Endowment Effect Theory, Subject Misconceptions and Enhancement Effect Theory: A Reply to Isoni, Loomes and Sugden

Charles R. Plott and Kathryn Zeiler

March 23, 2010

The purpose of Plott and Zeiler (PZ) (2005) was to investigate whether previously published experiments using consumption goods such as mugs and candy bars to measure gaps between willingness to pay (WTP) and willingness to accept (WTA) support endowment effect theory (EET). Our results demonstrate that the gap for commodities can be turned on and off by implementing procedures designed to control for subject misconceptions about the value elicitation procedures.

Following experiments traditionally used to demonstrate the endowment effect, we used mug valuations to test EET. We used lottery rounds only to provide our subjects with paid practice using the value elicitation device prior to employing that device to elicit subjects' mug valuations. In a footnote, we report evidence of contamination in the lottery data that rendered it inappropriate for our purposes. (PZ, 2005, fn. 15) In a data supplement provided to all who request our lottery data, we report gaps in lotteries and provide more detail about footprints of misconceptions we identified.

ILS claim that our 2005 paper is misleading and has mislead researchers. They are concerned that our paper produces, and has been interpreted as producing, a set of procedures sufficient to remove all gaps, including gaps in lotteries. To justify their concern they focus on the wording of our abstract and overlook the underlying themes of our analysis and the context of paragraphs from which they lift quotes to support their thesis. Unfortunately the lifting of quotes from context presents a misleading picture of the unchallenged facts that we report in our paper, namely that mug gaps disappear after we implement controls for misconceptions and that none of our data provides support for EET. In Section 1 we demonstrate that we did not make broad claims about our procedures.

ILS make three additional claims. First, they claim that endowment effect theory "appears to be a loosely-related family of theories…" (ILS, p. 8) From their immediately subsequent discussion, one might infer that testing EET is a futile endeavor. In Section 2 we demonstrate that endowment effect theory is well-specified and makes clear predictions given our experimental environment. We also emphasize that ILS's and our mug data do not support EET. When controls for misconceptions about the elicitation device are implemented in mug experiments, the valuation gap disappears.

Second, as they push our focus to the side, they substitute their own focus: to determine whether our procedures generally eliminate gaps. ILS recognize that our procedures contaminated the lottery data and claim to have adjusted our procedures to control for *all* contamination that concerned us. In Section 3 we demonstrate that, despite their attempt to implement a complete set of controls, ILS's lottery data are contaminated in the ways we described in our paper (PZ 2005, footnote 15) and in our data supplement. A closer look reveals that ILS's lottery data signals obvious footprints of contamination and that they were too quick to summarily dismiss our concerns as a mere "ex post conjecture." (ILS, p. 32) Subject misconceptions are evident in the substantial number of irrational valuations reported by subjects. Furthermore, misconceptions about random devices appear to cause a shifting in subject beliefs about lottery outcomes that is correlated with selling (WTA) and buying (WTP) roles. This implies that the lottery for which WTA is elicited is not the same as the lottery for which WTP is elicited, resulting in a failure to measure a gap arising from the valuation of the same good, the only type of gap of interest for testing EET. For this reason, the use of lotteries to study valuation gaps appears to be fundamentally flawed.

In Section 4 we discuss some of the vast literature, some of which is not typically associated with gaps, that provides clues about how to control for misconceptions related to random devices. This experimental literature suggests that our procedures are ill-equipped to elicit lottery valuations for the purpose of testing EET (or any preference-based theory). We provide the beginnings of an analysis that would be required to generate a set of controls to remove misconceptions related to random devices.

Third, despite ILS's claim that they "are not concerned with testing any particular theoretical account of WTP-WTA gaps," (ILS, p. 10) they take up a number of explanations in their conclusion section. In Section 5 we discuss the two most prominent alternative conjectures ILS offer to explain our mug data. The first—the house money conjecture—cannot be tested properly using ILS's data and is inconsistent with patterns of data found in the literature. The second ILS theory is grounded in the impact of procedures, including emphasis on the role of buying and selling and placement of the good, and focuses on how such procedures produce perceptions of loss. This theory is problematic because the observations do not rule out enhancement effects.

Section 6 offers concluding remarks and makes clear that our analysis of the lottery data is consistent with our observations related to the nature of different goods (PZ, 2005, p. 531).

I. We did not claim that ours procedures will eliminate gaps in all contexts and that gaps were eliminated in lottery rounds

The main claim of PZ (2005) is that the mug experiments that are widely advertized as supporting endowment effect theory do not. When procedures are implemented to control for subject misconceptions about the elicitation device, no WTA-WTP gap is observed, contrary to the prediction of EET. ILS attempt to saddle us with having made additional claims, but the text of our paper and our data supplement, which we sent to ILS upon request early on,[1] both make clear that these additional claims cannot be attributed to us. The supplement clearly stated that we observed gaps in the lottery data. In addition, it, together with the text of our paper, identified footprints of misconceptions, which ILS chose to summarily dismiss even though the footprints provide the key to a deeper understanding of what the lottery data can teach us about valuation elicitation.

ILS's statement of their perceived problem with our study seriously mischaracterizes the nature of our procedures. In their abstract, they assert "…other data from the same study, not published in that paper, exhibit a significant and persistent disparity when the same experimental procedures are applied to lotteries." This statement is flawed in a major (albeit subtle) way. The lottery data, in fact, were not produced using "the same experimental procedures." These data were produced during practice sessions implemented to remove misconceptions about the elicitation device used to elicit mug valuations, which eventually would be used to test EET. We never intended to use the lottery data to test EET. Therefore, we did not apply revealed theory methods to determine the appropriate set of controls for misconceptions in the lottery rounds.

[1] The data supplement includes all lottery data and an explanation of the problems we discovered when we analyzed them.

ILS failed to consider our key procedure: implementation of revealed theory methodology necessary to develop a properly controlled design.

In this Section we clarify three important points in an effort to alleviate ILS's concern that readers will (and have) misinterpreted our conclusions, and, most importantly, to make sure readers understand the implications of PZ (2005), which is consistent with ILS's and our results. First, readers of the full text of PZ cannot reasonably conclude that we claimed the procedures are sufficient to control for misconceptions related to lotteries. Second, readers cannot reasonably conclude that we looked for but found no gaps in the lottery data. Third, the studies referenced by ILS as examples of misinterpretations seem not to misinterpret our claims, contrary to ILS's suggestions.

a. PZ (2005) do not claim that the procedures remove all gaps

ILS claim that PZ (2005) "has been widely cited as providing experimental support for the hypothesis that the PZ elicitation procedure eliminates WTP–WTA disparities *in general.*" This claim rests on two sentences in our abstract, which might mislead the reader if read in isolation. The sentences in the PZ abstract read, "Experiments were conducted using both lotteries and mugs, goods frequently used in endowment effect experiments. Using the modified procedures, we observe no gap between WTA and WTP." The abstract should have read, "Experiments were conducted using lotteries as training tools. Endowment effect theory was tested using mugs, goods traditionally used for the development and testing of endowment effect theory."

When PZ is taken as a whole, it is clear we did not claim that the procedures are sufficient to control for misconceptions in all contexts. In fact, we went to great lengths to caution against such a broad interpretation of our results. In several other parts of the paper, we explicitly state that we make no claims about having discovered a set of procedures that remove all sources of misconceptions (e.g., "...we have neither a general theory of what might constitute misconceptions nor a set of operational definitions characterizing them. Constructing a full set of procedures to control for them could be very difficult." (p. 543); "…the concept of misconceptions has not been operationalized formally and certainly not quantified. In fact, its meaning changes from one experimental environment to another, and from experimental study to experimental study. Consequently, a theory of misconceptions has not been developed." (p. 531)). We included these statements in the paper specifically to ensure that readers would not interpret claims made about the procedures too broadly. We also make it clear in footnote 15 that our procedures were insufficient to control for contamination in the lottery rounds. As a general matter, we encourage readers to develop impressions based on our full text.[2]

b. PZ (2005) did not claim that gaps were eliminated in lottery rounds

While ILS assert that readers of PZ (2005) believe that we looked for and found no gaps in the lottery data, a closer look demonstrates that readers who read parts of the paper not quoted by ILS along with ILS's selective quotes are very unlikely to draw such a conclusion. While ILS focus on the text in our Abstract and literature review to claim that readers will interpret us as claiming that we eliminated gaps in the lottery rounds (ILS, p. 5), in other parts of the paper that focus on the results themselves, we make it clear that we did not include the lottery rounds in our analysis. Our footnote 15 openly states that the lottery data collected during training rounds were not used (and should not be used) to test EET. We described features of the data suggesting that they should not be used to test EET and would be a challenge to *any preference-based theory*.

---

[2] As a warning to ILS readers, we note here that ILS tend to quote from context, omitting key qualifying phrases. The parts left out substantially change the meaning of the included parts. For example, ILS (p. 3) include only part of the following quote:

> "The 'primary conclusion' they derive from the data they report is that 'observed WTP-WTA gaps do not reflect a fundamental feature of human preferences. *That is, endowment effect theory does not seem to explain observed gaps.*" ([PZ,] p. 542).

ILS omit the last italicized sentence thereby substantially changing the meaning of our text. This sentence, along with many other statements in our paper, make it clear that we properly focus on experiments using commodities that allow for clean tests of EET. The full quote clearly indicates that we made no claim about all WTP-WTA gaps in all circumstances. Sections 3 and 4 of this reply provide evidence that lotteries are not useful for testing EET.

ILS (p. 3) continue to support their thesis by quoting out of context from text appearing several paragraphs later:

> "[T]he thesis of their paper is that, to the contrary, 'observed gaps are symptomatic of subjects' misconceptions about the nature of the experimental task' in which valuations are elicited (p. 542)."

Again when taken from the context, the quoted sentence suggests that we made claims we simply did not make. In fact, in the paragraph following the sentence ILS lifted from context, we write: "our thesis is not especially satisfying because we have neither a general theory of what might constitute misconceptions nor a set of operational definitions characterizing them" and that "we have no direct evidence that our procedures actually eliminate all misconceptions about how revealed valuations map into payoffs." (PZ, p. 542-3) In fact, while misconceptions are a thesis of our project, our claims are not confined to a single interpretation of the data. Several alternative theories reveal themselves, and we conclude by stating that "[w]e do not take a stand on which of these interpretations is valid or answerable by our literature review and experimental results. In fact, we disagree on this point." (PZ, p. 544) We do agree, however, that our mug data do not support EET and that claims made in the literature that gaps are unrelated to experimental procedures are misleading. (PZ, p. 544)

c. Revisiting references to PZ (2005) by other researchers

ILS (p. 6) cite several papers to support the claim that our "paper has been widely cited as providing experimental support for the hypothesis that the PZ elicitation procedure eliminates WTP–WTA disparities *in general* and that such disparities are artifacts of insufficiently controlled experiments." Nothing in the quotes ILS report (including the one from PZ (2007)) imply that PZ's procedures—specifically designed to control misconceptions about the valuation elicitation device in mug rounds—would control all misconceptions, including those that arise from subjects' difficulties in valuing lotteries, goods they are not accustomed to valuing. In addition, none of the cited papers explicitly claims that the procedures eliminate gaps in general or that our procedures eliminated gaps in lottery rounds. None of the quotes mention lottery rounds and some explicitly mention mug rounds. The quotes simply summarize our claim that observed gaps cannot be used to support EET and the data suggest misconceptions are key.

A quick glance at our paper by any reader of the discussions ILS cite would make clear the lack of generalizability of our claim (e.g., "[the] meaning [of misconceptions] changes from one experiment environment to another and from experimental study to experimental study." (PZ, 2005, p. 531) If readers of PZ (2005) were fuzzy on this point due to language ILS point to in our abstract and literature review, our footnote 15 unambiguously clears up the fuzz by describing evidence that subjects had difficulties valuing lotteries in the absence of training on lotteries. In addition, nothing in ILS's quotes from others (p. 6) suggest that PZ tested for gaps in the lottery data and found no gaps. This is likely because footnote 15 also makes clear that we did not include the lottery data in our analysis.

II. ILS's mug data reject endowment effect theory

Before we address ILS's main concern, it is important to remind the reader about our study's major claims and purposes and to emphasize that ILS's mug round data unambiguously reject EET. While ILS claim not to be concerned with explaining observed gaps[3] (or at least with

[3] "[W]e are not concerned with testing *any* particular theoretical account of WTP–WTA gaps." (ILS, p. 10)

testing EET[4]), the purpose of our study clearly is to test EET.[5] We find that gaps observed in Kahneman, Knetsch and Thaler's (KKT) (1990) mug experiments are not due to a kink in the utility function at the endowment as posited by EET. When the experiment is controlled for subject misconceptions about the elicitation device the valuation gap disappears. ILS conduct experiments with mugs using our procedures. Their mug data confirm our basic conclusion.

Our study was designed specifically to focus on particular experiments developed to demonstrate EET at work—the case of mug experiments, where all alternative explanations seem to be ruled out. Data from both PZ and ILS demonstrate that experiments advanced in the literature as support for EET, i.e. experiments asking subjects to state valuations for goods such as candy and mugs, fail to produce gaps when controls designed to address misconceptions about the elicitation device are implemented. Thus, ILS contribute to the literature by replicating (using modified procedures) the key PZ result, adding to the data from other experiments that unambiguously reject EET as it was originally developed and applied.[6]

ILS skirt around a discussion of the relationship between their data and EET by claiming that "PZ do not set out EET as a specific formal theory." (ILS, p. 7) They argue "[a]s used by PZ, 'EET' appears to refer to a loosely-related family of theories of reference-dependent preferences which has evolved and diverged over time…" (ILS, p. 7), concluding that it is impossible to determine whether data support or reject the theory. They justify their avoidance of the issue by arguing that "[s]ince EET is not a sharply-defined concept, engagement with those issues would be an unhelpful distraction from the point of our paper." (ILS, p. 7)

This characterization of EET and our study is inaccurate. First, our paper refers to two of many specific descriptions of the purported link between valuation gaps and prospect theory. (PZ, p. 531) Second, contrary to ILS's claim, EET is a precisely stated theory with unambiguous predictions that apply generally. KKT (2008) provide a clear formulation of EET. Figure 1 is a reproduction of the figure KKT present as a formal model of EET as Tversky and Kahneman

[4] "[W]e are *not* concerned with PZ's interpretation of what they call EET. Nor are we concerned with whether PZ's or our own data are consistent with EET." (ILS, p. 9-10) Although later in the paper, ILS design experiments to test a house money effect explanation and other explanations for gaps in mug rounds. We take this up in Section 5.
[5] We clearly state in our title, abstract and throughout the paper that our experiment is designed specifically to test EET.
[6] E.g., Kovalchik et al. (2005) conducted similar experiments using healthy elderly individuals (average age 82) and found no WTP-WTA gap for mugs when misconceptions for the elicitation device are controlled. These results support the conclusion that EET does not explain observed gaps.

(1991) formally developed it.[7] The theory apparently has remained unchanged since then. Individual preferences in the buying role are modeled in the northwest quadrant, where a move from the origin involves gaining the good and giving up money. When the same individual is in the selling role, we use the southeast quadrant to predict behavior. A move from the origin as seller involves the giving up of the good and a gain in the money dimension. The kink in the curve at the origin illustrates loss aversion: losses hurt more than gains of the same size help. To round out the possibilities, the theory also models marginal rates of substitution in gains and losses when individuals are offered choices. EET's prediction in our context (valuations for mugs) is straightforward: EET is supported only if WTA > WTP.

While ILS claim that EET generates no unambiguous predictions,[8] the developers and appliers of the theory harbor no such hesitations. For example, Knetsch et al. (2001, p. 257) conclude, "The endowment effect and loss aversion have been among the most robust findings of the psychology of decision making. People commonly value losses much more than commensurate gains independent of transactions costs, income effects or wealth constraints." KKT (1990, p. 1345) also express confidence in the reliability with which the theory can be applied, holding that its consequences can be observed in property rights acquired by historic accident or fortuitous circumstances, government licenses, landing rights, transferable pollution permits, divisions, plants, product lines, etc. Those applying EET in legal contexts have made similar claims.[9]

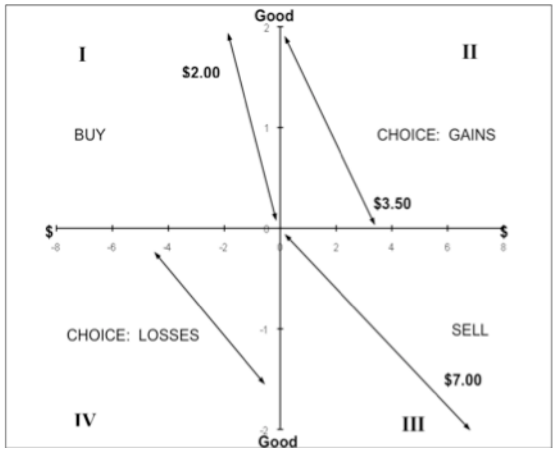
---

[7] The Figure is derivable directly from the model presented in Tversky and Kahneman (1991). Note that EET is a theory of preferences as opposed to a theory of a decision process as developed by Gigerenzer and Todd (2008) or a hybrid of preference and process as postulated by the discovered preference hypothesis (Plott, 1996). When describing the experiments with mugs, KKT are very clear about this feature of the theory: "To conclude, the evidence reported here offers no support for the contention that observations of loss aversion and the consequential evaluation disparities are artifacts; nor should they be interpreted as mistakes likely to be eliminated by experience, training, or 'market discipline.' Instead, the findings support an alternative view of endowment effects and loss aversion as fundamental characteristics of preferences." (KKT, 1990, p. 1346)

[8] KKT (1990) do mention some contexts in which EET might not predict a gap, e.g., goods held for resale and induced-value tokens, but none of these contexts are relevant for our mug experiment.

[9] "Researchers in behavioral decision theory have developed a growing line of evidence that people appear to value a good that they own much more than an identical good that they do not own…. Researchers have used several different procedures to demonstrate the endowment effect." (Rachlinski and Jourden, 1998) Rachlinski and Jourden go on to argue that endowment effect theory explains why litigants expend resources to appeal judicial decisions that strip them of property rights even in cases in which the court awards damages to compensate for the loss. Similarly, Huang (2004) suggests that valuation disparities impact settlement behavior of litigants. Sanchirico (2004) argues that WTA-WTP gaps might preclude parties to a nuisance suit from bargaining to an efficient outcome, although he questions the meaning of an efficient outcome when preferences depend on entitlements. Reuben (2004) suggests the right to trial by jury might push disputants to favor litigation over arbitration because litigants wish to avoid the disutility that accompanies giving up one's right to trial. Bell and Parchomovsky (2002) claim that endowment

Figure 1: EET as described by KKT (2008)



To summarize, contrary to ILS's claim, KKT (1990) offered a formal version of EET and suggested that it applies under many complex and varying circumstances. It has been applied widely in economics and other fields, and gaps observed in the laboratory are used as evidence of the reliability of the theory as it is applied in the field. It is important to note that the applications assume that a kink exists at the endowment as opposed to other contextual reference points. Our experiment design was meant specifically to test whether the laboratory evidence conventionally used to support EET (i.e., mug experiments) would hold up when misconceptions about the elicitation device were controlled. Our data, ILS's data and others' data clearly reject EET when put to that test.

III. ILS's experiment failed to control lottery data contamination

The purpose of our lottery rounds was to provide our subjects with paid practice to help them understand the properties of the Becker–DeGroot–Marschak (BDM) (1964) mechanism and the meaning of and consequences of expressing bids and asks in the BDM context. During the practice rounds, we elicited valuations for four degenerate lotteries in which the outcome was a

---

effect theory strengthens an adverse possessor's claim to disputed land. Law reviews house hundreds of similar applications of endowment effect theory to law.

small amount of money for certain,[10] in addition to two small stakes, non-degenerate lotteries and eight larger stakes, non-degenerate lotteries, the outcomes of which were determined randomly. Our review of the data from these training rounds implies that they were contaminated by an experimental design inappropriate for the study of lottery gaps.

ILS repeatedly assert that they have developed a set of experimental procedures that removed “all contamination” from lottery choices that we identified in our paper. Indeed, ILS define as their major purpose to determine whether a WTP-WTA gap remains in the lottery rounds after controlling for all sources of contamination.[11] They describe their experiment as an “uncontaminated replication of PZ’s experiment.” (ILS, p. 10) As it turns out, their experiment is neither a replication of any PZ experiment nor does it control for contamination we identified and discussed. Furthermore, as will be discussed in this section, the ILS experiments contain footprints of misconceptions that signal the types of contamination we identified. ILS rejected our analysis of misconceptions as ad hoc theorizing, but they did not check their data for the footprints we identified. These footprints of misconceptions challenge the foundation of ILS’s thesis.

Relative to our concept of what constitutes lottery data contamination, ILS declare a narrower definition of “contamination.” ILS use the term to describe only the order of the lottery rounds and the personal communications between subjects and the experimenter during the practice rounds.[12] Our paper makes clear that the lottery data were collected during training rounds, and

[10] Given ILS’s inferences from a paragraph taken from our instructions (“This passage seems to be advising subjects to use the small stakes lottery tasks as practices for the large stakes lottery tasks and possibly (in the case of Treatments 1 and 3) also for the mug task…” (ILS, p. 15), we worry that readers will make the unsupported assumption that the four degenerate and two other small stakes lottery rounds are sufficient for training. First, ILS’s interpretation is incorrect. We did not intend to communicate anything to the subjects about the purpose of various lottery rounds relative to other lottery rounds. Rather, our goal was simply to motivate them to pay attention during the small stakes rounds. Second, because we did not intend to use the lottery rounds to measure gaps, we want to make clear that we did not make any claims about how many rounds are required to properly train subjects on how to value lotteries when facing the BDM mechanism. As we discuss in the following section, training about probabilities and randomness is more involved than training for the elicitation device alone. For example, Grether, et al. (2007) report evidence suggesting that eight rounds of experience using a second price auction elicitation device is insufficient to remove misconceptions harbored by many subjects.

[11] “In mounting our study, our primary objective was to apply PZ’s elicitation procedure to both mugs and lotteries while ensuring that none of the paid tasks was contaminated.” (ILS, p. 7)

[12] “All WTA rounds were conducted prior to the WTP rounds,” and “Mistake corrections, public answers to questions and other procedures were also employed continuously….” (PZ, 2005, fn. 15) In other words, they interpret our footnote 15 to mean that if these two sources of contamination are properly controlled (or otherwise explained away), the lottery data will be useful for determining whether gaps are linked to misconceptions. It is important to note that ILS provide no measure or test of whether contamination was successfully removed even

our definition of contamination is broader than the one ILS adopt. We designed the lottery rounds only to provide subjects with paid practice with the elicitation device that would subsequently be used to elicit mug valuations to test EET.[13] Our footnote 15 begins with a warning—"The lottery round data…are contaminated by a design that was developed only for training and not for purposes of measuring a gap"—and then continues with a list of both sources and hints of misconceptions related to the elicitation device, the nature of randomization and the concept of probability. Our concept of "contamination" naturally includes misconceptions, a fundamental component of our analysis as demonstrated by the title of our paper and our use of the concept throughout. Despite this, ILS incorrectly assume it to encompass only the ordering of the rounds and the personal interactions between experimenter and subjects.

We elaborated on the concepts introduced in footnote 15 in a detailed data supplement that we supplied to ILS upon request. The supplement reports gaps in the lottery data in addition to footprints of misconceptions about random devices that we identified by analyzing the lottery data in light of subject debriefs that provided hints about where to look. In line with the thesis of our paper, we warned that the identified lottery data contamination rendered the data inappropriate for testing EET. More specifically, we reported evidence of irrational valuations, asymmetric valuations on the boundaries of the lottery value support, and inexplicable risk preference patterns. ILS dismiss our concerns about these footprints as post hoc theorizing. Upon systematic investigation, however, we find the same footprints of misconceptions in the ILS lottery data. Closer study reveals patterns in the data that challenge ILS's enterprise of using lottery valuations to measure WTP-WTA gaps. The footprints of misconceptions we identify might explain why so few have used lotteries to test EET.[14]

---

though it is possible to examine the lottery data for misconceptions. Their paper rests entirely on the unwarranted assumption that the procedures they added were successful.

[13] Given the role of the training rounds, lotteries appealed to us. Using lotteries for training was more efficient than consumer goods because we did not have to have a large number of different commodities on hand for each training decision of each session.

[14] In addition to the data issues we describe below, according to the theoretical literature at the time, EET applies only if the reference point is some sure amount, as ILS note (ILS, fn. 5). EET makes no prediction when the reference point is set by a lottery. Although KKT (1990, p. 1343) briefly mention "risky prospects" in their discussion section, Tversky and Kahneman (1991), which KKT cite as a foundational model of loss aversion, state that their theory applies to riskless choice. After we produced our results, others developed theories that are alternatives to EET and might be applied to lottery environments (Sugden (2003); Koszegi and Rabin (2006)). Our experiments, however, were not designed to test these alternative theories, which were not available at the time we were writing.

While we presented no general theory of misconceptions, the existence of footprints of misconceptions about the random device implies that misconceptions cannot be ruled out as an explanation of patterns observed in the lottery data. Since subjects were not trained on the nature of randomness and the associated independence of outcomes across rounds, a broad concept of misconceptions is a worthy candidate for further study. Importantly, ILS fail to investigate whether the data provide evidence of misconceptions; they simply assume that their modified procedures controlled for all contamination. In the remainder of this Section, we provide an analysis of the lottery data that reveals clear footnotes of misconceptions similar to the ones we reported in our data supplement.

a. Irrational Choices

The natural places to look for evidence of misconceptions are those in which the indicators are uncontroversial and clear. For this reason, we first examine subject valuations of the four lotteries with certain payoffs. Under the BDM procedure, the dominant strategy is to value degenerate lotteries at the certain value. Presumably, any irrational choices expressed in these lotteries reflect misconceptions, misunderstanding or confusion on the part of the subjects.

A close look at ILS's lottery data reveals substantial irrationality, both in general and relative to our lottery data. As ILS explain, their subjects' and fixed offers were restricted to multiples of £ .05. Thus, the only rational valuations for a lottery with a certain payoff of X are X and X + .05 for sellers (WTA) and X and X - .05 for buyers (WTP). Table 1 reports the proportion of irrational valuations reported by ILS's subjects and our subjects for the four lotteries with certain outcomes. For example, in Lottery #1, 28% of ILS's subjects failed to state a rational value for a lottery that paid 20 pence with certainty. On average, in each degenerate lottery round roughly 23% of ILS's subjects exhibited irrationality in this seemingly simple task. The irrationality displayed by ILS's subjects is approximately seven times higher than that displayed by our subjects; roughly 3.4% of our subjects' valuations, on average per round, were irrational. ILS assume that their modifications of our procedures removed all contamination, but their data clearly demonstrate that they did not.

Table 1: FREQUENCY OF IRRATIONAL VALUATIONS FOR CERTAIN LOTTERIES				
	Lottery #1	Lottery #2	Lottery #4	Lottery #5
Certain payoff	20 cents/pence	30 cents/pence	30 cents/pence	40 cents/pence
ILS	28%	24%	19%	22%
PZ	3%	3%	3%	5%

The difference between ILS's and our data is striking, and we can only speculate about the reasons. As we explained in our paper, we detected what appeared to us to be a substantial lack of understanding of BDM during the early training rounds and used hands-on methods to correct it. ILS, following our footnote, viewed the hands on procedures as a possible source of contamination from an experiment design perspective for purposes of measuring within-subject gaps, and employed computerized instructions that involved different methods. The levels of irrational behavior exhibited in ILS's data, however, suggest that either our training procedures did at least some work to reduce some misconceptions or the procedures they used generated additional misconceptions.

Irrational valuations of non-degenerate lotteries are similarly signaled by valuations that fall outside the support of a lottery value's probability distribution (when payoffs are positive). If a lottery has a zero probability of paying more than x and a zero probability of paying less than y, then valuations above x or below y violate rationality postulates. The ILS non-degenerate lottery data exhibit very little irrationality according to this definition, and thus stand in sharp contrast to their degenerate lottery data discussed above. Indeed, the change in behavior between degenerate and non-degenerate lotteries is dramatic. On average, in the non-degenerate lottery rounds, 3.6% of ILS's subjects reported irrational valuations, as defined in the subsection above,[15] whereas 22% exhibited irrational behavior in lottery #5, the last degenerate lottery round. While further investigation is required to determine the factors driving this phenomenon, the sudden change in behavior—substantial irrationality in a simple task to almost no irrationality in a more complex task—is another footprint of misconceptions. In the following section, we provide a more detailed examination of the nature of the misconceptions in both sets of data.

[15] Note that while we can attribute a portion of the difference to the fact that four out of the ten non-degenerate lotteries have a lower bound of zero, no more than 5% of the subjects bid above the highest rational value in each round. Therefore, it is not the inability to express one's low valuation that is driving the decrease in irrationality.

b. Boundary Valuations

All ILS and PZ lotteries take the same form. Each results in a payout of x with probability P(x) and a payout of y with probability 1- P(x). The preference elicitation procedures are designed to elicit the subject's lottery value, V. Under such conditions, standard models of risky choice, including the model ILS adopt (p. 13), hold that V = x if and only if P(x) =1, and V = y if and only if P(x) = 0. This property is true regardless of the level of risk aversion. Thus a valuation at the boundary of a lottery payoff support implies an extreme belief over the outcome probabilities. While this model assumes that possible lottery outcomes are continuous and that utility functions are smooth, if the mathematical properties are viewed only as approximations, boundary valuations imply beliefs that are inconsistent with the objective probabilities of the lottery value supports. In other words, the existence of boundary valuations suggests that subjects have misconceptions about randomness.

To investigate footprints of misconceptions in the non-degenerate large stakes lottery rounds, we focus on "boundary valuations." We define boundary valuations as valuations that lie exactly on the bounds of the support or irrationally outside the bounds. For example, ILS lottery #13 (a buying task) paid £ 3.5 with probability 0.5 and £ 1.5 with probability 0.5; thus, all valuations at £ 3.5 or above and all valuations at £ 1.5 or below are counted as boundary valuations. Even though valuations exactly on the bounds are not strictly irrational as defined in the previous subsection, these valuations are sufficiently extreme to challenge any reasonable assumptions related to belief formation. Table 2 displays the frequencies of boundary valuations for each large-stakes, non-degenerate lottery. In both experiments, subjects valued large stakes lotteries as sellers (WTA) in rounds 7-10 and as buyers (WTP) in rounds 11-14. In lottery #14, for example, 15% of ILS's subjects offered to pay no more than £1, and 3% offered to pay at least £5, for a lottery that paid £5 with a 30% chance and £1 with a 70% chance.

Several features of the data are instructive. First, ILS subjects more frequently reported boundary valuations. On average each round, 13.4% of ILS's subjects reported boundary valuations; 8.2% of our subjects, on average across rounds, reported such valuations (counting only lotteries with non-negative outcomes—lotteries #9 and #13 had a negative lower bound in our experiment).

Table 2: FREQUENCY OF BOUNDARY VALUATIONS FOR UNCERTAIN LARGE-STAKES LOTTERIES									
		WTA BOUNDARY VALUATIONS				WTP BOUNDARY VALUATIONS			
		#7	# 8	#9	#10	#11	#12	#13	#14
ILS (N=100)	lottery value support (£)	[0,3]	[0,2]	[0.5,2.5]	[0,4]	[1,4]	[1,3]	[1.5,3.5]	[1,5]
	upper	14%**	5%	5%	5%	5%	2%	5%	3%
	lower	0%	0%	7%	0%	8%***	18%***	15%**	15%***
PZ (N=74)	lottery value support ($)	[0,7]	[0,5]	[-4,8]	[0,10]	[1,8]	[1,6]	[-3,9]	[1,11]
	upper	8%**	7%**	0%	8%	1%	1%	0%	7%
	lower	1%	0%	4%	1%	7%**	7%**	1%	1%

Note: "Upper" ("lower") indicates the proportion of subjects who valued the non-degenerate lottery at or above (below) the upper (lower) bound of the lottery value support. The asterisks denote results of tests of equal proportions (* denotes significance at the 10% level, ** at the 5% level and *** at the 1% level). For upper boundary valuations, Ho: proportion of valuations at or above upper bound in WTA = proportion of valuations at or above upper bound in WTP; Ha: proportion at or above upper bound in WTA rounds > proportion at or above upper bound in WTP. For example, 14% at or above the upper bound in ILS's round 7 is compared to 5% at or above the upper bound in ILS's round 11. Similar tests are performed on lower boundary valuations (Ho: proportion of valuations at or below lower bound in WTA = proportion of valuations at or below lower bound in WTP; Ha: proportion at or below lower bound in WTA rounds < proportion at or below lower bound in WTP). For lotteries with negative or zero lower bounds, the frequency represents offers of $0.

While the ILS boundary valuations are more frequent than ours, both are suspiciously large and suggest loss of control over beliefs of outcomes. Given the experimental environment, these subjects do not exhibit sophisticated beliefs over lottery outcomes, which represent probability distributions over the state space. Specifically, they are acting as if they can accurately guess the lottery outcome. We reported this phenomenon in our data supplement. We also find it in the data reported by ILS, which demonstrates that their experiment did not remove contamination as we defined the term.

c. Boundary Valuation Asymmetries

A closer look at Table 2 reveals that subject beliefs appear to be directly influenced by whether they are asked to state their WTA or WTP. That is, the boundary valuations reveal a tendency for subjects to believe that the high payoff will occur with certainty when they are asked for a selling price (WTA) and that the low payoff will occur with certainty when they are asked how much they would pay (WTP). We reported our discovery of this phenomenon in our data supplement, and a quick search through the literature reveals that others have reported closely related phenomena elsewhere (e.g., Risen and Gilovich, 2007).

We define "boundary valuation asymmetry" as the asymmetric tendency for subjects to be more likely to report boundary valuations at or above the upper bound in selling (WTA) rounds and at or below the lower bound in buying (WTP) rounds. For example, 14% of ILS's subjects reported boundary valuations at or above the upper bound when valuing as sellers in round #7 and no subject chose the lower bound of zero. By comparison, when valuing the matched lottery as buyers in round #11, 5% reported boundary valuations at or above the upper bound and 8% chose the lower bound or below. The results of one-tailed tests of equal proportions are displayed in Table 2. Five of the eight tests using ILS's data resulted in a statistically significant difference (at the 5% level); four of the eight tests using our data produced similar results.

In short, when subjects report valuations as sellers, subjects are more likely to believe that the lottery outcome will be the higher payoff. Conversely, when subjects report valuations as buyers, they are more likely to believe that the lottery outcome will be the lower payoff.

Boundary asymmetries naturally imply the existence of "boundary gaps." An upper boundary gap is measured using only pairs of individual values for which the valuation as seller is on or above the upper bound and the valuation as buyer is below the upper bound. Similarly, to measure lower boundary gaps, we use only pairs for which the valuation as buyer is on or below the lower bound and the valuation as seller is above the lower bound. Table 3 displays gaps, as measured by ILS, using a variety of subsets of the data. The first row of each panel ("all ILS data" and "all PZ data") reports gaps using all the data. These correspond to the gaps reported by ILS.[16] The second row in each panel reports upper boundary gaps. The results demonstrate that the magnitude of gaps in the full dataset is driven substantially by relatively few individual ratios with seller valuations that are based on extreme beliefs. Likewise, the third row in each panel displays lower boundary gaps and reveals a similar impact by relatively few individual ratios with buyer valuations that are based on an extreme belief. The fourth row further demonstrates the impact of extreme asymmetric beliefs on lottery gaps. When those individuals exhibiting extreme beliefs of any kind as buyer or seller are removed from the data, the magnitudes of the gaps drop substantially.

Importantly, boundary gaps demonstrate evidence of misconceptions and do not imply the utility function assumed by endowment effect theory. An individual who values a lottery on the boundary of the value support should reveal the same value regardless of whether in the role of buyer or seller. This implies that boundary gaps are different from the standard interpretation of a valuation gap: a gap between WTA and WTA for the *same* good.[17] Individuals revealing

[16] ILS reported gaps separately for our Type A and Type B lotteries. We pooled these subsets to construct Table 3.
[17] Very few ILS subjects valued the lotteries at or above the upper bound as both buyer and seller (2 for L3/6 and L7/11; 1 for L8/12, L9/13 and L10/14) and at or below the lower bound as both buyer and seller (0 for L3/6, L7/11,

boundary gaps view the lottery from the buying perspective as a different lottery than the one considered from the selling perspective, even though the lotteries are identical (save the addition of some amount to both outcomes). In other words, boundary gaps imply that the actual properties of the lottery as perceived by the individual have changed from round to round and from selling perspective to buying perspective, presumably due to misconceptions about randomness. Hence, the very concept of a valuation gap is inappropriate.

TABLE 3: BOUNDARY GAPS

	L3/L6	L7/L11	L8/L12	L9/L13	L10/L14
ILS LOTTERY DATA					
MEAN WTA/WTP RATIO (MEDIAN RATIO WITH RESULTS FROM ONE-TAILED SIGNED RANK TEST[b])					
All ILS data	2.19 (1.33***) N=100	1.53 (1.26***) N=100	1.37 (1.16***) N=100	1.11 (1.]00) N=100	1.46 (1.11***) N=100
Upper boundary gap	10.58 (2.67**) N=7	2.42 (2.03***) N=12	2.76 (3.00**) N=4	2.20 (2.33**) N=4	4.38 (5.00**) N=4
Lower boundary gap	6.43 (2.50***) N=15	4.01 (4.00***) N=8	2.27 (2.24***) N=18	1.71 (1.67***) N=15	3.06 (2.50***) N=15
Both valuations inside the bounds	1.36 (1.15**) N=78	1.26 (1.16***) N=81	1.16 (1.04***) N=80	1.00 (1.00) N=83	1.16 (1.03) N=83
PZ LOTTERY DATA (POOLED ACROSS TYPE A AND B LOTTERIES)					
	L3/L6	L7/L11	L8/L12	L9/L13	L10/L14
MEAN WTA/WTP RATIO (MEDIAN RATIO WITH RESULTS FROM ONE-TAILED SIGNED RANK TEST)					
All PZ data	1.53 (1.15***) N=71[a]	1.56 (1.20***) N=74	1.50 (1.20***) N=74	1.68 (1.03***) N=73	1.32 (1.04***) N=74
Upper boundary gap	2.16 (1.60**) N=5	2.15 (2.00**) N=5	1.77 (1.44**) N=4	N=0	1.83 (1.47**) N=4
Lower boundary gap	N=0	4.40 (5.00**) N=5	4.00 (4.05**) N=5	N=0	8.00 (8.00) N=1
Both valuations inside the bounds	1.49 (1.10***) N=65	1.31 (1.15***) N=63	1.29 (1.17***) N=64	1.68 (1.03***) N=73	1.20 (1.00**) N=67

[a] Ratios for three subjects are undefined because lottery valued at $0 as buyer in Round 6. In Round 13, one potential buyer valued the lottery at $0. The sign test includes the $0 bids.

[b] H_o: adjusted WTA and WTP come from identical distributions (results from one-tailed test)

The existence of the boundary gaps and the substantial reduction in gap magnitude when only non-suspect data are included suggest that misconceptions about the random device play a central role in the lottery valuation gaps. Furthermore, the existence of boundary asymmetries

L8/12 and L10/14, and 2 for L9/13). The same is true for our subjects (upper bound: 1 for L3/6, L7/11 and L8/12; 0 for L9/13 and 2 for L10/14; 0 for all lotteries for the lower bound).

has substantial implications for the use of lotteries in the measurement of WTP-WTA gaps. If misconceptions about the random device cannot be controlled, and in particular, if the systematic violation of sophisticated beliefs cannot be avoided, then, as a matter of principle, lotteries valuations should not be used to measure WTA-WTP gaps, which assume that individuals value the same good as buyer and as seller. If the roles of seller and buyer trigger a change in subjective probabilities for the same lottery, then the lottery valued from the buying perspective is not the same as the lottery valued from the selling perspective. Unless subjects view the good as the same good from both perspectives, testing EET or even measuring a valuation gap using lotteries is impossible. Tversky and Kahneman (1991) were wise to move the science to the study of commodities when developing the foundations from which EET was constructed.

Finally, while gaps remain after removing suspect valuations, conclusions drawn from this observation would be unreliable. It is reasonable to believe that the forces that compel the reporting of suspect valuations might influence the valuations of subjects who stayed inside the bounds. More work is required to determine whether gaps would remain after properly controlling for misconceptions about random devices. In Section 4, we discuss how our revealed theory methodology illuminates avenues through which the research can proceed.

d. Preference Consistency and Risk Preferences

The final footprint of misconceptions we exam relates to risk preferences. Classical prospect theory holds that individual preferences are possibly risk averse in the gains and possibly risk seeking in the losses. (Kahneman and Tversky, 1979) The individual, however, is postulated to have a consistent (complete and transitive) preference given an endowment (or reference point). That is, the preference will change as the endowment changes, but, given the endowment, the preferences are of a classical type. Testing a theory such as EET with data that violate these fundamental assumptions makes little sense. Here we demonstrate that the lottery data fall into this category.

While modern generalizations of prospect theory might provide a theory of shifting preferences and shifting risk attitudes as expectations and experience evolves, such instability would not be welcome in a test of EET as it is traditionally developed and applied.[18] Both ILS's and our subjects failed to exhibit stable risk preferences across rounds. For example, only 38% of ILS

[18] See Section 2 for a summary of a formal model of EET.

subjects consistently displayed risk aversion in all of the large stakes buying rounds. Similarly, only 23% of our subjects consistently bid below the expected value of the lottery in every large stakes buying round.[19] While such proportions suggest that a large number of subjects had risk seeking preferences, very few subjects consistently bid the expected value or bid above the expected value in all of the large stakes buying rounds.[20] Thus, 53% of ILS's subjects and 64% of our subjects failed to exhibit consistent risk preferences in buying rounds. Such instability, the flopping back and forth between risk aversion and risk seeking behavior in WTP rounds, is a signal of loss of experimental control and suggests that the data are not appropriate for theory testing. The same instability is observed in selling rounds, with 70% of ILS's subjects and 78% of our subjects exhibiting inconsistent risk preference throughout those rounds.[21] These features of the data add to the list of reasons why ILS's and our lottery data are not useful for the purpose of our 2005 study. If subjects' perceptions of the lotteries fluctuate when the experimenter (and the theory) assume that the lotteries have not changed, support for a specific shape of any underlying preference becomes impossible.

In conclusion, it is important to remember that EET rests on a theory of preferences characterized by a specific kink at the reference point set by the endowment. Thus, if EET is tested using any data for which the reference point is known or assumed to be known, and if the data fail to exhibit consistent preferences, the theory is rejected. The lottery data might tell a useful story if applied to modern extensions of prospect theory, especially those assuming the reference point is set by something other than the endowment. Pointing to the lottery data as either support for or rejection of EET, however, would not be credible given that major features of the data disconfirm the basic assumptions of EET.

---

[19] Risk aversion is commonly reported in experiments, so a lack of evidence of risk aversion (especially in buying rounds as some versions of EET predict risk seeking behavior in selling rounds) signals a possible lack of experimental control. Of 400 large stakes WTP valuations in ILS's study, 91 (23%) were *above* expected value. The results are starker if we include the small stakes lotteries. In PZ's study, 61 (23%) of 296 large stakes WTP valuations were above expected value. This calls control into question in both experiments.

[20] Three percent (8%) of ILS (PZ) large stakes lottery buyers bid the expected value each round. Six percent (5%) consistently bid above expected value in these buying rounds.

[21] Eleven percent (5%) of ILS (PZ) large stakes lottery owners offered below expected value each round. Thirteen percent (12%) consistently offered above expected value in these rounds. Six percent (5%) consistently offered the expected value in these rounds.

IV. Revealed theory methodology and controlling for lottery misconceptions

Our study was based on the intuition that subject misconceptions about the preference elicitation device systematically impact valuations. Importantly, we recognized that there is no theory of misconceptions. To approach this seemingly intractable problem, we developed a "revealed theory" approach that rests on the assumption that procedures used by experimentalists to control for misconceptions in some sense reveal a theory of misconceptions. Our strategy was to cast a wide net by examining theories and employing procedures that others had suggested or used in relation to misconceptions and the preference elicitation device. Using this method, we incorporated them all into a single set of procedures with the hope that misconceptions about the elicitation device might be adequately controlled. As we did not intend to use the lottery data to test EET, we did not develop procedures to control for misconceptions about random devices associated with lotteries.

When we employed the revealed theory procedures in mug rounds, the valuation gap disappeared. Thus, our mug data demonstrate that the mug experiments do not support EET. Given that there is no well-defined theory of misconceptions that can be specified as a theoretical alternative to EET, however, the data fall short of supporting a conclusion that the gap in mug valuations is caused by some specific theory of misconceptions. Instead, application of the revealed theory method produces a collection of potential theories (PZ, p. 531). This implies that no clear theory of misconceptions can be formulated to apply in all circumstances and that the set of controls will depend on the features of the experiment that might lead to misconceptions, including features of the goods. For these reasons, to correctly apply revealed theory methodology, one must appeal to the literatures that specifically address features that might trigger misconceptions.

Neither PZ nor ILS apply revealed theory methodology in the lottery round design. Neither design considers the literature that explores possible misconceptions associated with special features of lotteries. It is well known that lotteries can trigger difficulties in addition to any misconception that might relate to the elicitation device and that these must be addressed before reported valuations can be used to test any theory of preference. Experimenters have studied important issues related to the random process and its role in influencing subjects' beliefs (i.e., subjective probabilities) about what actions might be optimal. Depending on the factors that influence beliefs as subjects face lotteries, values placed on the same lottery can change from round to round, as we demonstrated in the previous section. The experimenter, being unable to observe the subject's changing beliefs, unknowingly collects data that are contaminated for the purpose of testing any theory assuming consistent preferences. This problem is especially troublesome when the theory rests on a particular shape of preference, as does prospect theory.

To properly apply revealed theory methodology to remove misconceptions about random devices in order to test theories of preferences over lotteries, the literature warns that one must attend to numerous types of misconceptions related to random devices. The challenge seems to stem from difficulties some individuals have with the notion of randomness itself. For example, some of our subjects reported believing that they could accurately guess the lottery outcomes, a belief supported by ILS's and our lottery data. If a subject believes he can accurately predict the outcome of a particular lottery and the prediction depends on whether he owns the lottery, then the subject's (unobserved) subjective probabilities, and thus the subject's valuation of the same lottery, might change from round to round. Such changing beliefs can be motivated by imagined patterns in historical data, theories based on the physical properties of random devices, or even impulses that cause people to believe that some outcomes are more likely than others. Such beliefs are symptomatic of a disconnect between the probabilities as understood by the experimenter and the actual beliefs held by subjects.

The nature of this disconnect requires study and, in fact, has been studied in a variety of ways. One of the most dramatic examples of the disconnect is the number of offers to pay more for a lottery than the maximum it could pay (or asking for less than the minimum payoff), as observed by both ILS and us. Keren and Willemsen (2008) connect this mistake to misconceptions over random devices. Camerer (1989) and Bossaerts and Plott (2004) report findings suggesting that individuals tend not to accept the notion of independence. How to train subjects away from these tendencies is not obvious, but the literature suggests ideas including the use of formal training models (Offerman et al. 2009) and training methods that work through subject involvement in the lottery procedures (Fong and McCabe, 1999).

While on the one hand paradoxes (e.g., Allais (1953) and Ellsberg (1961)) and anomalies (e.g., overconfidence, beliefs relating to low probability events) abound, several findings suggest that attitude formation has an underlying structure that leads to the eventual understanding of or an effective way of dealing with random processes (Gaissmaier and Schooler, 2008; Gigerenzer et al., 2008). Camilleri and Newell (2009) give subjects an opportunity to value hundreds of lotteries.[22] Roth (2005) has directly studied whether observed lottery gaps are related to the underlying riskiness of lotteries. The application of revealed theory methodology to lottery rounds should take into account these and other such studies in fashioning controls for misconceptions about randomness.

V. Explanations for gaps: House money effect, enhancement effect theory and endowment effect theory

Despite ILS's claim that they are "not concerned with testing *any* particular theoretical account of WTP–WTA gaps" (ILS, p. 10), they include expansive discussions of possible explanations. Specifically, they offer two conjectures to explain the mug results reported both by us and by them. Both conjectures are problematic. We discuss them in turn and note that their second theory is simply a version of a more general theory we call enhancement effect theory, which holds that the process of acquisition and experience of ownership can create special features of goods that have independent value that can be confused with a kinked utility function.

a. The House Money Effect

The first conjecture is that the elimination of the gap in mug rounds is due to a house money effect triggered not by cash from lottery winnings but by the show up fee paid in cash at the beginning of the experiment. While this alternative explanation is worth considering, we note that it does not work to explain the broader set of results found in the literature. It also provides additional evidence to reject EET.[23]

While ILS's results are suggestive, one should be hesitant to accept the house money effect explanation on the basis of this limited evidence. First, as ILS admit, they do not formally test the conjecture. While it could be tested in many ways, the most straightforward would be to

[22] By comparison, ILS's and our lottery data are collected after only two unpaid practice rounds.

[23] Of course, others might argue that, rather than rejecting EET, the evidence simply suggests that multiple effects exist and, in some cases, one effect will dominate the others. If this were the case, however, EET would need to be modified to take these other effects into account. This adaptation of the theory, however, would seriously compromise the effect's robustness and would call into question the myriad applications of EET. See Zeiler (2010) for numerous examples of applications in law.

replicate KKT's gap result and then add a show up fee to the KKT procedures. The disappearance of the gap would serve as evidence of the house money effect conjecture. ILS, however, do not perform this test because they have not produced a baseline using KKT's original procedures, which do not include a show up fee but result in a gap. They claim that they performed a "controlled comparison between the PZ and KKT elicitation procedures," (ILS, p. 27) but they have not. Since ILS do not perform this experiment, the lack of a gap in their modified KKT experiment possibly is a result of a different subject pool or subtle differences between KKT's original procedures and ILS's modification of them. That is, it could be the case that KKT's results simply are not robust to changes in procedures, which would imply that EET does not explain gaps.

Second, their explanation does not work well to explain the pattern of results reported in the literature. Experiments conducted by Kovalchik et al. (2005) employ the PZ procedures and do not observe a gap in mug valuations, yet they do not employ a show up fee. This result is contrary to the predictions of the ILS conjecture that PZ's elimination of the gap is the result of a house money effect. In addition, the experiments performed by Smith (2008) employ KKT's procedures but add a show up fee. Contrary to the prediction of ILS's theory, Smith reports a valuation gap. Thus the broader literature meets the ILS house money effect explanation of our mug gap with skepticism.

b. Enhancement Effect Theory

The second theory ILS advance to explain gaps observed in the lab is based on a blurring of EET with what we call "enhancement effect theory." Enhancements are generated through the process of endowment or the experience of ownership. Enhancements create value that is added to value that exists in the absence of the enhancement.[24] For example, if the experimenter announces, "the mug is a gift" when endowing subjects, the nature of the good in the eyes of non-owners and owners might differ. (Plott and Zeiler, 2007) Potential buyers are deciding about a mug, while potential sellers are deciding about a mug that was a gift from someone who will observe

[24] In a separate study, we explore environments in which features of the experiment design generate enhancements that might add value to endowed goods. (Plott and Zeiler, 2007) Specifically, we tested what we now call enhancement effect theory against EET and found support for enhancement effect theory. While we focused on features that enhance the value of endowments in an effort to explain exchange asymmetries, we recognize the possibility that features of the environment might also *reduce* the value of the endowed good, possibly leading to buyer valuations that exceed seller valuations.

whether the "gift" is traded for something else. Sellers who benefit from signaling appreciation to the experimenter might be more likely to ask for more than they would if they were considering only the mug's consumption value.

Examples of valuation gaps taken from the field demonstrate the potential for blurring enhancements and endowments. In the field, ownership is often associated with enhancements that supplement, modify and shape the value of a good for potential sellers. For example, a potential home seller values his home from the perspective of someone who has lived in it for some time. The independent value from enhancements that shape the seller's overall value could emerge from many unobservable sources or could manifest in observable ways, e.g. through the good memories that living there create. Economic theory posits a number of sources of enhancements that might produce gaps, such as value associated with a trophy or the first dollar one earns, or values associated with information asymmetries (e.g., a risk-averse owner might place a higher value on the car she has been driving relative to an identical car that was previously owned by someone else simply because she is more certain about how her car impacts her utility).

Without the ability to control for enhancement effect theory as an alternative explanation, we are unable to determine whether a kink in the utility function at the reference point is causing gaps observed in complex field environments, or whether ownership (for more than just a few minutes) has somehow changed the nature of the good. It is hard to imagine an example taken from the field and used as support for EET that could not also be explained by unobserved enhancement effects. Of course, this calls into question the falsifiability of enhancement effect theory when applied in the field. A similar problem, however, plagues field applications of EET. Essentially any gap in the field also can be attributed to some unobserved reference point combined with loss aversion.

The problems of distinguishing EET from enhancement effect theory naturally spill into the laboratory. For example, ILS criticize our instructions as biasing the results against EET by using the word "offer" as opposed to words like "buy" and "sell." They assert that such language "reduc[es] the salience of the distinction between buying and selling tasks." (ILS, p. 27) They argue that, while our subjects are told that they own the good, "there is little else to flag up the difference between buying and selling, whereas other experiments draw more attention to this difference." (ILS, p. 28) Enhancement effect theory predicts, however, that subjects might perceive the experimenter's emphasis of roles as information about the value of the goods (PZ, 2007). Emphasis on "buyer" and "seller" roles can also trigger strategic instincts that cause sellers to offer high amounts and buyers to offer low amounts (PZ, 2005, p. 537-38). Therefore, if we emphasize buying and selling and we observe a gap, we can't be sure whether the result supports EET or enhancement effect theory (or reversion to basic instincts).[25] Similarly, ILS argue that placement of the good might impact the setting of the reference point, and that elimination of the gap in our experiment might be due to the fact that both sellers and buyers had mugs in front of them. (ILS, p. 28) Again, if we make such a change and observe a gap, enhancement effect theory cannot be ruled out because it suggests that placement of the goods might signal relative value. (PZ, 2007) Given that the theories are observationally equivalent in general settings, separation of them requires highly controlled laboratory settings.

VI. Concluding remarks

PZ (2005) draws a specific conclusion: EET does not explain the WTP-WTA gap observed in the classic mug experiments that were designed specifically to give the theory its best shot. ILS replicate this result, providing support for our central conclusion.

That EET does not explain gaps does not tell us what does explain them. Our analyses (PZ 2005 and PZ 2007) offer two conjectures, both of which are related to experiment procedures. The first posits that gaps are due to misconceptions related to the elicitation mechanism. The second advances enhancement effect theory as an explanation. Enhancement effect theory holds that gaps arise from enhancements predicted by economic theory and generated by the method of endowment and, perhaps, by the elicitation procedures. While it was not their purpose, ILS's results add to the evidence we present here that lottery gaps are related to misconceptions about randomness and beliefs about probabilities produced by preference elicitation procedures. Thus, the ILS experiments bring into focus a dimension of misconceptions that we posited and that, upon further exploration, have been found to raise questions about the appropriateness of

[25] We should add that the forms subjects used to record valuations mentioned buying or selling more than once immediately prior to the elicitation of the offer. For example, the forms used to elicit valuations from buyers read: "**Buyer** Record Sheet: The experimenter owns one Round X lottery ticket. I will offer to **buy** the lottery ticket for an amount equal to the maximum I am willing to pay for the ticket." (emphasis added)

interpreting observed lottery gaps as revealing preferences for the same item in the selling role and in the buying role.

PZ (2005) makes clear that we were curious about experimental procedures and associated misconceptions. Our observation related to the nature of goods was (and is) that others have focused on the properties of the endowed good, e.g. the difference between mugs and money or tokens for money, as an explanation for gaps, suggesting that the disparity in gap results that we highlighted in Table 1 in PZ (2005) is caused by the lack of comparability of different properties (Tversky and Kahneman (1991)). The conclusion of PZ (2005) was that the mug experiments do not support theories based on the properties of goods formulated to explain the disparity of results across experiments. The claims we make here about the different nature of lotteries are unrelated to our previous claim. Here we argue that properties of goods, such as the randomness associated with lotteries, might carry an inherent possibility of misconceptions, which might lead to gaps. Indeed, our analysis of ILS's and our lottery data supports the view that misconceptions indeed play a role in producing lottery gaps and that different goods might call for different controls for misconceptions. Just as we argued in PZ (2005), however, a full understanding and appropriate test of our conjecture will require an application of revealed theory methodology, which neither ILS nor we applied.

References

Allais, Maurice. 1953. "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine." *Econometrica* 21:503-46.

Becker, Gordon M.; DeGroot, Morris H. and Marschak, Jacob. 1964. "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science*, 9(3):226–32.

Bell, Abraham and Parchomovsky, Gideon. 2002. "Pliability Rules." *Michigan Law Review*, 101:1-xx.

Bossaerts, Peter and Plott, Charles R. 2004. "Basic Principles of Asset Pricing Theory: Evidence from Large-Scale Experimental Financial Markets." *Review of Finance*, 8:135-69.

Camilleri, Adrian R. and Newell, Ben R. 2009. "The role of representation in experience-based choice." *Judgment and Decision Making,* 4(7) 518-29.

Chou, Eileen; McConnell, Margaret; Nagel, Rosemarie and Plott, Charles R. 2009. "The Control of Game Form Recognition in Experiments: Understanding Dominant Strategy Failures in a Simple, Two Person 'Guessing' Game." *Experimental Economics*, 12(2): 159-79.

Camerer, Colin. 1989. "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review*, 79:1257-61.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75(4):643–69.

Fong, Christina and McCabe, Kevin. 1999. "Are decisions under risk malleable?" *Proceedings of the National Academy of Science*, 96:10927-32.

Gaissmaier, Wolfgang and Schooler, Lael J. 2008. "The Smart Potential Behind Probability Matching." *Cognition*, 109:416-22.

Gigerenzer, Gerd; Hertwig, Ralph; Hofferage, Ulrich and Sedlmeier, Peter. 2008. "Cognitive Illusions Reconsidered." In *Handbook of Experimental Economic Results*, ed. Charles R. Plott and Vernon L. Smith, 1018-33. North Holland.

Gigerenzer, Gerd and Todd, Peter M. 2008. "Rationality the Fast and Frugal Way: Introduction." In *Handbook of Experimental Economic Results*, ed. Charles R. Plott and Vernon L. Smith, 977-85, North Holland.

Grether, David M.; Plott, Charles R.; Rowe, Daniel B.; Sereno, Martin and Allman, John M. 2007. "Mental Processes and Strategic Equilibration: An fMRI Study of Selling Strategy in Second Price Auctions." *Experimental Economics*, 10(2):105-22.

Huang, Peter H. 2004. "Lawsuit Abandonment Options in Possibly Frivolous Litigation Games." *The Review of Litigation*, 23:47-xx.

Isoni, Andrea; Loomes, Graham and Sugden, Robert. 201x. "The Willingness to Pay-Willingness to Accept Gaps, the 'Endowment Effect,' Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: A Reassessment," *American Economic Review*, vol(issue):##-##.

Kahneman, Daniel; Knetsch, Jack L. and Thaler, Richard H. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98(6):1325-48.

Kahneman, Daniel; Knetsch, Jack L. and Thaler, Richard. 2008. "The Endowment Effect: Evidence of Losses Valued More Than Gains." In *Handbook of Experimental Economic Results*, ed. Charles R. Plott and Vernon L. Smith, 939-48, North Holland.

Kahneman, Daniel and Tversky, Amos. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2):263-92.

Knetsch, Jack L.; Tang, Fang-Fang and Thaler, Richard H. 2001. "The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing?" *Experimental Economics*, 4:257-69.

Keren, Gideon and Willemsen, Martijn C. 2008. "Decision Anomalies, Experimenter Assumptions, and Participants' Comprehension: Revaluating the Uncertainty Effect." *Journal of Behavioral Decision Making*, 22(3):301-17.

Koszegi, Botond and Rabin, Matthew. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4):1133-65.

Kovalchik, Stephanie; Camerer, Colin F.; Grether, David M.; Plott, Charles R. and Allman, John M. 2005. "Aging and Decision Making: A Comparison Between Neurologically Healthy Elderly and Young Individuals." *Journal of Economic Behavior and Organization*, 58:79-94.

Offerman, Theo; Sonnemans, Joep; Van De Kuilen, Gijs and Walker, Peter P. 2009. "A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes." *Review of Economic Studies*, 76(4):1461-89.

Plott, Charles R. 1996. "Rational Individual Behavior in Markets and Social Choices Processes: The Discovered Preference Hypothesis." In *The Rational Foundations of Economic Behavior*, eds. Kenneth J. Arrow, Enrico Colombatto, Mark Perlman and Christian Schmidt, IEA Conference Vol. 114: 225-50. London: Macmillan.

Plott, Charles R. and Zeiler, Kathryn. 2005. "The Willingness to Pay/Willingness to Accept Gap, the 'Endowment Effect,' Subject Misconceptions and Experimental Procedures for Eliciting Valuations." *American Economic Review*, 95:530-45.

Plott, Charles R. and Zeiler, Kathryn. 2007. "Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory." *American Economic Review*, 97(4):1449-66.

Rachlinski, Jeffrey J. and Jourden, Forest. 1998. "Remedies and the Psychology of Ownership." *Vanderbilt Law Review*, 51:1541-82.

Reuben, Richard C. 2004. "Democracy and Dispute Resolution: The Problem of Arbitration." *Law and Contemporary Problems*, 67:279-xx.

Risen, Jane L. and Gilovich, Thomas. 2007. "Another Look at Why People Are Reluctant to Exchange Lottery Tickets." *Journal of Personality and Social Psychology*, 93(1):12-22.

Roth, Gerrit. 2005. "Predicting the Gap Between Willingness to Accept and Willingness to Pay." PhD diss. Munich Graduate School of Economics.

Sanchirico, Chris W. 2004. "Evidence, Procedure, and the Upside of Cognitive Error." *Stanford Law Review*, 57:291.

Smith, Alex. 2008. "Lagged Beliefs and Reference Dependent Utility." PhD diss. University of Arizona Department of Economics.

Sugden, Robert. 2003. "Reference-dependent subjective expected utility." *Journal of Economic Theory*, 111:172-91.

Tversky, Amos and Kahneman, Daniel. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, 106(4):1039-61.

Zeiler, Kathryn. 2010. "Willingness to Pay-Willingness to Accept Gaps: The Legal Implications of Recent Empirical Developments." Unpublished.