an opportunity arises. This is predicted when reference points are expectation-based, but not when they are determined by the status quo. Our second experiment provides a quantitative measure of the effect of expectations on subjects' monetary valuation of an item. We randomly assign subjects a high (80%) or low (10%) probability of obtaining an item for free, and then elicit their willingness-to-accept for the item conditional on receiving it. Being in the high probability treatment increases valuation of the item by 20-30%. These results shed light on the circumstances under which loss aversion will affect individual behavior and market outcomes, and reconcile conflicting findings regarding the existence of endowment effects in different settings.

1 Introduction

Evidence from a variety of settings indicates that people are "loss averse": they dislike losses much more than they like equal-sized gains. However, losses and gains are vague constructs, since they are only defined relative to some reference point. It is often unclear what will be considered a loss, and thus how loss aversion will affect behavior. Specifically, the predicted effects of loss aversion may depend on whether people compare outcomes a) to their current endowment, or b) to what they expect to have. Consider a homeowner who receives an unsolicited $1M bid for his house. If he did not own a house, he would slightly prefer to receive $1M instead of receiving this house. However, if he is loss averse relative to his current endowment (he owns the house), he will reject the $1M bid for his house. The loss of the house looms much larger than the gain of money, so the owner's loss aversion blocks the transaction. On the other hand, suppose the homeowner is a retiree who has long planned to sell the house and move to Florida. In that case, his expectation—to be living in Florida—may be his reference point, and staying in his current position may feel like a loss. As a consequence, the homeowner perceives *not* selling to be a loss, and may accept a lower offer than if he were not loss averse. With an expected Florida relocation as the reference point, the owner's loss aversion effectively facilitates the sale of his old house.

In this paper, we report two experiments in which we manipulate expectations separately from current endowments (the status quo) and test whether expectations affect subsequent behavior. Our findings suggest that people's reference points are not necessarily fixed by the status quo—rather, people appear loss averse around their expected outcomes. As a consequence, loss aversion need not impede transactions that participants expect to occur; for instance, when they enter a market planning to trade. However, in other situations, loss aversion may have powerful effects on people's behavior, even if they are not formally losing anything compared to their status quo.

Although reference-dependent preferences and loss aversion have been invoked to explain various economic behaviors[1], there is no general agreement on how reference points are determined. Kahneman and Tversky's (1979, 1991) influential prospect theory, where loss aversion was first introduced, left the reference point imprecise. Subsequently, it has often been assumed that ei-

---

[1]Well-known field applications include individual investment and trading behavior (Benartzi and Thaler, 1995; Odean, 1998; Genesove and Mayer, 2001), labor supply (Camerer et al., 1997; Fehr and Götte, 2007), and consumer demand (Hardie, Johnson, and Fader, 1993).

ther the status quo (Samuelson and Zeckhauser, 1988) or some previous outcome determines the reference point. The alternative view that reference points are determined by recent expectations about outcomes, which need not correspond to the status quo, has been advocated by Kőszegi and Rabin (2006, 2007, 2009) (henceforth KR).[2] Since the status quo often determines expectations, the two alternatives are sometimes indistinguishable; yet as illustrated above, the implications of loss aversion are often very different depending on the assumed reference points. Evidence from controlled experimental environments can help distinguish between the different specifications.

Our first experiment tests whether expectations affect exchange behavior. The key feature of our design is that it induces expectations that differ across subjects while keeping current ownership status fixed. We do this by exogenously manipulating the probability that subjects, who are all endowed with the same item (a mug), will have the possibility to exchange that item for an alternative (a pen). While some subjects have a 90% probability of having the option to exchange if they so desire, others only have a 10% probability. A few minutes after endowing the subjects with their mug and informing them of the potential exchange option, we ask them whether they would like to exchange in case the option materializes. If preferences are reference-independent, or if current endowments (which are the same for all subjects) determine the reference point, there should be no difference in exchange propensities across the two treatments. However, we find that subjects that are more likely to *expect* to keep their endowed item (because they have a low probability of being allowed to exchange) are more likely to *choose* to keep their item if given the opportunity to exchange: only 23% of subjects in the low-probability treatment would like to exchange, compared to 56% of subjects in the high-probability treatment. This finding is predicted if individuals have expectation-based reference points and suggests that the expectation of continued ownership, rather than current formal ownership per se, induces a reluctance to part with possessions (a phenomenon Thaler, 1980 called the "endowment effect").

The large impact of expectations on exchange propensities in the first experiment could be driven by a small effect of expectations on valuation, if subjects were otherwise near-indifferent between the two objects. We thus design a second experiment with the goal of measuring the quantitative magnitude of the effect of expectations on subjects' monetary valuation of an item.

---

[2]A related assumption made in earlier papers such as Bell (1985), Loomes and Sugden (1986), or Gul (1991) is to take the expected value or the certainty equivalent of a gamble as the reference point to which outcomes are compared.

Subjects are not initially endowed with any item, but learn that they will receive a mug for free with either high (80%) or low (10%) probability. All subjects also know that with probability 10%, they instead have a choice between the mug and money. We elicit their mug/money choices for varying amounts of money, which provides us with a measure of their willingness-to-accept (WTA) for the mug. Our results show that subjects who were randomly assigned to have a high expectation of leaving with the mug value the mug about 20-30% higher than subjects who were less likely to leave with the mug, an effect that is both statistically and economically significant.

We additionally conduct a variation of our second experiment to untangle whether expectations matter because people experience gains and losses with respect to a reference point, or because likely ownership of an item increases individuals' estimates of its consumption utility. According to the former theory, expectations should only affect the valuation of the mug currently in a person's reference point. The latter theory, which we refer to as "motivated taste change," instead predicts that expecting to own one mug should also increase the valuation of a second, identical mug. We do not find any evidence supporting the motivated taste change theory.

Expectation-based loss aversion has various implications for consumer behavior and for policy. As an example, imagine a potential car buyer who forms an expectation that she will buy a particular model. Loss aversion may lead the buyer not to cancel her purchase when she learns at the car dealership that the car gets less than 20 miles per gallon or that the financing options are less advantageous than she had thought, even though if she had known this information from the beginning she would not have planned on buying that car. Thus, the timing of such disclosures can be very important for how they affect demand when expectations determine reference points, unlike when people do not display loss aversion or are loss averse with respect to the status quo. More generally, firms' pricing and marketing strategies will attempt to instill in consumers an expectation of purchase, which may increase their willingness to pay.[3] Policymakers also benefit from "expectations management," as they may face less resistance against a policy change if it was expected than if it comes as a surprise.

Apart from distinguishing between different theories of reference point formation, our experi-

---

[3]The theoretical implications of expectation-based reference points for optimal pricing strategies are studied by Heidhues and Kőszegi (2008, 2010) and Herweg (2010). Other recent theoretical work applies the KR model to study optimal incentive contracts (Herweg, Müller, and Weinschenk, 2009; Macera, 2010) or bidding behavior in auctions (Lange and Ratan, 2010).

ments contribute to a recent debate regarding the existence and interpretation of endowment effects in exchange experiments (Knetsch, 1989; Plott and Zeiler, 2007; Knetsch and Wong, 2009) and experiments comparing WTA of current owners to willingness-to-pay (WTP) of current non-owners (Kahneman, Knetsch, and Thaler, 1990; Plott and Zeiler, 2005). In these experiments, procedural details may influence subjects' perceptions of the likelihood that they will be able to trade their endowed item for an alternative item or money, or that they might also be given the alternative item. Most discussions of such experiments implicitly assume that reference points are given by current endowments, in which case these perceptions should not matter for behavior. As a consequence, the results by Plott and Zeiler (2005, 2007), who find no endowment effect under their preferred experimental procedures, can be interpreted as evidence against the view that utility is kinked at the current endowment. However, they should not be interpreted as evidence against reference dependence and loss aversion more generally: if Plott and Zeiler's procedures led subjects to expect the trade possibility, the absence of an endowment effect is still compatible with reference-dependent preferences with expectation-based reference points.[4] Indeed, in our experiments such expectations matter a lot for behavior, in ways consistent with a kink in the utility function around *expected* outcomes.

There is little other experimental research that directly examines the effect of expectations on reference points.[5] Knetsch and Wong (2009) conduct exchange experiments using a variety of procedures, and discuss their results in terms of the KR theory. They do not manipulate expectations explicitly as we do in our first experiment, and do not directly test the theory of expectation-based reference points. The study by Smith (2008) is close to our second experiment, in that he tests whether a higher (lagged) probability of receiving an item increases subjects' valuation of that item (as measured by WTA of subjects that end up receiving the item and WTP of subjects who do not). The data cannot reject the null hypothesis that lagged expectations do not matter for valuation, but standard errors are large enough that there would not have been

---

[4]As discussed by DellaVigna (2009), a similar argument could explain List's (2003) finding that experienced sports card traders, who may expect to trade the item they get endowed with, do not seem to display the endowment effect. Engelmann and Hollard (2010) instead argue that trading experience reduces "trade uncertainty" (the perception of how risky the act of trading is), and that trade uncertainty is a source of observed endowment effects.

[5]Some recent field evidence on cab drivers' labor supply (Crawford and Meng, 2010), effort choice of professional golf players (Pope and Schweitzer, 2009), risky choices in "Deal or No Deal" (Post et al., 2008), domestic violence (Card and Dahl, 2010), and police performance after final offer arbitration (Mas, 2006) is also consistent with reference points being determined by expectations.

enough power to detect reasonably sized effects (e.g. a \$1, or 15-25%, change in valuation).[6] In our second experiment, we attempt to proxy for idiosyncratic factors that may affect valuation, which increases our statistical power to detect a treatment effect.

Abeler, Falk, Götte, and Huffman (2009) conduct a real-effort experiment in which subjects are equally likely to receive a fixed payment or to be paid according to their effort, and only learn which case applies after they stop working. When the amount of the fixed payment is higher, subjects work more. This is predicted when expectations determine reference points, as subjects want to avoid being disappointed by potentially earning less than their fixed payment.[7] In this study, subjects are not informed that the level of the fixed payment (high/low) is randomly assigned. As a consequence, subjects may make (potentially mistaken) inferences from the amount of the fixed payment about the "appropriate" amount of effort to provide, and this could contribute to the observed treatment effect (e.g., if the fixed payment is perceived as an indication of how much the experimenter expects participants to work).[8] We provide evidence that transparent randomization can be important: in pilot sessions of our first experiment, subjects were not aware that their probability of being allowed to exchange was randomly determined, and they made inferences about the relative value of the items which then influenced their behavior. Our two experiments make randomization transparent, and these inferences disappear. We thus confirm that expectations affect reference points, and furthermore provide new evidence that expectations matter not only for effort provision, but also for exchange behavior and the valuation of goods.

2   Experiment 1: Expectations and Exchange Behavior

In our first experiment, we endow subjects with an item (a university travel mug) and randomly and transparently manipulate the probability with which a subject will have the opportunity to exchange her item for an alternative (a silver metal university pen).[9] The experiment is a variation

---

[6]Another potential reason why no effect is found may be that valuations are elicited after subjects learn whether or not they receive the item for free, so that the reference point may already have adjusted to having the item (or not having it) for sure.

[7]See also a related study by Gill and Prowse (2010) who find evidence for disappointment aversion in a two-person sequential-move tournament.

[8]The concern that subjects might conform to their interpretation of appropriate effort is distinct from anchoring or gift-exchange motives, which Abeler et al. control for and find no evidence of in additional treatments. See Section 4.2 for more detail.

[9]Both items have a retail value of around \$8. In a typical endowment effect experiment, half the subjects would be endowed with a mug and the other half with a pen. We chose to endow all our subjects with the same item (which

of a classic experiment by Knetsch (1989), with many procedural details following Plott and Zeiler's (2007) "loss emphasis treatment." We first derive the theoretical predictions, then describe our experimental procedures, and present the results.

2.1 Theoretical Predictions for Behavior

In this section, we compare how expectations affect the exchange behavior of three types of individuals: an individual with classical preferences (no loss aversion), one with loss aversion around a reference point given by the status quo, and one with loss aversion around a reference point determined by expectations (as in KR 2006).[10] Consider an individual who is endowed with a mug and expects to have the option to exchange this mug for a pen with probability p. After being given some time to think about the situation, she is then asked to register a decision: if the option to exchange materializes, would she like to exchange?

Let individuals have utility functions $u(c|r)$ that can depend both on consumption c and reference levels r. Consumption and reference levels have multiple dimensions; for this setting, only the "mug" dimension and the "pen" dimension are relevant. Utility on the kth dimension is composed of direct consumption utility u_k and gain-loss utility with respect to that dimension's reference level of utility u_k^r. Total utility from a consumption outcome c is then given by:

$$u(c|r) = \sum_{k \in \{mug, pen\}} \left\{ \underbrace{u_k}_{\text{Consumption utility}} + \underbrace{\mu(u_k - u_k^r)}_{\text{Gain-loss utility}} \right\},$$

where μ has the properties of the Kahneman-Tversky value function. For simplicity, we follow Section IV of KR (2006) and assume μ to be a piecewise linear function with a kink at zero that captures loss aversion: $\mu(x) = \eta x$ for $x > 0$ and $\mu(x) = \eta \lambda x$ for $x \leq 0$, where $\eta \geq 0$ is the weight on gain-loss utility and $\lambda > 1$ is the individual's loss aversion coefficient. This specification nests classical preferences that do not feature gain-loss utility ($\eta = 0$). Throughout, when referring to individuals with reference-dependent preferences, we assume $\eta > 0$.

Now we consider two different specifications of the reference point. If the reference point is

---

we randomly picked by flipping a coin before our very first session) in order to get sufficient power for testing the hypothesis that expectations affect exchange behavior.

[10] A more detailed exposition is provided in the Theory Appendix.

given by the status quo, which is that the individual owns the mug but not the pen, then we simply have $u_{mug}^r = u_{mug}$ and $u_{pen}^r = 0$.

In contrast, in KR's specification, the reference point is determined by the individual's probabilistic beliefs about outcomes, or in other words, her expectations. KR assume that individuals calculate an outcome's gain-loss utility as follows: they compare each dimension's consumption utility from the actual outcome to that of each possible outcome, producing a series of gain-loss $\mu(\cdot)$ terms. Each possible outcome's $\mu(\cdot)$ is then weighted by the probability with which the individual expected that outcome to occur.[11] Furthermore, an individual's reference point is *endogenous* to her plans, as the probabilities of different states of the world occurring may depend at least in part on her decisions.

In our example, if the KR individual plans not to exchange, then her reference point is the same as in the status quo case—she keeps the mug for sure. If instead she plans to exchange conditional on having the possibility to do so, the two states of the world she considers for her gain-loss utility are $r_{end\ w/pen} = \{1\text{ pen},\ 0\text{ mugs}\}$ and $r_{end\ w/mug} = \{0\text{ pens},\ 1\text{ mug}\}$, which occur with probability p and $1-p$, respectively. Her utility of consumption bundle c given her expectations is then $pu\left(c|r_{end\ w/pen}\right) + (1-p)\,u\left(c|r_{end\ w/mug}\right)$. If the exchange occurs, then she gains or loses nothing with respect to $r_{end\ w/pen}$, but with respect to $r_{end\ w/mug}$, she gains u_{pen} and loses u_{mug}. Thus, her total utility of this outcome would be her consumption utility u_{pen} plus her gain-loss utility $(1-p)\,\eta\,(u_{pen} - \lambda u_{mug})$. Similarly, if the exchange does not occur, her total utility would be $u_{mug} + p\eta\,(u_{mug} - \lambda u_{pen})$. Following KR, we assume that the individual acts so as to maximize her expected utility given her reference point (which itself is determined by her plan), and chooses the plan that leads to the highest expected utility, given her rationally-forecasted future actions.[12]

Now, let individuals vary in their value for the mug and for the pen; assume that (u_{mug}, u_{pen}) are distributed according to some population distribution function. The following proposition shows that increasing the probability p with which an individual is allowed to exchange increases the expected proportion of individuals who choose to exchange when reference points are determined

---

[11]In this application, no predictions would change if reference points were instead simply given by the expected consumption utility in each dimension, in the spirit of Bell (1985).

[12]In the language of KR, we assume that the individual plays the "preferred personal equilibrium." However, the theoretical prediction on the effect of p is the same when using the less restrictive "personal equilibrium" concept, or Kőszegi and Rabin's (2007) "choice-acclimating personal equilibrium," which assumes that reference points are determined by the individual's decision, not her plan. See the Theory Appendix for more details.

by expectations, but not when reference points are given by the status quo or when individuals do not have gain-loss utility.

Proposition 1 *If reference points are given by the status quo or otherwise unaffected by expectations, or if there is no gain-loss utility ($\eta = 0$), the probability p of having the option to exchange does not affect the decision to exchange. If reference points are determined by expectations, then increasing p increases the expected proportion of individuals who choose to exchange.*

Proof. When reference points are given by the status quo, the expected utility from choosing to exchange is greater than the utility of keeping the mug if and only if $p\left[u_{pen} + \eta\left(u_{pen} - \lambda u_{mug}\right)\right] + (1-p)u_{mug} > u_{mug}$, which holds if and only if $u_{pen} > \frac{1+\eta\lambda}{1+\eta}u_{mug}$ and thus does not depend on p. A similar condition, not dependent on p, holds for any fixed reference point. When $\eta = 0$, the individual strictly desires to exchange if and only if $u_{pen} > u_{mug}$.

However, when reference points are determined by expectations, an individual desires to exchange if and only if the expected utility of planning to exchange (and following through on the plan) is greater than the expected utility of planning to keep the mug (and keeping it). As the Theory Appendix shows, this is true when $(u_{pen} - u_{mug}) + \eta\left(1 - p\right)\left(1 - \lambda\right)\left(u_{pen} + u_{mug}\right) > 0$. Note that $u_{pen} > u_{mug}$ is a necessary condition for the individual to desire to exchange, as for $p < 1$ the second term is negative due to loss aversion. The second term gets less negative as p increases. Therefore, an individual who chooses to keep the mug under one value of p will always do so for lower values of p, but may instead choose to exchange for higher values of p. ■

Intuitively, for an individual who has classical preferences, the decision of whether to exchange is influenced only by the consumption utilities she derives from the two items. On the other hand, an individual who is loss averse around the status quo may not be willing to give up the mug he owns for the pen, even if the pen has higher consumption utility, due to the loss he feels from moving away from the status quo. However, this loss, and therefore the decision, is independent of whether the chance of the exchange occurring is high or low.

Meanwhile, for an individual with KR preferences, the cost imposed by the gain-loss utility in case she decides to exchange must be outweighed by the gain in consumption utility from getting the pen instead of the mug, and this is less likely to happen the lower is p. On the other hand, as $p \to 1$, gain-loss utility becomes relatively less important. In the extreme, if such an individual

9

were certain to be able to exchange, she would prefer to do so if and only if the consumption utility from the pen exceeds the one from the mug, exactly like a classical individual.

2.2 Procedures[13]

Upon arrival at the lab, each subject is seated at a carrel with a mug and a pen on it. We then flip a coin in front of each subject individually. The coin's sides are labeled "1" and "9", and we give each subject an index card with their resulting number on it. Subjects then start reading instructions on their computer screen. Subjects are told that they own the mug in front of them, and that each participant will leave the experiment with either a mug or a pen. They are then informed that at the end of the session, they may have the option to exchange their mug for the pen, if they so desire, and that this option will occur if and only if a ten-sided die that we roll individually for each subject at the end of the study comes up lower than or equal to the number on their index card. Thus, subjects whose coin came up "1" have a 10% chance of having the option to exchange (we will refer to them as being in *Treatment T^L, for low probability of being able to trade) while subjects who got a "9" have a 90% chance of being able to exchange their mug for the pen if they would like (*Treatment T^H*).

After this first round of instructions, and after an experimenter reads the most important parts out loud, subjects answer some demographic questions and then fill out the first part of a 44-item "Big Five" personality questionnaire (John, Donahue, and Kentle, 1991). The purpose of this questionnaire is to distract subjects from the main decision we are interested in, and also to provide them with some time to plan their decision as to whether or not to exchange the mug. After they finish answering the first 22 questions, we remind them of the instructions and procedures for the (possible) mug-pen exchange, in order to make sure they understand, and also to make them think about their choice. Then, after they answer the second 22 questions of the questionnaire, subjects are asked to make a choice *conditional* on the die coming up lower than or equal to the number on their index card.[14] This allows us to observe a decision for each subject, not just the ones for which the decision actually turns out to apply, which is similar to the strategy method often used

---

[13]More details on both experiments are provided in the Methods Appendix, available from the authors.

[14]They are asked to check one of the following: If the the die comes up [1 in Treatment T^L; 1 to 9 in Treatment T^H]: □ I want to keep the mug; or □ I want to trade the mug for the pen.

in experimental games.[15] Before rolling the die, we ask some additional questions, as described in Section 4.1.2.

The experiments were run at the Harvard Decision Science Laboratory. A total of 45 subjects (23 females; mean age 21), all of them undergraduate students at the university, participated. We conducted 10 sessions with between 3 and 7 subjects per session. Half the sessions were run on one day in late October 2009, the other half over two days in early November. Subjects received a show-up fee of $10, and the experiment took about 20 minutes.[16]

2.3 Results

RESULT 1: *Subjects that have a 10% chance of being able to exchange are significantly less likely to be willing to exchange than subjects that have a 90% chance. The proportion of subjects who choose to exchange the mug for the pen is 22.7% in Treatment T^L and 56.5% in Treatment T^H (p=0.033, two-sided Fisher's exact test).*

Thus, we confirm the prediction of KR (2006) that reference points matter for choice, and that they are determined by expectations. The effect is large: subjects in Treatment T^H are 34 percentage points, or one and a half times, more likely to choose to exchange than subjects in Treatment T^L. As a test of the robustness of the result, Table 1 reports the estimated marginal effects from probit regressions that predict the probability a subject chooses to exchange from a treatment indicator and other covariates that may be related to their choice. Among the covariates we consider, gender significantly predicts the desire to exchange: column (2) shows that females seem to like the pen relatively more than males. However, the gender effect does not drive our result, as our treatments were perfectly gender-balanced. Column (3) shows that when both treatment and gender indicators are included as regressors, the indicator for Treatment T^H is still significant

---

[15]As is further discussed in the Theory Appendix, eliciting a subject's decision conditional on a choice set being reached is theoretically equivalent to asking the subject to choose once the choice set is reached, if by then the subject has made her plan. This is because KR (2006) posit that reference points are determined by lagged expectations. An interesting avenue for future research would be to test this prediction and investigate how quickly reference points adapt. However, if in our experiment we were to only elicit decisions after the uncertainty was resolved, we would need a prohibitively large number of subjects to examine decisions in the 10% possibility-of-trade condition.

[16]In the November sessions, the subjects afterwards participated in a second, completely unrelated experiment. The fact that these experiments were unrelated was made very clear to participants, and they were told that the first experiment involved the mugs and pens while the second involved choosing between various payment amounts. The results below demonstrate that behavior was similar in the October and November sessions.

at $p < 0.02$. Furthermore, column (4) shows that adding subject age and an indicator for the day on which the session took place does not change the result either. If anything, the estimated treatment effect becomes even larger.[17]

3 Experiment 2: Expectations and Valuation

3.1 Theoretical Predictions for Behavior

3.1.1 Effect of Expectations on WTA

To provide a quantitative measure of the effect of expectation-based reference points, we now consider how WTA for an item is affected by expectations of getting that item. As in the case of exchange behavior, this section shows that expectations affect WTA of individuals with loss aversion around a reference point determined by expectations, but not that of individuals with reference-independent preferences or with a reference point given by the status quo.

Consider an individual who is told the following: she will receive a mug for free with probability p, which varies between individuals and is either high ($p_H = 0.8$) or low ($p_L = 0.1$). With probability q (=0.1) the individual will receive the opportunity to choose between receiving the mug or various amounts of money. With probability $1 - p - q$, the individual expects to get nothing. Individuals' WTA in case they have the opportunity to trade the mug for money is elicited using an incentive-compatible modified Becker-DeGroot-Marschak (1964) (BDM) mechanism. How do expectations affect WTA under different theories of the reference point?

As before, let consumption utility have multiple dimensions, here a "mug" dimension and a "money" dimension. We parameterize the consumption utility of money as being linear, but this is not crucial. Proposition 2 shows that the probability p of getting the mug for free alters WTA for the mug if reference points are determined by expectations, but has no effect under other theories of the reference point.

Proposition 2 *The probability p of getting the mug for free does not affect WTA for the mug if reference points are given by the status quo or otherwise unaffected by expectations, or if there is*

---

[17]The robustness of the effect across subsets of sessions is strong: the proportions of subjects in Treatments T^L and T^H who wanted to exchange were 23.1% and 53.9% in our October sessions (13 subjects per treatment) and 22.2% and 60.0% in our November sessions (9 subjects in Treatment T^L, 10 subjects in Treatment T^H).

no gain-loss utility ($\eta = 0$). If reference points are determined by expectations, then increasing the probability p affects WTA for the mug. For any values of the parameters $\eta > 0$ and $\lambda > 1$, there is a q small enough such that increasing p unambiguously increases WTA for the mug.

Proof. If the status quo (no mug, no money) is the reference point, or if $\eta = 0$, then WTA for the mug simply equals the consumption utility of the mug. For any fixed reference point, WTA may depend on λ and η, but will not depend on p. The Theory Appendix shows that with expectation-based reference points and $q = 0$, WTA for the mug is given by $\left(1 + \frac{\eta(\lambda-1)}{1+\eta}p\right) u_{mug}$, which is increasing in p.[18] The Appendix also considers the case when $q > 0$, and shows that a sufficient condition for WTA to be increasing in p is $q < \frac{1+\eta}{2\eta(\lambda-1)}$. ∎

Denote by WTA_{mug} an individual's WTA for the mug, which is affected by p under KR preferences. The intuition underlying Proposition 2 is that a higher probability of getting the mug for free increases the weight on the mug in the individual's reference point. However, WTA_{mug} is also affected by q, the probability that the results of the BDM mechanism are implemented, because the endogenous probability of receiving money or a mug in this case affects the reference point. The Appendix shows that WTA_{mug} is unambiguously increasing in p for q small enough; for $q = 0.1$, as in our experiment, it is sufficient that $\lambda \leq 5$. This is a mild restriction, as it is unlikely that losses loom more than five times larger than gains.

Examining the limit as q goes to zero ($q = 0.1$ is small in our experiment) clarifies the intuition: when $p = 0$, WTA_{mug} simply gives the consumption utility of the mug, u_{mug}. When $p = 1$, getting the mug is the full reference point, and $WTA_{mug} = \frac{1+\eta\lambda}{1+\eta}u_{mug}$. For intermediate values of p, we have $WTA_{mug} = \left(1 + \frac{\eta(\lambda-1)}{1+\eta}p\right) u_{mug}$. Thus, the effect of changing expectations is proportional to u_{mug}, suggesting the use of the logarithmic functional form to identify the effect of p as a percentage change in WTA.

3.1.2 Improving Statistical Power

Previous studies such as Smith (2008) have found large variance in subjects' WTA or WTP for items similar to university mugs, which might make it difficult to detect even a reasonably sized effect of expectations on WTA. To improve statistical power, we control for individual-specific

---

[18]This is the case considered by Smith (2008).

valuation of university merchandise by eliciting WTA for an unrelated item, a university pen. Holding expectations constant, WTA for both the mug and pen will be affected by similar factors, such as individual wealth and attachment to the university. Hence, by controlling for WTA for the pen, the unexplained variance in WTA for the mug should be reduced and a more precise estimate of the effect of expectations may be obtained.

Consider the following situation. After the individual provides her WTA for the mug, but before she learns whether her choice applies, she is given a surprise decision. She is shown a pen and her WTA for the pen, WTA_{pen}, is elicited using the same modified BDM mechanism as for the mug. The pen-related BDM mechanism will be implemented with probability q, in addition to the probability q that the mug-related BDM mechanism will be implemented.[19] Proposition 3 verifies that WTA_{pen} is unaffected by p for theories of the reference point that do not depend on expectations, and that if a higher p increases WTA_{mug}, then WTA_{pen} will not decrease in p. Thus, controlling for individual WTA_{pen} will not lead to a predicted effect of p on WTA_{mug} when there is no such effect.

Proposition 3 *The probability p of getting the mug for free has no effect on WTA for the pen if reference points are given by the status quo or otherwise unaffected by expectations, or if there is no gain-loss utility. If reference points are determined by expectations and WTA for the mug is increasing in p, then WTA for the pen will weakly increase in p.*

Proof. As in Proposition 2, if the reference point is the status quo (no mug, no money, no pen), then WTA for the pen simply gives the utility of the pen. For any fixed reference point, WTA may depend on λ and η, but will not depend on p.

Now suppose reference points are determined by expectations. If mugs and pens are the same dimension of utility, then the derivation is the same as for Proposition 2, and raising p increases WTA_{pen}. If mugs and pens are independent dimensions of utility, then when $q = 0$, the utility of getting the pen is $u_{pen} + \eta (u_{pen} - p\lambda u_{mug})$ and that of getting \$$WTA_{pen}$ is $WTA_{pen} + \eta (WTA_{pen} - p\lambda u_{mug})$. At the indifference point, these expressions must be equal; p cancels and WTA_{pen} is independent of the treatment. The Appendix takes the case where $q > 0$ and shows

---

that WTA_{pen} does not depend directly on p but does depend positively on WTA_{mug}, as WTA_{mug} will affect the amounts of money the individual expects to receive. This effect is proportional to q. We use the implicit function theorem to show that $\frac{dWTA_{pen}}{dWTA_{mug}} > 0$ for interior values of WTA_{pen}, WTA_{mug}. ∎

The proposition shows that there is a tradeoff: while controlling for WTA_{pen} would not give us a false positive (a positive estimate when the true effect is zero), doing so might attenuate our estimate of the effect of expectations on WTA_{mug}. If reference points are determined by expectations, WTA_{pen} may be affected by p for two reasons. First, mugs and pens may overlap in dimensions of utility, if for instance the relevant utility dimension were "university merchandise." In this case, p will have a similar effect on WTA_{pen} and WTA_{mug}. Second, p may affect WTA_{pen} indirectly: since p affects WTA_{mug}, it alters individuals' expectations of how much money they will receive. This indirect effect on WTA_{pen} is of the same sign as the effect on WTA_{mug}, and is proportional to q, which is small in our context and thus likely undetectable.[20]

Our results below show that controlling for WTA_{pen} increases the precision of our treatment effect on WTA_{mug} without reducing the point estimate. We cannot statistically reject a zero effect of p on WTA_{pen}, consistent with "mug" and "pen" being in different utility dimensions and the indirect effect through the expected money receipt being small (due to q being low).

3.2 Procedures

Upon arrival at the lab, each subject is seated at a carrel with a mug on it. We then flip a coin in front of each subject individually. The coin's sides are labeled "1" and "8", and we give each subject an index card with their resulting number on it. Subjects then start reading instructions on the computer screen in front of them. Subjects are told that if a ten-sided die that we roll individually for each subject at the end of the study comes up lower than the number on their index card, they will receive the mug in front of them for free and leave with it at the end of the experiment.[21] They are also told that if the die comes up 9, they will have the option to keep the mug or exchange it for a randomly determined amount of money between \$0 and \$10.

---

[20]Another way in which WTA_{mug} could affect WTA_{pen} is via anchoring, which would raise WTA_{pen} when p is high and also attenuate our estimated effect. We find no evidence of this.

[21]While the ten-sided die we used for Experiment 1 had numbers from 1 to 10, the die we used for this study had numbers from 0 to 9.

Thus, subjects whose coin came up "1" have a 10% chance of receiving the mug without making a choice (we will refer to them as being in *Treatment M^L*, for low probability of getting the mug) while subjects who got an "8" have an 80% chance of receiving the mug without making a choice (*Treatment M^H*). Subjects in both treatments have a 10% chance of having a choice between the mug and money.

The next phase of the experiment is then identical to Experiment 1. The experimenter reads the most important parts of the instructions out loud; subjects fill out a personality questionnaire in two parts, and get reminded in the middle that they may get the mug for free or have the possibility to choose between the mug and money. They are then told that they will now make choices (on a list) between different monetary amounts or keeping the mug (and not getting the money), and that, if the die comes up 9, they will receive their choice from one randomly selected row. Subjects then make their choices for dollar amounts ranging from $0 to $9.57, in increments of $0.33.[22] They are also told that their decisions will only be revealed to the experimenters in case the die comes up such that one of their choices applies.

Once they have made their choices for the case in which the die comes up 9, they get presented with a page of instructions that informs them that if the die comes up 8 (in which case they previously expected to get neither the mug nor money), they will get their choice between a pen and a randomly determined dollar amount. It is made very clear to the subjects that their choice for this contingency does not in any way influence what they will get in case the die does not come up 8. They are handed a pen to inspect, and are then again asked to make choices between keeping the pen and different dollar amounts ($0 to $9.57, in $0.33 increments).

The experiments were again run at the Harvard Decision Science Laboratory. A total of 112 subjects (66 females; mean age 22), all of them either undergraduate (83 subjects) or graduate (29 subjects) students at Harvard, participated. We conducted 16 sessions with between 3 and 11 subjects per session, across five days in April and May 2010. Subjects received a show-up fee of $10, and the experiment took about 20 minutes.[23]

---

[22]The increments were chosen in order to avoid focal numbers as much as possible. The software enforced a single switching point and required that subjects make a choice in all the rows. For more details, see the Methods Appendix.

[23]As in Experiment 1, the subjects afterwards participated in a second, completely unrelated pilot experiment, and the fact that these experiments were unrelated was made very clear to subjects.

3.3 Results

RESULT 2: *Subjects that have a high chance of leaving with the mug value the mug significantly higher than subjects that have a low chance of leaving with the mug. We estimate that being in treatment M^H instead of M^L increases willingness-to-accept for the mug by 20-30%.*

Based on the individual coin flips, 52 subjects were in treatment M^H and 60 subjects in treatment M^L. We start by simply comparing the mean WTA for the mug across treatments.[24] On average, subjects in the M^H treatment request \$4.12 to give up the mug, while the mean WTA of subjects in the M^L treatment is \$3.74. While these averages are in the right direction to support our theoretical predictions, they are far from statistically significantly different ($p=0.44$, t-test[25]), due to the large between-subjects variation (the standard deviation of WTA_{mug} is around 2.5 in both treatments).

However, as explained in Section 3.1.2, we also elicit WTA for the pen, WTA_{pen}, to control for individual-specific variation in WTA for university memorabilia. Looking at subject-level differences between the amounts people are willing to accept in exchange for the mug or for the pen, $WTA_{mug} - WTA_{pen}$, the treatment effect indeed moves closer towards significance: on average, subjects in the M^H treatment require \$0.92 more to give up the mug than to give up the pen, while the corresponding average for the M^L treatment is only \$0.02. This treatment difference is borderline significant ($p=0.058$, t-test).[26]

Differences in absolute values between WTA for the mug and the pen are, however, not the correct functional form according to the theory. As described in Section 3.1.1, the effect of a high probability of getting the mug should be approximately proportional to the consumption utility of the mug, or, as we will implement it, linear in logs. The mean $\ln(WTA_{mug})$ is 1.30 in Treatment

---

[24]In our main analysis, we measure WTA as the midpoint between the highest amount for which the subject prefers keeping the item and the lowest amount at which she prefers returning the item in exchange for the money. For subjects who prefer keeping the item for all amounts, we set their WTA to \$9.57 (the highest amount we offer them). We control for this censoring in some of our regressions (by using Tobit or interval regressions) and find that it does not materially affect our results.

[25]All tests reported in this paper are two-sided. While in this section we report t-tests (assuming unequal variances), nonparametric Fisher-Pitman permutation tests in all cases give very similar results, often with slightly smaller p-values. (The Fisher-Pitman test is a more powerful alternative to the better known Wilcoxon-Mann-Whitney test; see Kaiser, 2007 for further discussion.)

[26]The average WTA for the pen, WTA_{pen}, is somewhat higher in the M^L treatment (\$3.72) than in the M^H treatment (\$3.20), but the means are not statistically significantly different ($p=0.27$, t-test; $p=0.61$ if instead we use $\ln(WTA_{pen})$).

M^H and 1.11 in Treatment M^L; a difference that is not quite significantly different (p=0.17, t-test). However, when we consider the subject-level difference between $\ln(WTA_{mug})$ and $\ln(WTA_{pen})$, we find a mean of 0.33 for subjects in the M^H treatment and 0.01 for subjects in the M^L treatment. This difference is statistically significant (p=0.03, t-test).[27]   Thus, controlling for idiosyncratic variation in WTA enables us to reduce noise and detect the treatment effect.

Regression analysis allows for a more flexible relationship between $\ln(WTA_{mug})$ and $\ln(WTA_{pen})$, and allows us to control for demographic characteristics and potential day-specific effects that may influence individuals' preference for the mug versus the pen. It can also address the censoring that occurs in a few cases at the top or the bottom of the WTA scale. Our main regression equation is

$$\ln\left(WTA_{mug,i}\right) = \alpha + \beta \cdot M_i^H + \gamma \ln\left(WTA_{pen,i}\right) + \psi' X_i + \epsilon_i$$

where M_i^H is an indicator that equals one for subjects in treatment M^H. If expecting to get the mug with a high probability increases WTA for the mug, as predicted by the KR theory, we expect $\beta > 0$. If instead preferences are reference-independent, or if reference points are given by the status quo, we expect $\beta = 0$.

Table 2 shows that we find strong support for the KR prediction. Column (1), where we regress $\ln(WTA_{mug,i})$ on the treatment indicator M_i^H only, restates the result that without controlling for WTA_{pen}, β is positive but not quite statistically significant. However, as predicted in the discussion in Section 3.1.2, once we add $\ln(WTA_{pen,i})$ as an additional regressor in order to reduce idiosyncratic noise, β increases in magnitude to 0.27 and becomes significant at $p < 0.05$ (column (2)). It remains significant and increases even further in magnitude (to 0.3) if we add demographic and day controls (column (3); note that the signs of the coefficients on age and female are consistent with what was found in Experiment 1) or use a Tobit that accounts for the few censored observations (at $WTA_{mug} = 9.57$; column (4)). In column (5), we allow for a more flexible relationship between $\ln(WTA_{mug,i})$ and $WTA_{pen,i}$ by using a cubic function in $WTA_{pen,i}$ instead of $\ln(WTA_{pen,i})$. This also enables us to include the two subjects with $WTA_{pen,i} = 0$ in the sample. The point estimate of the treatment effect decreases slightly but remains significant at $p < 0.05$.

---

[27]The nonparametric Fisher-Pitman test yields p=0.02; the less powerful Wilcoxon-Mann-Whitney test p=0.09. Subjects that indicate a \$0 WTA for either the mug or the pen, meaning they would rather have nothing than leave with the item, get dropped when we use $\ln(WTA)$. There are five such subjects in M^H (two of which indicate a \$0 WTA for both items) and three in M^L.

Taken together, these results indicate that WTA for the mug increases by 20-30% when subjects have a high probability of getting the mug. In Table A.2 in the Empirical Appendix, we conduct a number of additional robustness checks which show that our result is robust to adding an interaction term $M_i^H \cdot \ln(WTA_{pen,i})$, session fixed effects, or to the use of an interval regression that accounts for the fact that we only observe bounds on subjects' WTA. Also, we show that if we run regressions in levels rather than in logs, the estimated treatment effect is generally borderline significant (at $p < 0.1$, and at $p < 0.05$ if we use interval regression) once WTA_{pen} and demographic characteristics are added as regressors.

4 Discussion

4.1 Psychological Mechanisms

4.1.1 Gain-Loss Utility or Motivated Taste Change?

The results from both our experiments show that individuals display a reluctance to give up items they expect to own. Our findings are predicted by theories of expectation-based reference points, such as KR, in which expectations affect individuals' gain-loss utility. However, expectations could also directly affect perceived consumption utility, that is, how desirable one finds an item. Following Strahilevitz and Loewenstein (1998), we will refer to this theory as "motivated taste change."[28] This theory presumes that preferences are to some extent (consciously or subconsciously) "manipulated" to maximize a utility function that may include things such as self-image, which could in part be determined by the self-perceived desirability of one's possessions. One could formalize such a theory as saying that the perceived consumption utility from an item i that an individual will own with probability p_i is given by $u_i(p_i)$, with $du_i/dp_i > 0$.

These two types of theories are observationally equivalent in our experiments, as well as in many other economically-relevant settings involving valuation of goods. However, the KR theory

---

[28]Strahilevitz and Loewenstein also attempt to test whether their findings of an effect of ownership on valuation are driven by shifting reference points or motivated taste change, and argue that the former theory provides a better explanation of their findings. Using subjective attractiveness questionnaires, they find some evidence for motivated taste change for goods owned over a long period of time (about an hour), but not when ownership had only lasted for a few minutes, as in our experiment. Some researchers in psychology who use similar questionnaires find evidence that people who expect continued ownership of an item like it better even within a few minutes (e.g., Beggan, 1992, Gilbert and Ebert, 2002), while others do not find similar results (e.g., Barone et al., 1997). See Rick (2010) for a recent survey of related topics that also includes a discussion of neuroeconomic evidence on loss aversion.

makes many additional, largely untested, predictions in different domains, with expectation-based reference points and loss aversion as crucial ingredients (see e.g. Kőszegi and Rabin, 2007 on risky choice). If motivated taste change were to explain our experimental results, they would provide less support for the application of theories of expectation-based reference points in other domains.

In an attempt to untangle the two theories, we conduct a variation of Experiment 2 for which the two theories make distinct predictions. In this new experiment, we continue to manipulate the subjects' probability of obtaining one mug for free, p_1. However, we now measure subjects' WTA for a *second*, identical mug in a state of the world in which they are given one mug for free. Thus, diminishing marginal utility of a second mug should not affect WTA differently across treatments.[29]

The gain-loss utility account formalized by KR predicts that WTA for a second mug should not be affected by p_1, since a second mug is a gain with respect to both a reference point of zero mugs or one mug (see Theory Appendix). In contrast, the motivated taste change theory predicts that WTA for the second mug is also increased by a high p_1. This is because in this theory, individuals' beliefs about the desirability of such mugs are affected by their probability of getting one.

The procedures for this experiment are nearly identical to the ones of our valuation experiment described in Section 3.2 (see Methods Appendix). The exception is that subjects are now informed that in case the die comes up 9, they will receive the mug in front of them for free, plus have the possibility to choose between a second, identical mug and a randomly determined amount of money between $0 and $10.

Data from 56 subjects (29 in the high probability treatment, $\tilde{M}^H$) show no support for motivated taste change. Mean WTA for the second mug (WTA_{mug2}) is slightly lower in $\tilde{M}^H$ than in the low-probability treatment $\tilde{M}^L$ ($2.28 v. $2.39), as is the difference between WTA_{mug2} and WTA_{pen} ($-0.89 v. $-0.71).[30] While $\ln(WTA_{mug2})$ is slightly higher in the high-probability treatment (0.75 v. 0.72), the mean subject-level difference between $\ln(WTA_{mug2})$ and $\ln(WTA_{pen})$ is lower in treatment $\tilde{M}^H$ (-0.29) than in treatment $\tilde{M}^L$ (-0.20). As a reminder, in Section 3.3 the

---

[29]A similar experiment is conducted by Morewedge et al. (2009), who report that people who already received a mug for free have a significantly higher WTP for a second mug than non-owners do for a first mug (or per mug if given the opportunity to buy two), consistent with the idea that owners like the mug better. The reasons behind our differential findings could be due to differences in how subjects are randomized into treatments, or to the fact that in Morewedge et al.'s setting ownership was certain, not probabilistic.

[30]Unsurprisingly, WTA for a second mug is thus quite a bit lower than WTA for the first mug found in Section 3.3, which averaged $3.92 across the two treatments. One might be concerned that our finding of no treatment differences is due to many subjects having very low WTA_{mug2}; however, only seven subjects state WTA_{mug2} below $0.50.

mean log-difference between WTA for one mug and WTA for the pen was 0.32 higher in the high-probability treatment. We can reject the null hypothesis that the effect of being in the high-probability treatment on $\ln(WTA_{mug2}) - \ln(WTA_{pen})$ equals 0.32 with $p < 0.04$ (t-test). Similarly, the coefficient on $\tilde{M}^H$ in a regression parallel to column (2) in Table 2 equals -0.03, and the null hypothesis that this coefficient equals 0.266 (the treatment effect of M^H estimated in Experiment 2) is rejected at the 10% level.

Thus, the results of this experiment provide no evidence in favor of the hypothesis that a higher probability of ending up with a mug increases subjects' perceived consumption utility from the mug: the estimated treatment effect mostly goes in the opposite direction. This evidence suggests that the findings in our main experiments were driven by gain-loss utility around expectation-based reference points, not motivated taste change.

4.1.2   How Do Reference Points Form?

The discussion in the previous subsection leaves open the question of what psychological mechanism is behind the formation of expectation-based reference points and gain-loss utility. One candidate mechanism is that a higher likelihood of getting an item may increase the time or intensity with which an individual thinks about the item. This could increase the weight on the item in the person's reference point, which she compares her subsequent position to. In turn, this makes it more likely that the reluctance to incur a loss compared to this reference point outweighs the potential consumption utility surplus the alternative item may provide. A related mechanism is proposed by "Query Theory" (Johnson et al., 2007), which argues that values are constructed by posing queries to oneself and that the order of these queries matters. In our context, individuals with high expectations of getting the mug may pose themselves different queries and focus more on the mug than on the alternative (pen or money), producing a shift of the reference point and a reluctance to give up the mug.

Questions posed to subjects in Experiment 1 provide suggestive evidence consistent with these mechanisms. After they made their choices (but before the die was rolled to determine whether their decision applied), subjects were asked to indicate agreement or disagreement with the following statements on a scale from 1 (disagree strongly) to 5 (agree strongly): (i) I like the mug better than the pen; (ii) Since the beginning of the session, I have spent some time thinking about how I

21

would use the pen; (iii) Since the beginning of the session, I have spent some time thinking about how I would use the mug; (iv) Since the beginning of the session, I have spent more time thinking about the mug than about the pen.

Table A.1 in the Empirical Appendix summarizes the results. Consistent with the choice evidence, subjects in Treatment T^L (mildly) significantly more strongly agree with the statement that they like the mug better than the pen. Subjects in both conditions are about equally likely to agree to having spent *some* time thinking about using each item. However, subjects in Treatment T^L more strongly agree to having spent *more* time (between the moment we explained the decision situation and the moment they made their decision) thinking about the mug than about the pen (means: 3.95 v. 3.17; Wilcoxon-Mann-Whitney $p=0.056$). This result is consistent with the psychological mechanisms discussed above.

4.2 Subject Misconceptions and Transparent Randomization

Subjects in experiments make inferences from the decision environment they are put in. While this fact is true in general, Plott and Zeiler (henceforth PZ) (2007) suggest that it is particularly important to control for such inferences in experiments that try to test whether reference-dependent preferences affect exchange behavior. They argue that if subjects in the typical exchange experiment are not told that which item they receive first was determined randomly, they may (mistakenly) infer that this item is superior to the one they can exchange it for, which makes them reluctant to exchange. To prevent this, PZ explicitly tell subjects that which item they get first was randomly determined, as will be discussed further in the next subsection.

In a pilot for our exchange experiment, we found evidence that subject misconceptions from non-transparent randomization do indeed matter for behavior. Before settling on the design reported in Section 2, we ran sessions in which we did not make the random assignment to treatments obvious to subjects. A total of 63 subjects participated in 15 sessions conducted at the end of July 2009; 32 were randomly assigned (without their knowledge) to Treatment $\tilde{T}^L$ and 31 to Treatment $\tilde{T}^H$. Subjects in these treatments had the same probabilities of being able to exchange as in T^L and T^H, respectively, and received very similar instructions, except that they were not told the source of the probability they would be permitted to exchange (see Methods Appendix for more details).

The results went in the opposite direction of the ones in our main sessions reported in Section 2.3.

In Treatment $\tilde{T}^H$, 29.0% of subjects chose to exchange, while in Treatment $\tilde{T}^L$, 62.5% chose to exchange, and the difference in proportions is statistically significant at $p = 0.011$ (Fisher's exact test).

After the first few sessions, we realized that this may have been due to a value inference effect: subjects who were given a low (10%) probability of being able to exchange their mug for the pen may have inferred that the pen must therefore be more valuable. We then added a question to the debriefing survey in which we asked subjects which item they believe has higher retail value, on a five-point scale from "definitely pen" to "definitely mug." Consistent with the value inference hypothesis, among the 19 subjects in Treatment $\tilde{T}^L$ who answered the question, 13 (or 68.4%) indicated that they believed that the pen "definitely" or "probably" had higher retail value, while the same was true for only four out of 14 (28.6%) of subjects in Treatment $\tilde{T}^H$. A Wilcoxon-Mann-Whitney test indicates that the distributions of responses to the question are significantly different at $p < 0.03$.

This value inference effect disappears when randomization is made transparent by flipping a coin in front of each subject.[31] Thus, we conclude that, in the pilot, our test of expectation-based reference points was confounded by an experimental design that created subject misconceptions about the relative values of the two items. The pilot may also speak to the relative strength of expectation-based reference points versus value inference. While we find evidence in favor of the KR theory in our clean treatments, the effect of loss aversion can be more than outweighed by (perceived) value signals provided by probabilities. This observation was part of the motivation behind Experiment 2, the goal of which was to get a quantitative measure of the strength of the effect of expectation-based reference points on valuation.

Misconceptions from non-transparent randomization may affect the results of other research. For instance, in Abeler et al. (2009), subjects are randomized into a low or high fixed payment condition. They are then equally likely to be paid a piece rate or to receive this fixed payment and only learn which case applies after they stop working. Abeler et al. find that subjects in the high fixed payment condition work more and that a significant fraction of subjects stop working right

---

[31]The final row of Appendix Table A.1 shows that the answers to the debriefing question about relative retail value do not differ significantly across treatments in our main sessions, suggesting that our explicitly random assignment to treatments prevented subjects from inferring anything about the relative values of the two items. We asked the same question after Experiment 2, also finding only small and insignificant treatment differences (means: 2.75 in Treatment M^H, 2.68 in Treatment M^L; Wilcoxon-Mann-Whitney $p=0.71$).

when their piece rate earnings equal the fixed payment they may receive. These findings, like our results, are consistent with subjects having reference-dependent preferences with a reference point determined by expectations. However, because randomization into treatments was not transparent to subjects in Abeler et al.'s experiment, their results could also be driven in part by subjects' (mistaken) inference. In particular, subjects may have taken the fixed payment as a signal of the expected effort level in the experiment (for instance, they could think that the fixed payment they may receive instead of their piece rate earnings was calibrated to be close to the piece rate earnings the average participant accumulates), and their behavior may (consciously or unconsciously) be affected by this.[32]

4.3 Reconciling Previous Experiments

4.3.1 Exchange Behavior

Our findings, along with the theory of expectation-based reference points, can help reconcile the varied findings of previous experiments on the presence or absence of the endowment effect and on WTP-WTA gaps. The classic early experiment demonstrating the endowment effect was conducted by Knetsch (1989), who endowed subjects with either a mug or a chocolate bar and found that substantially fewer subjects exchanged their item for the alternative than predicted by classical theories. Such exchange asymmetries have usually been interpreted as resulting from loss aversion around a reference point given by current endowments. However, a natural interpretation in terms of expectation-based reference points is that, until the opportunity to exchange is offered to participants, they fully expect to leave the experiment with the item they were endowed with, so that expectations and endowments coincide.

PZ (2007) argue that Knetsch's findings (and those of other researchers who replicated Knetsch's experiment) were largely driven by certain features of his experimental procedures. PZ alter Knetsch's procedures in various ways and demonstrate that such changes can have a large impact on the existence and magnitude of exchange asymmetries. They interpret this result as evidence

---

[32]Note that this concern is different from the alternative concern that subjects stop working when their piece rate earnings equal the fixed payment because the amount of the fixed payment is particularly salient and may thus serve as a focal point or anchor. Abeler et al. can rule out this explanation for their findings through some clever additional treatments. They also conduct a further additional treatment to show that their result is not driven by gift exchange motives (i.e., subjects wanting to reciprocate in response to the generosity of the possible fixed payment).

that the endowment effect observed in earlier studies is not due to non-standard features of subjects' preferences. However, when experimental procedures are altered, subjects' expectations of which item(s) they will leave the experiment with may also change. In particular, PZ's procedures may have made subjects believe that they would also be given a pen, or that there would be an opportunity to exchange the mug for the pen.[33] In such a case, the expectation-based reference point would be the same, regardless of initial endowment, and no exchange asymmetry is predicted by KR.[34]

Thus, while PZ show that the endowment effect can disappear when certain experimental procedures are used, they do not directly manipulate or measure subjects' expectations that they will leave with a mug or pen. They argue that previously observed exchange asymmetries are due solely to classical preferences interacting with experimental procedures. However, their results are also consistent with the endowment effect being driven by reference-dependent preferences, but with a reference point determined by expectations, not current endowments.

Other work has found an "endowment effect" even when subjects do not formally own an item. Knetsch and Wong (2009) find exchange asymmetries in an experimental treatment in which they give subjects an item, but tell them they do not yet own the item (but will own it at the end of the session). They interpret this finding as evidence in favor of KR, and interpret the PZ procedures as providing only a "weak reference state." In this treatment, there is no upfront announcement to subjects that they will be able to alternatively get the other item at the end. In another treatment, Knetsch and Wong closely follow the PZ procedures, except that they explicitly tell subjects that at the end they will have the option to exchange, and confirm the absence of an endowment effect under these conditions.[35]

---

[33]Their two treatments in which no exchange asymmetries are observed, the "full set of controls" and the "loss emphasis treatment," begin as follows (p. 1459): "We began these sessions by informing the subjects that mugs and pens would be used during the experiment. Subjects were then told that a coin was flipped before the start of the experiment to determine which good, the mug or the pen, would be distributed first. We then distributed mugs to the subjects and announced, 'These mugs are yours'." In their "loss emphasis treatment," PZ use somewhat stronger language to convey subjects' entitlement to the endowed good; they say "The mug is yours. You own it."

[34]A subject who expects to get both items will necessarily feel a loss in one dimension when she gets surprised by the announcement that she can only leave with one of them, so she will choose the one with the higher consumption utility. Similarly, a subject who expects to be able to exchange for sure will plan on leaving with the item with the higher consumption utility (this is the $p \to 1$ case in Proposition 1), independently of which item she is endowed with.

[35]Knetsch and Wong have a third treatment, in which they find a significant exchange asymmetry even though subjects are told early on that they will have the option to exchange their item if they would like. They speculate that this finding may be due to the randomization procedure used, but it could also be due to the classic endowment effect.

4.3.2　Valuation Behavior: WTP and WTA

A robust debate continues on when WTP-WTA gaps exist and whether they should be interpreted as evidence for reference-dependent preferences (see Isoni, Loomes, and Sugden 2010 and the reply by PZ 2010). In a classic paper, Kahneman et al. (1990) find that initial ownership of an item seems to affect valuation of that item. In a series of laboratory experiments, they observe large gaps between WTP and WTA for mugs, even though initial endowments were randomly assigned. These results have been interpreted as evidence for prospect theory and loss aversion. PZ (2005) reexamine this paradigm and dispute that the mere fact of endowment affects preferences. They instead argue that subject misconceptions drive the WTP-WTA gap and show that with extensive training in the BDM mechanism and anonymity, there is no statistically significant gap between WTP and WTA for mugs.[36] As in the previous subsection, reference points that are determined by expectations can reconcile the differential findings. Under this view, WTP-WTA gaps are predicted only when buyers and sellers of the good have different expectations of keeping the good. In the original Kahneman et al. experiments, endowments may (or may not) have induced an expectation of continued ownership of the good. Either way, we agree with PZ that many factors other than reference dependence—bargaining heuristics, misunderstanding the BDM mechanism, etc.—may contribute to differences between WTP and WTA in these experiments. Yet just as in the exchange experiments, the procedures and training used by PZ may not have induced a difference between buyers and seller in their expectations of keeping the mug, as subjects may anticipate the possibility of trade. Thus, the absence of WTP-WTA gaps in their experiments is not surprising when reference points are determined by expectations.

Our second experiment, which directly manipulates expectations and holds everything else constant across treatments, provides clean evidence of a statistically and economically significant effect of reference dependence on valuation. We cannot directly speak to whether WTP-WTA gaps observed elsewhere are due to differences in expectations as opposed to a possible direct effect of endowment or other factors, but our results suggest that researchers examining WTP-WTA gaps should directly induce or at least elicit buyers' and sellers' expectations that they end up with the good they are buying/selling.

---

[36]However, both Plott and Zeiler (2005) and Isoni et al. (2010) do find WTP-WTA gaps in lotteries. Isoni et al. and Plott and Zeiler (2010) debate various reasons for this differential finding.

In sum, PZ (2005, 2007) show that the mere fact of endowment, independent of expectations, may not lead to increased valuation of a good. However, their results should not be taken as an indication that in general, reference-dependent preferences do not matter for exchange and valuation behavior, since the reference point may be determined by expectations which do not necessarily correspond to current endowments. Our results suggest that to the extent endowments affect expectations (as they often do outside the lab), endowments will lead to increased valuation of a good.

5 Conclusion

In two simple experiments, increasing subjects' expectations of leaving with an item has an effect on exchange behavior (they are more likely to choose to keep the item) and valuation (they demand more compensation to give up the item). We thus provide evidence that individuals' reference points are determined, at least in part, by expectations.

Our findings support the assumption of expectation-based reference points made by Kőszegi and Rabin (2006) and subsequent applications of the theory. Furthermore, they suggest that the findings of recent research that identifies settings in which the endowment effect does not appear should not necessarily be interpreted as evidence against the importance of reference-dependent preferences and loss aversion. The results from our experiments indicate that an "endowment effect" of some sort is real, but operates via expectations instead of formal ownership.

The implications of expectation-based reference points for policy are often similar to those of status-quo-based reference points, since ownership often carries with it the expectation of keeping a good. Yet there may be consequential differences. For instance, loss aversion may not inhibit trade in markets where individuals expect to trade, such as markets for tradeable pollution (e.g. carbon) permits. Moreover, in contingent valuation studies, individuals' prior perception of the probability a policy will be implemented may influence their WTP or WTA for the policy. This should be considered when interpreting the results from such studies for policy evaluation.

While we show that expectations are an important determinant of reference points, we cannot rule out that other factors, such as social norms, aspirations, salience, and history, may also influence the reference point. Untangling these factors suggests an interesting direction for future research.

References

ABELER, J., A. FALK, L. GÖTTE, AND D. HUFFMAN (2009): "Reference Points and Effort Provision," *American Economic Review*, forthcoming.

BARONE, M. J., T. A. SHIMP, AND D. E. SPROTT (1997): "Mere Ownership Revisited: A Robust Effect?" *Journal of Consumer Psychology*, 6, 257–284.

BECKER, G. M., M. H. DEGROOT, AND J. MARSCHAK (1964): "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science*, 9, 226–32.

BEGGAN, J. K. (1992): "On the Social Nature of Nonsocial Perception: The Mere Ownership Effect," *Journal of Personality and Social Psychology*, 62, 229–237.

BELL, D. E. (1985): "Disappointment in Decision Making Under Uncertainty," *Operations Research*, 33, 1–27.

BENARTZI, S. AND R. THALER (1995): "Myopic loss aversion and the equity premium puzzle," *Quarterly Journal of Economics*, 110, 73–92.

CAMERER, C., L. BABCOCK, G. LOEWENSTEIN, AND R. THALER (1997): "Labor Supply of New York City Cabdrivers: One Day At A Time," *Quarterly Journal of Economics*, 112, 407–441.

CARD, D. AND G. B. DAHL (2010): "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics*, forthcoming.

CRAWFORD, V. P. AND J. MENG (2010): "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, forthcoming.

DELLAVIGNA, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

ENGELMANN, D. AND G. HOLLARD (2010): "Reconsidering the Effect of Market Experience on the "Endowment Effect"," *Econometrica*, forthcoming.

FEHR, E. AND L. GÖTTE (2007): "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 97, 298–317.

GENESOVE, D. AND C. MAYER (2001): "Loss aversion and seller behavior: Evidence from the housing market," *Quarterly Journal of Economics*, 116, 1233–1260.

GILBERT, D. T. AND J. E. J. EBERT (2002): "Decisions and Revisions: The Affective Forecasting of Changeable Outcomes," *Journal of Personality and Social Psychology*, 82, 503–514.

GILL, D. AND V. PROWSE (2010): "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *Discussion Paper No. 1006, Economics Division, School of Social Sciences, University of Southampton.*

GUL, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686.

HARDIE, B., E. JOHNSON, AND P. FADER (1993): "Modeling loss aversion and reference dependence effects on brand choice," *Marketing Science*, 12, 378–394.

HEIDHUES, P. AND B. KŐSZEGI (2008): "Competition and price variation when consumers are loss averse," *American Economic Review*, 98, 1245–1268.

——— (2010): "Regular Prices and Sales," *Working Paper, UC Berkeley.*

HERWEG, F. (2010): "Uncertain Demand, Consumer Loss Aversion, and Flat-Rate Tariffs," *Bonn Econ Discussion Papers.*

HERWEG, F., D. MÜLLER, AND P. WEINSCHENK (2009): "Binary Payment Schemes: Moral Hazard and Loss Aversion," *American Economic Review*, forthcoming.

ISONI, A., G. LOOMES, AND R. SUGDEN (2010): "The Willingness to Pay–Willingness to Accept Gap, the "Endowment Effect", Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: A Reassessment," *American Economic Review*, forthcoming.

JOHN, O. P., E. M. DONAHUE, AND R. L. KENTLE (1991): "The Big Five Inventory—Versions 4a and 54," *Berkeley: University of California, Berkeley, Institute of Personality and Social Research.*

JOHNSON, E. J., G. HÄUBL, AND A. KEIMAN (2007): "Aspects of Endowment: A Query Theory of Value Construction," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 461–474.

KAHNEMAN, D., J. L. KNETSCH, AND R. H. THALER (1990): "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98, 1325–1348.

KAHNEMAN, D. AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292.

KAISER, J. (2007): "An exact and a Monte Carlo proposal to the Fisher–Pitman permutation tests for paired replicates and for independent samples," *The Stata Journal*, 7, 402–412.

KŐSZEGI, B. AND M. RABIN (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133–1165.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047–1073.

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99, 909–936.

KNETSCH, J. L. (1989): "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *American Economic Review*, 79, 1277–1284.

KNETSCH, J. L. AND W.-K. WONG (2009): "The endowment effect and the reference state: Evidence and manipulations," *Journal of Economic Behavior & Organization*, 71, 407–413.

LANGE, A. AND A. RATAN (2010): "Multi-dimensional reference-dependent preferences in sealed-bid auctions—how (most) laboratory experiments differ from the field," *Games and Economic Behavior*, 68, 634–645.

LIST, J. A. (2003): "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 118, 41–71.

LOOMES, G. AND R. SUGDEN (1986): "Disappointment and Dynamic Consistency in Choice under Uncertainty," *Review of Economic Studies*, 53, 271–282.

MACERA, R. (2010): "Intertemporal Incentives Under Loss Aversion," *Working Paper, UC Berkeley*.

MAS, A. (2006): "Pay, Reference Points, and Police Performance," *Quarterly Journal of Economics*, 121, 783–821.

MOREWEDGE, C., L. SHU, D. GILBERT, AND T. WILSON (2009): "Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect," *Journal of Experimental Social Psychology*, 45, 947–951.

ODEAN, T. (1998): "Are investors reluctant to realize their losses?" *Journal of Finance*, 53, 1775–1798.

PLOTT, C. R. AND K. ZEILER (2005): "The Willingness to Pay–Willingness to Accept Gap, the 'Endowment Effect,' Subject Misconceptions, and Experimental Procedures for Eliciting Valuations," *American Economic Review*, 95, 530–545.

——— (2007): "Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory?" *American Economic Review*, 97, 1449–1466.

——— (2010): "Endowment Effect Theory, Subject Misconceptions and Enhancement Effect Theory: A Reply to Isoni, Loomes and Sugden," *American Economic Review*, forthcoming.

POPE, D. G. AND M. E. SCHWEITZER (2009): "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, forthcoming.

POST, T., M. VAN DEN ASSEM, G. BALTUSSEN, AND R. THALER (2008): "Deal or no deal? Decision making under risk in a large-payoff game show," *American Economic Review*, 98, 38–71.

RICK, S. (2010): "Losses, Gains, and Brains: Neuroeconomics Can Help to Answer Open Questions about Loss Aversion," *Journal of Consumer Psychology*, forthcoming.

SAMUELSON, W. AND R. ZECKHAUSER (1988): "Status quo bias in decision making," *Journal of Risk and Uncertainty*, 1, 7–59.

SMITH, A. (2008): "Lagged Beliefs and Reference-Dependent Utility," *Working Paper 08-03, Department of Economics, University of Arizona.*

STRAHILEVITZ, M. AND G. LOEWENSTEIN (1998): "The Effect of Ownership History on the Valuation of Objects," *Journal of Consumer Research*, 25, 276–289.

THALER, R. (1980): "Toward a positive theory of consumer choice," *Journal of Economic Behavior & Organization*, 1, 39–60.

TVERSKY, A. AND D. KAHNEMAN (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106, 1039–1061.

Table 1: Determinants of Desire to Exchange Mug for Pen

Pr(choose to exchange)	(1)	(2)	(3)	(4)
Treatment T^H	.338**		.353**	.402**
	(.137)		(.139)	(.154)
Female		.249*	.270*	.291*
		(.141)	(.145)	(.163)
Age				.103
				(.066)
Day indicators	No	No	No	Yes
# observations	45	45	45	45

Notes: Displayed coefficients are predicted marginal effects from probit regressions (in case of dummy variables, for a discrete change from 0 to 1). Dependent variable: Indicator variable = 1 if subject wants to exchange. Standard errors in parentheses.

Level of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

Table 2: Determinants of Willingness-to-Accept for the Mug

	(1)	(2)	(3)	(4)	(5)
Treatment M^H	.194	.266**	.306**	.308**	.290**
	(.143)	(.124)	(.128)	(.126)	(.13)
$\ln(WTA_{pen})$		.524***	.560***	.568***	
		(.091)	(.093)	(.091)	
Female			-.199	-.213*	-.191
			(.131)	(.128)	(.133)
Age			-.052*	-.050*	-.046*
			(.027)	(.026)	(.027)
Day indicators	No	No	Yes	Yes	Yes
Other controls					Cubic in WTA_{pen}
# observations	106	104	104	104	106

Notes: Dependent variable: $\ln(WTA_{mug})$. (1), (2), (3) and (5) are OLS regressions; (4) displays predicted marginal effects from a Tobit that takes into account censoring at $WTA_{mug} = 9.57$. All regressions contain a constant, and except for (1) also an indicator for $WTA_{pen} = 9.57$ (which indicates that a subject's WTA for the pen may be censored; $p > 0.3$ in all regressions). Regressions with $\ln(WTA_{pen})$ as explanatory variable drop two subjects with $WTA_{pen}=0$. Standard errors in parentheses.

Level of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

A   Theory Appendix

A.1   General Theory

The main objective of KR (2006) is to extend Kahneman and Tversky's (1979) prospect theory and to bring it closer to "standard" economic theory. In their formulation, the utility of a decision-maker (henceforth DM) depends both on her consumption c and on how this consumption compares to a reference level r. They assume separability between "consumption utility" $m(c)$ and "gain-loss utility" $n(c|r)$. Consumption and reference levels can have multiple dimensions, so that $c = (c_1, c_2, ..., c_K)$ and $r = (r_1, r_2, ..., r_K)$, and it is assumed that both forms of utility are additively separable across dimensions, such that $m(c) = \sum_{k=1}^{K} m_k(c_k)$ and $n(c|r) = \sum_{k=1}^{K} n_k(c_k|r_k)$. Total reference-dependent utility (henceforth RDU) of an outcome c given reference level r is then given by

$$u(c|r) = \underbrace{\sum_{k=1}^{K} m_k(c_k)}_{cons.\ utility\ m(c)} + \underbrace{\sum_{k=1}^{K} \mu(m_k(c_k) - m_k(r_k))}_{gain-loss\ utility\ n(c|r)},$$

where $\mu(\cdot)$ is continuous and strictly increasing, and has the properties of the Kahneman-Tversky value function: loss aversion (a kink at zero) and diminishing sensitivity. In what follows, we will use the piecewise linear specification (as in Section IV of KR, 2006) $\mu(x) = \eta x$ for $x > 0$ and $\mu(x) = \eta \lambda x$ for $x \leq 0$, where $\eta > 0$ is the DM's weight on gain-loss utility and $\lambda > 1$ her loss aversion coefficient.

In KR, the reference point is given by probabilistic beliefs about future outcomes, so that it is in fact a reference lottery given by probability measure $G : \mathbb{R}^K \to \mathbb{R}$ over consumption utility in each of the k dimensions. Thus the RDU of outcome c given expectations G is

$$U(c|G) = \int u(c|r)dG(r).$$

This formulation implies that when evaluating a dimension c_k of a consumption outcome c, the DM compares it separately to each possible value that this dimension could take according to the reference lottery, and weights it by the probability of this outcome in the reference lottery.

But what exactly determines the reference lottery? KR employ the concept of *personal equilibrium* (PE), which requires the following:

- the DM has a plan for every contingency that she can possibly face in a given decision situation;

- her reference lottery is based on her expectation to put these plans into action once she knows which contingency applies; and

- her actions maximize expected RDU, given her reference lottery.

The timing of a decision situation in which PE applies is as follows:

- $t = -1$: DM starts focusing on the decision; knows all the possible choice sets she might face at $t = 1$ and the probabilities with which they occur;

- $t = 0$: DM makes plans for each possible choice set, thereby setting her reference lottery;

- $t = 1$: 'Nature' determines the choice set the DM actually faces. The DM follows through on her plan for this choice set, which then yields either a deterministic or a probabilistic outcome.

The PE concept requires the DM's actions at $t = 1$ to be optimal given her beliefs at $t = 0$, beliefs formed at $t = 0$ to be rational given her plans, and her plans to be "credible" in the sense that at $t = 1$ the DM has no incentive to deviate from them (akin to subgame perfection in standard multi-player game theory).

KR emphasize that it is possible for multiple PEs to exist in a given situation. An example: assume the consumption utilities resulting from actions x and y are not too different. Then, one equilibrium may be for the DM to plan on x and indeed do x, but another one may be for her to plan on y and do y (because if she planned on y and ended up doing x, she would feel a loss from not getting the consequence of y, which may outweigh a possible consumption utility advantage of x). In such a case KR's preferred way of equilibrium selection is to take the PE that gives the DM the highest expected RDU, the *preferred personal equilibrium* (PPE).

A.2 Derivation of KR Portion of Proposition 1

Denote a subject's consumption utility from the mug by u_{mug} and the consumption utility from the pen by u_{pen}.[1] The treatment variable is p, the probability with which the subject expects to have the option to exchange the mug (which she is endowed with) against the pen. The subjects know the choice that they may face, but are unable to commit to a decision until the end of the session. In terms of the timeline from the previous subsection, the moment when we announce to subjects that they own the mug and may have the possibility to exchange the mug against a pen comprises $t = -1$. $t = 0$ is then the time span during which the subjects plan what they are going to do at $t = 1$, and this plan determines their reference lottery.

We now examine how the conditions on u_{mug} and u_{pen} for a subject to choose "exchange" in a PE or (more restrictively) the PPE vary with our treatment variable p.[2]

If a subject plans to choose "exchange" if given the choice, then she expects $r_{end\ w/pen}$ to occur with probability p and $r_{end\ w/mug}$ to occur with probability $1 - p$. Her utility of outcome c, given her expectations, is then $U\left(c|\text{"exchange"}\right) = pu\left(c|r_{end\ w/pen}\right) + (1 - p)u\left(c|r_{end\ w/mug}\right)$.

If she follows through with her plan, her gain-loss utility in the different outcome cases is as follows:

- if her decision applies and she gets the pen (which happens with probability p):

  $p \cdot 0 + (1 - p) \cdot \eta(u_{pen} - \lambda u_{mug})$;

- if it does not apply and she keeps the mug (which happens with probability $1 - p$):

  $p \cdot \eta(u_{mug} - \lambda u_{pen}) + (1 - p) \cdot 0$.

Therefore, if she reaches the state of the world in which she can exchange, and follows through with her plan, her utility is

$$U\left(\{1\text{ pen, }0\text{ mugs}\}|\text{"exchange"}\right) = u_{pen} + (1 - p)\eta(u_{pen} - \lambda u_{mug}), \tag{1}$$

---

[1] Throughout, we assume $u_{pen}, u_{mug} \geq 0$, i.e. that subjects weakly prefer both receiving the mug for sure and receiving the pen for sure to receiving nothing.

[2] We limit our focus to pure-strategy equilibria.

while if she deviated and chose to keep the mug instead, her utility would be

$$U\left(\{0 \text{ pens}, 1 \text{ mug}\} \,|\, \text{``exchange''}\right) = u_{mug} + p\eta(u_{mug} - \lambda u_{pen}). \tag{2}$$

Her plan to exchange if possible is credible and thus a PE if and only if (1) $\geq$ (2), or

$$u_{pen} \geq u_{mug} \underbrace{\frac{1 + \eta(\lambda + p(1 - \lambda))}{1 + \eta(1 - p(1 - \lambda))}}_{\equiv X(p)}. \tag{3}$$

As $\lambda > 1$ and $\eta > 0$, we have that

$$X'(p) < 0, \ X(0) = \frac{1 + \eta\lambda}{1 + \eta} > 1, \ X\left(\frac{1}{2}\right) = 1, \ \text{and } X(1) = \frac{1 + \eta}{1 + \eta\lambda} < 1.$$

In words, this means that as the probability p of getting the possibility to exchange the mug for the pen increases, the consumption utility a subject gets from the pen necessary for "exchange" to be a PE decreases (relative to u_{mug}), such that "exchange" is a PE for more (u_{mug}, u_{pen}) pairs.

Note that in our experiment we elicit a decision from all subjects, independently of whether their decision to exchange or not will in fact matter. In other words, our subjects are required to make a conditional choice before knowing whether they are in the state of the world in which they have the option to exchange. The initial instructions given to subjects did not specify whether they would make their choice before or after the die roll; we now discuss why either way it does not make a difference for the theoretical prediction.

If subjects expected to be asked to make an unconditional choice (i.e., only after the die is rolled), then in terms of the KR theory, eliciting a conditional choice is no different from the alternative in which we first roll the die and then only elicit a decision if the appropriate number comes up. At this point, the plan is made, and whether we elicit it as a plan or as an actual choice should not make a difference, given KR's assumption that reference points are given by lagged expectations, and that in equilibrium, plans have to be consistent with rationally expected behavior (i.e., the subject has no incentive to deviate from her plan once she knows which choice set she actually faces).

We can also see that asking subjects to make a conditional choice does not alter the PE condition

by considering the subject's expected utility from sticking to the plan to say "exchange" once she has learned that she is required to make a conditional choice:

$$
\begin{aligned}
\mathrm{E}U(\text{"exchange"}\,|\,\text{"exchange"}) \;=\;& p\,[U\,(\{1\text{ pen, }0\text{ mugs}\}\,|\,\text{"exchange"})] \\
& + (1-p)\,[U\,(\{0\text{ pens, }1\text{ mug}\}\,|\,\text{"exchange"})] \\
=\;& p[u_{pen} + (1-p)\eta(u_{pen} - \lambda u_{mug})] \\
& + (1-p)[u_{mug} + p\eta(u_{mug} - \lambda u_{pen})].
\end{aligned}
$$

For "exchange" to be optimal, this has to be greater or equal to $U\,(\text{"keep"}\,|\,\text{"exchange"})$, given by (2). This condition boils down to (3). Furthermore, the exact same condition would also be required for "exchange" to be a PE in case subjects expected from the beginning that they would be asked to make a conditional choice.

The above expression for $X(p)$ implies that if p is high $(p > \frac{1}{2})$, "exchange" may be a PE even if $u_{pen} < u_{mug}$. However, in such a case, "keep" is a PE as well, and it may indeed be the PPE (that is, the PE with the highest ex-ante expected utility).[3]

The condition for "exchange" to be the PPE is that $\mathrm{E}U(\text{"exchange"}\,|\,\text{"exchange"}) \geq \mathrm{E}U(\text{"keep"}\,|\,\text{"keep"}) = u_{mug}$, or

$$
p(u_{pen} - u_{mug}) + \eta[p(1-p)(u_{pen} - \lambda u_{mug}) + (1-p)p(u_{mug} - \lambda u_{pen})] \geq 0,
$$

which simplifies to

$$
(u_{pen} - u_{mug}) + \eta\,(1-p)\,(1-\lambda)\,(u_{pen} + u_{mug}) \geq 0.
$$

For $p < 1$, "exchange" can be the PPE only if the pen gives higher consumption utility ($u_{pen} > u_{mug}$), as the second term is negative due to loss aversion ($\lambda > 1$). Crucial for our hypothesis, note that as p is increased, "exchange" is the PPE for more pairs of u_{pen} and u_{mug} (the required consumption utility surplus from the pen becomes smaller), as increasing p makes the second term less negative.

Kőszegi and Rabin (2007) introduce an alternative equilibrium concept, the "choice-acclimating

---

[3]It is easy to show that "keep" is a PE if $u_{pen} \leq \frac{1+\eta\lambda}{1+\eta}u_{mug}$, which is always the case if $u_{pen} < u_{mug}$ and if "exchange" is not a PE.

personal equilibrium" (CPE), which applies in situations where the DM determines the reference lottery by her actual choice, instead of with her plan. In general, the CPE may yield different predictions from the (P)PE concept (which KR 2007 rename "unacclimating personal equilibrium"). However, the CPE applies only in cases in which uncertainty is resolved long after actions are committed to. In both our experiments, subjects cannot commit to their decision before the end of the decision situation, but are given ample time to think about it between the beginning of the experiment and the moment they make their choice, making (P)PE the appropriate concept. Furthermore, in this experiment, CPE would make the exact same prediction as PPE, as the expected utility from saying "keep" is simply u_{mug}, while the expected utility from saying "exchange," knowing that this will then determine the reference lottery, is the same as given in the expression for $EU($"exchange"$|$"exchange"$)$. Thus, the prediction that a higher p should increase the proportion of subjects that indicate that they would like to exchange is robust to the use of all the different equilibrium concepts proposed by KR.

A.3 Derivation of KR Portion of Propositions 2 and 3

In the second experiment, subjects are put in the following situation: at $t = -1$, they are told that at $t = 1$, they will receive a mug with probability $p_i \in \{p_L, p_H\}$. With probability q, they instead enter a Becker-DeGroot-Marschak (1964) (BDM) mechanism where for different amounts of money between \$0 and \$10 they have to indicate whether they prefer receiving the mug or the money. With probability $1 - p - q$, they do not get the mug or a choice between mug and money. $t = 0$ is the time span during which subjects plan what to do at $t = 1$, which determines their reference lottery.

A subject's plan here corresponds to an amount at which the subject is indifferent between receiving \$$W^*_{mug}$ and receiving the mug. Then, when entering the BDM mechanism, the subject (consistent with her plan) chooses the mug for any amount below \$$W^*_{mug}$ and the money otherwise. Proposition 2 states that for any pair of parameters $\eta > 0$ and $\lambda > 1$, there exists a q small enough such that W^*_{mug} unambiguously increases in p for both the PE and the PPE equilibrium concepts.

We give the proof of the proposition for the continuous BDM case; subjects actually make a discrete decision for different amounts, but this does not affect the results. Assume that a price

x is drawn from a uniform distribution over [\$0, \$10].[4] Subjects indicate an indifference point W_{mug}^*, their WTA. (All subjects indicate their W_{mug}^*, even though their decision only matters with probability q. As in Experiment 1, eliciting their plan is theoretically equivalent to eliciting their actual choice if the BDM choice set is reached.) If $x < W_{mug}^*$ the subject keeps the mug and gets no money; if $x \geq W_{mug}^*$, the subject gets \$x but no mug.

A PE requires that subjects' choices must be consistent: a plan will be of the form "choose 'mug' if $x < W_{mug}^*$, choose 'money' if $x \geq W_{mug}^*$," where W_{mug}^* is assumed to be in the interior of the interval $(0, 10)$.[5] Call this plan Ω. For plan Ω to be a PE, it must be implementable, and thus the subject must be indifferent between receiving the mug and receiving \$$W_{mug}^*$ (because if she were not, she would have an incentive to deviate from her plan). Thus, it must be the case that

$$U\left(\{1 \text{ mug}, \$0 | \Omega\}\right) = U\left(\left\{0 \text{ mugs}, \$W_{mug}^* | \Omega\right\}\right)$$

Under plan Ω, the possible outcomes are:

- $\{1 \text{ mug}, \$0\}$ with probability $p + \frac{q}{10} W_{mug}^*$

- $\{0 \text{ mug}, \$0\}$ with probability $1 - p - q$

- $\{0 \text{ mug}, \$x\}$ for $x \in \left[W_{mug}^*, 10\right]$ with probability $\frac{q}{10}\left(10 - W_{mug}^*\right)$

Getting a mug involves a gain of a mug relative to getting nothing and a loss relative to getting \$$x \geq W_{mug}^*$. Similarly, getting \$$W_{mug}^*$ involves a loss of a mug and gain of \$$W_{mug}^*$ relative to getting a mug, a gain of \$$W_{mug}^*$ relative to getting nothing, and a loss of $x - W_{mug}^*$ relative to getting \$$x > W_{mug}^*$. Continue to denote subject i's consumption utility from the mug by u_{mug}, and let utility of money be linear with scale normalized to one. Using the piecewise linear gain loss

---

[4]The statement in Proposition 2 also holds for other distributions, which however make the proof less tractable.

[5]Given our focus on interior equilibria, we can restrict u_{mug} to be smaller than 10, as it can be shown that if $u_{mug} \geq 10$, there cannot exist an equilibrium with $W_{mug}^* < 10$.

function described earlier, we have

$$U\left(\{1\text{ mug}, \$0|\Omega\}\right) = u_{mug} + \eta\left[(1-p-q)\left(u_{mug}\right) + \left(\frac{q}{10}\right)\int_{W^*_{mug}}^{10}(u_{mug} - \lambda x)\,dx\right]$$

$$U\left(\{0\text{ mugs}, \$W^*_{mug}|\Omega\}\right) = W^*_{mug} + \eta\left[\begin{array}{c}(1-p-q)\left(W^*_{mug}\right) + \left(p+\frac{q}{10}W^*_{mug}\right)\left(W^*_{mug} - \lambda u_{mug}\right) \\ -\left(\frac{q}{10}\right)\lambda\int_{W^*_{mug}}^{10}\left(x - W^*_{mug}\right)dx\end{array}\right]$$

where the integrals are the loss from not getting $x > W^*_{mug}$. The indifference condition at W^*_{mug} is then satisfied if the two expressions are equal.

Consider the case when $q = 0$. Then, there is a unique PE $W^*_{mug} = u_{mug}\frac{1+\eta(1+(\lambda-1)p)}{1+\eta}$, which is therefore the PPE. This is the unique solution and it does not depend on the subject's planned action.[6]

When $q > 0$, it is possible that multiple PE exist. Rearranging the indifference condition gives a quadratic equation for W^*_{mug} :

$$\left(W^*_{mug}\right)^2\frac{q}{10}\eta\left(\lambda-1\right) - W^*_{mug}\left[1+\eta\left[1+(\lambda-1)q\left(1-\frac{u_{mug}}{10}\right)\right]\right] + u_{mug}\left[1+\eta\left(1+(\lambda-1)p\right)\right] = 0$$

Write this as $AW^2 + BW + C = 0$. Hence, there are at most two roots of this equation, and only two possible values of W consistent with a PE. Call these values W^+ and W^-. We restrict attention to the lower root W^-, as W^+ gives the perverse case in which WTA for the mug decreases in the consumption utility of the mug ($dW^+/du_{mug} < 0$). Moreover, a sufficient condition for W^+ to be a non-interior root ($W^+ > 10$) is $\frac{1+\eta}{\eta(\lambda-1)} > q\left[1 + \frac{u_{mug}}{10}\right]$, which is satisfied whenever q is low enough. Now, assume the lower root W^- is real, so that we have an interior solution. Then, simple algebra shows that $\frac{dW^-}{dp} > 0$, as $\frac{dW^-}{dp} = \frac{u_{mug}[\eta(1+(\lambda-1))]}{(B^2-4AC)^{1/2}2A} > 0$.

For Proposition 3, consider the surprise WTA elicitation for the pen, the results of which will be implemented with probability q, replacing part of the probability space in which the subject expected to get nothing. Since it is a surprise, subjects give their indifference point W^*_{pen} while reference points stay fixed (because they are assumed to be given by lagged expectations). In this case, reference points are the same as the mug case derived above. Then, for the indifference

---

[6]Hence, W^*_{mug} would also be a CPE: a CPE must maximize expected utility given that the plan determines the reference-point. But when $q = 0$, the reference point is not affected by plans.

condition to hold at W_{pen}^*, we must have $U\left(\{1 \text{ pen}, \$0|\Omega\}\right) = U\left(\{0 \text{ pens}, \$W_{pen}^*|\Omega\}\right)$. We have

$$U\left(\{1 \text{ pen}, \$0|\Omega\}\right) = u_{pen} + \eta\left[u_{pen} + \left(p + \frac{q}{10}W_{mug}^*\right)(-\lambda u_{mug}) + \left(\frac{q}{10}\right)\int_{W_{mug}^*}^{10}(u_{pen} - \lambda x)\,dx\right]$$

$$U\left(\left\{\begin{array}{c} 0 \text{ pens}, \$W_{pen}^* \\ |\Omega, W_{mug}^* > W_{pen}^* \end{array}\right\}\right) = W_{pen}^* + \eta\left[\begin{array}{c} \left(1 - q + \frac{q}{10}W_{mug}^*\right)W_{pen}^* \\ -\left(p + \frac{q}{10}W_{mug}^*\right)\lambda u_{mug} + \frac{q}{10}\lambda\int_{W_{mug}^*}^{10}\left(W_{pen}^* - x\right)dx \end{array}\right]$$

$$U\left(\left\{\begin{array}{c} 0 \text{ pens}, \$W_{pen}^* \\ |\Omega, W_{mug}^* < W_{pen}^* \end{array}\right\}\right) = W_{pen}^* + \eta\left[\begin{array}{c} \left(1 - q + \frac{q}{10}W_{mug}^*\right)W_{pen}^* \\ -\left(p + \frac{q}{10}W_{mug}^*\right)\lambda u_{mug} \\ +\frac{q}{10}\int_{W_{mug}^*}^{W_{pen}^*}\left(W_{pen}^* - x\right)dx - \frac{q}{10}\lambda\int_{W_{pen}^*}^{10}\left(x - W_{pen}^*\right)dx \end{array}\right]$$

Take $W_{mug}^* > W_{pen}^*$. Then at the indifference condition we have, after simplification:

$$(1 + \eta)\,u_{pen} = W_{pen}^* + \eta\left[\left(1 + (\lambda - 1)q + \frac{q}{10}(1 - \lambda)W_{mug}^*\right)W_{pen}^*\right]$$

which does not depend directly on p. Applying the implicit function theorem gives

$$\frac{q}{10}\eta W_{mug}^* W_{pen}^*(\lambda - 1) = \frac{dW_{pen}^*}{dW_{mug}^*}\cdot\left[1 + \eta\left[\left(1 + (\lambda - 1)q - \frac{q}{10}W_{mug}^*(\lambda - 1)\right)\right]\right]$$

which yields $\frac{dW_{pen}^*}{dW_{mug}^*} > 0$ since $W_{mug}^* < 10$.

Similarly, for the case in which $W_{mug}^* < W_{pen}^*$, simplifying the indifference condition and using the implicit function theorem gives

$$\eta\left[\frac{q}{10}\left((\lambda - 1)\left(W_{mug}^*\right)\right)\right] = \frac{dW_{pen}^*}{dW_{mug}^*}\cdot\left[1 + \eta\left[(1 + (\lambda - 1)q) - \frac{q}{10}\left(W_{pen}^*\right)(\lambda - 1)\right]\right]$$

which again gives $\frac{dW_{pen}^*}{dW_{mug}^*} > 0$ since $W_{mug}^* < 10$.

A.4 Derivation of KR Prediction for Experiment in Section 4.1.1

In the experiment in Section 4.1.1, subjects are put in the following situation: at $t = -1$, they are told that at $t = 1$, they will receive one mug with probability $p_i \in \{p_L, p_H\}$. With probability q, they receive one mug for sure and additionally enter a Becker-DeGroot-Marschak (BDM) mechanism

where for different amounts of money between \$0 and \$10 they have to indicate whether they prefer receiving a *second* mug or the money. With probability $1 - p - q$, they do not get any mug or a choice between a second mug and money.

In the main text, we claim that the KR theory predicts no effect of p on WTA for the second mug. To formally prove this, we need to show that any PE for the BDM stage is independent of p. A subject's plan here corresponds to a monetary amount $\$W^*_{mug2}$ at which the subject is indifferent between receiving that amount and receiving the second mug. Then, when entering the BDM mechanism, the subject (consistent with her plan) chooses the mug for any amount below $\$W^*_{mug2}$ and the money otherwise.

Following a reasoning similar to the one in the proof of Proposition 2, the necessary condition for a plan $\tilde{\Omega}$ to be an interior PE is:

$$U\left(\left\{2 \text{ mugs}, \$0|\tilde{\Omega}\right\}\right) = U\left(\left\{1 \text{ mug}, \$W^*_{mug2}|\tilde{\Omega}\right\}\right).$$

Under plan $\tilde{\Omega}$, the possible outcomes are:

- $\{1 \text{ mug}, \$0\}$ with probability p

- $\{0 \text{ mug}, \$0\}$ with probability $1 - p - q$

- $\{2 \text{ mugs}, \$0\}$ with probability $\frac{q}{10} W^*_{mug2}$

- $\{1 \text{ mug}, \$x\}$ for $x \in \left[W^*_{mug}, 10\right]$ with probability $\frac{q}{10}\left(10 - W^*_{mug2}\right)$

Denote subject i's consumption utility from the mug by u_{mug}, her utility from two mugs (where we assume the second mug to be in the same consumption dimension as the first mug) by u_{2mugs}, and let utility of money be linear with scale normalized to one.[7] The two expressions of interest for the indifference condition are given by:

---

[7]If the second mug were in a different consumption utility dimension from the first mug, the claim that p does not affect the valuation of the second mug trivially holds.

$$U\left(\left\{2 \text{ mugs}, \$0|\tilde{\Omega}\right\}\right) = u_{2mugs} + \eta \left[\begin{array}{l} (1-p-q)\left(u_{2mugs}\right) + p(u_{2mugs} - u_{mug}) \\ + \left(\frac{q}{10}\right) \int_{W^*_{mug2}}^{10} \left((u_{2mugs} - u_{mug}) - \lambda x\right) dx \end{array}\right]$$

$$U\left(\left\{1 \text{ mug}, \$W^*_{mug2}|\tilde{\Omega}\right\}\right) = u_{mug} + W^*_{mug2} + \eta \left[\begin{array}{l} (1-p-q)\left(u_{mug} + W^*_{mug2}\right) + pW^*_{mug2} \\ + \frac{q}{10}W^*_{mug2}\left(W^*_{mug2} - \lambda(u_{2mugs} - u_{mug})\right) \\ - \left(\frac{q}{10}\right)\lambda \int_{W^*_{mug2}}^{10}\left(x - W^*_{mug2}\right) dx \end{array}\right]$$

The indifference condition at W^*_{mug2} is then satisfied if the two expressions are equal. What we need to show is that this indifference condition is independent of p, the probability of getting only one mug. Collecting all the terms that do not contain p in constants κ_1 and κ_2 for the two expressions above, the expressions can be simplified to

$$U\left(\left\{2 \text{ mugs}, \$0|\tilde{\Omega}\right\}\right) = \kappa_1 + \eta\left[(1-p-q)\,u_{2mugs} + p(u_{2mugs} - u_{mug})\right]$$
$$= \kappa_1 + \eta\left[(1-q)\,u_{2mugs} - pu_{mug}\right]$$
$$U\left(\left\{1 \text{ mug}, \$W^*_{mug2}|\tilde{\Omega}\right\}\right) = \kappa_2 + \eta\left[(1-p-q)\left(u_{mug} + W^*_{mug2}\right) + pW^*_{mug2}\right]$$
$$= \kappa_2 + \eta\left[(1-p-q)\,u_{mug} + (1-q)W^*_{mug2}\right]$$
$$= \kappa_2 + \eta\left[(1-q)(W^*_{mug2} + u_{mug}) - pu_{mug}\right]$$

When setting $U\left(\left\{2 \text{ mugs}, \$0|\tilde{\Omega}\right\}\right) = U\left(\left\{1 \text{ mug}, \$W^*_{mug2}|\tilde{\Omega}\right\}\right)$, $-pu_{mug}$ cancels and the indifference condition is independent of p. Thus, W^*_{mug2} is independent of p, as claimed. As in Proposition 2, it is again instructive to inspect the limit case $q = 0$: in this case, the unique PE is $W^*_{mug2} = u_{2mugs} - u_{mug}$, exactly like with classical preferences.

B Empirical Appendix

B.1 Experiment 1

Table A.1 summarizes the questionnaire responses from subjects in our exchange experiment. Questions 1 to 4 were asked after the subjects made their conditional choice as to whether to exchange

or not, but before they knew whether their choice would apply or not. Question 5 was asked as part of the debriefing questionnaire at the end of the study.

Table A.1: Subject Evaluations of the Mugs and Pens in Experiment 1

| | *Treatment* | | |
Question	L	H	*p*-value
1) "I like the mug better than the pen."	3.95 (1.17)	3.26 (1.25)	.061
2) "Since the beginning of the session, I have spent some time thinking about how I would use the pen."	2.95 (1.46)	3.35 (1.23)	.366
3) "Since the beginning of the session, I have spent some time thinking about how I would use the mug."	4.09 (0.87)	4.00 (1.09)	.960
4) "Since the beginning of the session, I have spent more time thinking about the mug than about the pen."	3.95 (1.13)	3.17 (1.40)	.056
5) "Which item do you think has higher retail value?"	3.50 (1.14)	3.26 (1.01)	.387

Notes: Questions 1) to 4) were asked after subjects made their decision as to whether they would like to trade, and are answered on a scale from 1 (strongly disagree) to 5 (strongly agree). Question 5) was asked as part of the debriefing questionnaire, and is answered on the following scale (with the corresponding number in brackets): "Definitely Pen" (1), "Probably Pen" (2), "About the same" (3), "Probably Mug" (4), "Definitely Mug" (5). Standard deviations in parentheses. p-values are from a two-sided nonparametric Wilcoxon-Mann-Whitney test.

B.2   Experiment 2

Table A.2 contains multiple robustness checks of the results discussed in Section 3.3. Regressions (1) to (3) use $\ln(WTA_{mug})$ as the dependent variable, as in the analysis in the main text, while regressions (4) to (9) instead use WTA_{mug} in levels.

In regression (1), we add an interaction term $M_i^H \cdot \ln(WTA_{pen,i})$ to our main regression to allow for the possibility that the relation between $\ln(WTA_{mug})$ and $\ln(WTA_{pen})$ differs across treatments. The coefficient on this interaction term is negative but not significant. Regression (2) uses session dummies instead of date dummies as in the main text, which increases the estimate of the treatment effect somewhat. Column (3) shows the results from an interval regression that explicitly allows for the fact that we only know bounds on subjects' WTA.[8] The treatment effect remains significant at $p < 0.05$, with a nearly unchanged point estimate compared to the analysis

---

[8]The different intervals in which subjects' WTA can lie are: \$0 (assuming nobody would be willing to pay to not have to take an item); [\$0, \$0.33]; [\$0.33, \$0.66]; ... [\$9.24, \$9.57]; [\$9.57, ∞).

in the text.

Regression (4) shows that simply comparing WTA_{mug} across treatments does not yield a significant treatment effect. However, once WTA_{pen} is added as a regressor, the estimated treatment effect increases in magnitude, as in the log case, and is significant at $p < 0.1$ (column (5)). Column (6) adds demographic controls, while column (7) displays the marginal effects from a Tobit regression that takes into account censoring at $WTA_{mug} = 0$ and $WTA_{mug} = 9.57$. Columns (8) and (9) display results from interval regressions, and show that once demographic characteristics are added as regressors, the treatment effect is significant at $p < 0.05$. Overall, we conclude that the results from estimating our main equation (in logs) are robust to a variety of different regression specifications.

Table A.2: Determinants of Willingness-to-Accept for the Mug: Robustness

	Dependent variable: $\ln(WTA_{mug})$			*Dependent variable:* WTA_{mug}					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment M^H	.559** (.228)	.37*** (.134)	.297** (.13)	.596 (.496)	.765* (.426)	.81* (.419)	.777* (.448)	.774* (.425)	.814** (.413)
$\ln(WTA_{pen})$	.616*** (.115)	.438*** (.095)	.531*** (.093)						
Tr. $M^H \times \ln(WTA_{pen})$	-.252 (.179)								
WTA_{pen}					.536*** (.099)	.577*** (.098)	.571*** (.104)	.547*** (.098)	.587*** (.096)
Female						-.759* (.432)			-.796* (.426)
Age						-.16* (.087)			-.156* (.086)
#observations	104	104	104	112	112	112	112	112	112

Notes: All regressions other than (2) (which uses session indicators) also contain indicators for the different days on which sessions were run (none of which ever reach a significance level $p < 0.2$) and an indicator for $WTA_{pen} = 9.57$ (which indicates that a subject's WTA for the pen may be censored; $p > 0.3$ in all regressions). Standard errors in parentheses.

Level of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$