The Willingness to Pay–Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations

By CHARLES R. PLOTT AND KATHRYN ZEILER*

We conduct experiments to explore the possibility that subject misconceptions, as opposed to a particular theory of preferences referred to as the "endowment effect," account for reported gaps between willingness to pay ("WTP") and willingness to accept ("WTA"). The literature reveals two important facts. First, there is no consensus regarding the nature or robustness of WTP-WTA gaps. Second, while experimenters are careful to control for subject misconceptions, there is no consensus about the fundamental properties of misconceptions or how to avoid them. Instead, by implementing different types of experimental controls, experimenters have revealed notions of how misconceptions arise. Experimenters have applied these controls separately or in different combinations. Such controls include ensuring subject anonymity, using incentive-compatible elicitation mechanisms, and providing subjects with practice and training on the elicitation mechanism before employing it to measure valuations. The pattern of results reported in the literature suggests that the widely differing reports of WTP-WTA gaps could be due to an incomplete science regarding subject misconceptions. We implement a "revealed theory" methodology to compensate for the lack of a theory of misconceptions. Theories implicit in experimental procedures found in the literature are at the heart of our experimental design. Thus, our approach to addressing subject misconceptions reflects an attempt to control simultaneously for all dimensions of concern over possible subject misconceptions found in the literature. To this end, our procedures modify the Becker-DeGroot-Marschak mechanism used in previous studies to elicit values. In addition, our procedures supplement commonly used procedures by providing extensive training on the elicitation mechanism before subjects provide WTP and WTA responses. 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. Therefore, our results call into question the interpretation of observed gaps as evidence of loss aversion or prospect theory. Further evidence is required before convincing interpretations of observed gaps can be advanced.

A subtle controversy exists in the literature. At issue is the existence and interpretation of a possible gap between willingness to pay ("WTP") and willingness to accept ("WTA").[1]

* Plott: Department of the Humanities and Social Sciences, California Institute of Technology, M/C 228-77, Pasadena, CA 91125 (e-mail: cplott@hss.caltech.edu); Zeiler: Law Center, Georgetown University, 600 New Jersey Ave., NW, Washington, DC 20001 (e-mail: kmz3@law.georgetown.edu). The financial support of the Laboratory of Experimental Economics and Political Science is gratefully acknowledged. We thank Colin Camerer, Daniel Klerman, Robert Sherman, Eric Talley, Richard Thaler, and Leeat Yariv for helpful discussions and comments. In addition, we are grateful for comments provided during presentations of earlier versions of this work at the University of Southern California Law School and the California Institute of Technology's Experimental Economics Workshop. All errors are ours.

[1] The WTP-WTA gap refers to a tendency for an individual to state a minimum amount for which that individual is willing to "sell" an item that is greater than the maximum amount the same individual is willing to pay to "buy" the item. Under conditions of sufficiently smooth preferences and the absence of wealth effects, the two magnitudes should (theoretically) be the same. Some attribute observed gaps to loss aversion: the notion that the loss of the item due to a sale is more pronounced than a gain of the same item due to a purchase. For this reason, WTP-WTA gaps have

Such a gap is frequently reported in the literature, and many broad claims have been made regarding the robustness of the gap and its implications. For example, summarizing experimental findings about the pervasiveness of the gap in a recent survey of the WTP-WTA gap literature, John K. Horowitz et al. (2000) state, "Previous authors have shown that WTA is usually substantially larger than WTP, and almost all have remarked that the WTP/WTA ratio is much higher than their economic intuition would predict." Furthermore, claims about the power of a particular theory to explain the gap are appearing with increasing frequency. Specifically, the interpretation of the gap as an "endowment effect" rests on a special theory of the psychology of preferences associated with "prospect theory." In particular, Jack L. Knetsch et al. (2001) conclude, "The endowment effect and loss aversion has been one of the most robust findings of the psychology of decision making—people commonly value losses much more than commensurate gains." Such claims regarding the nature and robustness of the gap have seeped into other areas of research, including law and economics, and specific interpretations of the WTP-WTA gap accompany the claims. Jeffrey J. Rachlinski and Forest Jourden (1998) begin their discussion of the implications of the WTP-WTA gap for legal doctrine by claiming, "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." The research reported here suggests that this broad discussion found in the literature is based on an incorrect interpretation of experimental results.

In spite of the enthusiastic interpretations of the WTP-WTA gap as a fundamental feature of human preferences (e.g., referring to the gap as an "endowment effect"), in fact no consensus exists on whether the literature, considered in its entirety, supports such interpretations. While many experimenters have observed a WTP-WTA gap, others have failed to observe it. This variation in experimental results seriously undermines the claim that the gap is a fundamental feature of human preferences. Recognizing this, scholars who accept the psychological explanation of the gap have sought to explain the variation in terms of the commodity used in experiments (e.g., mugs, lotteries, money, candy, etc.). Some suggest that the existence and magnitude of the endowment effect depend on the commodity employed in the experiment. This observation led us to conclude that further examination of the nature of observed gaps is necessary before any convincing interpretation is possible.

Although our review of the experimental literature revealed no consensus about the appropriate interpretation of observed WTP-WTA gaps, we did note an important consensus about the experimental procedures used to measure gaps. Implicitly, researchers unanimously agree that experimental procedures should be designed to minimize or avoid "subject misconceptions." Like its close cousin "confusion," however, 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. Nevertheless, controlling for subject misconceptions is necessary to determine whether they play an important role in the lack of consensus about the nature of the gap.

Our approach in the face of this difficulty is to employ a "revealed theory methodology" to infer an operational meaning of subject misconceptions revealed by the myriad procedures experimentalists adopt to control for them. The procedures implicitly reflect different ideas about the form(s) that subject misconceptions might take. Our approach is to assume that, unless all controls are exercised simultaneously, one cannot conclude that subject misconceptions have been eliminated.

This observation leads to our main research question: If we design an experiment that completely controls for subject misconceptions as implicitly defined by the literature (i.e., an

come to be called "endowment effects." We refer to this explanation of gaps as "endowment effect theory" to denote that the terminology is not simply a label for a particular phenomenon, but rather implies a theoretical explanation of the observed phenomenon.

experiment that includes every procedure used in previous experiments to control for misconceptions), will we observe a WTP-WTA gap?

This question led us toward a natural experiment design. The first step is to conduct experiments using procedures that frequently produce gaps. If we replicate the gap we can conclude that there is nothing special about the subject pools, or about us as experimenters, that might eliminate the gap. The second step is to conduct experiments in which subject misconceptions are completely controlled by incorporating the union of procedures found in the literature. If a gap is observed under this treatment, the results would strongly support interpreting observed gaps as support for endowment effect theory. On the other hand, if a gap is not observed, then the results would support the conjecture that the procedures themselves produce gaps and that gaps are unrelated to the nature of preferences, loss aversion, and prospect theory. If the gap can be turned off and on using different sets of procedures, then it likely does not reflect an asymmetry between gains and losses as posited by loss aversion.

The paper is organized as follows. Section I reviews much of the WTP-WTA gap literature and is designed to document two facts. First, no consensus exists about the nature and robustness of the gap. Second, experimenters have employed a wide range of different and sometimes overlapping sets of procedures to control for subject misconceptions. Such differences in procedures could account for variations in the results if the specific procedures employed control for some facets of misconception but fail to control for others. More importantly, given that no theory of "misconceptions" exists, if subject misconceptions influence the existence or magnitude of the gap, then the optimal method of investigating the influence of misconceptions is to implement the union of controls. Without a theory about the relationship between gaps and subject misconceptions, our objective is to cast a large net using a revealed theory methodology in the hopes of determining whether the procedures themselves cause gaps between WTA and WTP.

Section II reports the results from our replication of an experiment reported by Daniel Kahneman et al. (1990) ("KKT"). KKT's results are cited widely as support for endowment effect theory. Using their procedures, we replicate the gap with roughly the same magnitude they report.

Section III describes in detail the experimental procedures we employ to study whether subject misconceptions account for observed gaps. The procedures reflect the conjecture that observed gaps are related to subjects' misconceptions about the valuation task. We expound on exactly how and why we developed and employed specific procedures. These procedures represent the study's central contribution. Section III also reports the striking result we obtain when we incorporate controls for subject misconceptions. When an incentive-compatible mechanism is used to elicit valuations, and subjects are provided with (a) a detailed explanation of the mechanism and how to arrive at valuations; (b) paid practice using the mechanism; and (c) anonymity, we observe no WTP-WTA gap. To investigate one conjecture regarding which procedures have the greatest impact, we designed a second experiment that is identical to the first, except subjects are not provided paid practice using the mechanism. Again, we observe no statistically significant WTA-WTP gap. Section IV offers concluding remarks.

I. Experimental Procedures and the Literature

We begin our exploration of experimental procedures by examining the literature, relying on the so-called "revealed theory methodology." As will become evident, there is no single view of how to control for subject misconceptions. Table 1 provides a categorization of WTP-WTA gap experiments by experiment procedures. As endowment effect theory[2] is our primary focus, we restrict our attention to the reported studies that investigate the possible existence or nature of an endowment effect, as opposed to mere WTP-WTA gaps. Several of these studies focus on the possible influences of certain procedures (e.g., experience with the elicitation mechanism, practice using the mech-

[2] To be clear, we use "endowment effect theory" to refer to the theory that observed gaps can be explained by some feature of human preferences that leads owners to resist selling goods because (a) selling is perceived as "losing" the endowed good, and (b) individuals are generally loss averse.

TABLE 1—SUMMARY OF THE LITERATURE BY EXPERIMENTAL DESIGN

	Result reported	Optimal responses explained	Practice rounds performed	Valuations elicited using incentive compatible mechanism	Valuations elicited using market environment with some incentives	Gap measured directly using valuations
Knetsch and Sinden (1984; test 1)	gap					
Knetsch and Sinden (1984; test 3)	gap					
Corsey et al. (1987; part 1)	gap					
Corsey et al. (1987; part 2)	gap					
Brookshire and Corsey (1987; exp 1)	gap					
Singh (1991; test 1 before learning)	gap					
Dubourg et al. (1994)	gap					
Brookshire and Corsey (1987; exp 2)	gap					
Knetsch (1989; test 1)	gap				binary choice	
Bateman et al. (2001)	gap				binary choice	
Shogren et al. (1994; stage 1, round 1)	gap				Vickrey	
Boyce et al. (1992)	gap		hypothetical		BDM	
Knetsch (1989; test 2)	gap		hypothetical		BDM	
Morrison (1997; part 2)	gap		random/pooled		BDM	
Shogren et al. (1994; stage 2)	gap		random/pooled		Vickrey	
Bateman (et al. (1997; exp 2)	gap		random/pooled		BDM	
Bateman et al. (1997; exp 1)	gap		random/pooled		BDM	
Knetsch and Sinden (1984; test 2)	gap					
Brookshire and Corsey (1987; exp 3)	gap				Smith auction	
Kahneman et al. (1990; exp 6 & 7)	gap	incorrectly suggested			sealed bid	
Franciosi et al. (1996; exp 1)	gap	incorrectly suggested			sealed bid	
Kahneman et al. (1990; exp 1 & 2)	gap	incorrectly suggested	random		sealed bid	
Kahneman et al. (1990; exp 4)	gap	incorrectly suggested	random/pooled		sealed bid	
Kahneman et al. (1990; exp 3)	gap	incorrectly suggested	pooled		sealed bid	
Loewenstein and Issacharoff (1994; exp 1)	gap	suggested			BDM	
Kahneman et al. (1990; exp 5)	gap	suggested	hypothetical		BDM	
Shogren et al. (2001; BDM)	gap	suggested	random/pooled		BDM	
Knetsch and Sinden (1984; test 4)	no gap					
Singh (1991; test 2 before learning)	no gap					
Singh (1991; tests 1 and 2 after learning)	no gap				DA	
Shogren et al. (1994; no available substitute)	no gap		random/pooled		Vickrey	
Corsey et al. (1987; part 3)	no gap				Vickrey	
Morrison (1997; part 1)	no gap		random/pooled		BDM	
Shogren et al. (1994; stage 1, rounds 2–5)	no gap		random/pooled		Vickrey	
Shogren et al. (1994; available substitute)	no gap		random/pooled		Vickrey	
Arien et al. (2002)	no gap		random/pooled		binary choice	
Shogren et al. (2001; Vickrey)	no gap		random/pooled		Vickrey	
Loewenstein and Issacharoff (1984; exp 2)	no gap	suggested			BDM	
Harless (1989)	no gap		pooled		Vickrey	

Notes: Optimal response explained: If blank, no explanation was provided. "Incorrectly suggested" indicates that the experimenter used a non-incentive-compatible elicitation mechanism but told subjects that revealing true valuations was the optimal strategy. "Suggested" indicates that the experimenter correctly advised subjects that the optimal strategy called for truthful valuation revelation. If shaded, the experimenter provided a detailed explanation of the optimal response.

Practice rounds performed: If blank, then no practice rounds were performed. "Hypothetical" indicates that practice rounds were not paid. "Random" indicates that randomly selected rounds were paid. "Pooled" indicates that the measurement of the gap includes valuations measured in the first round (without experience) and valuations measured in later rounds (after experience).

Valuations elicited using incentive-compatible mechanism: If blank, then non-incentive-compatible mechanism used to elicit valuations. If shaded, then incentive-compatible mechanism used to elicit valuations.

Valuations elicited using market environment with some incentives: If blank, elicitation was not conducted in a market environment. If shaded, then elicitation was conducted in a market environment with some incentives. The type of market environment is indicated for each experiment.

Gap measured directly using valuations: If blank, then gap measured using number of trades relative to predicted number of trades. If shaded, then gap measured using mean or median of actual WTP and WTA responses.

anism, etc.). In addition, our analysis includes mainly experiments that provide some sort of incentive for truthful revelation of valuations. Studies using purely hypothetical methodologies were not included unless they focused specifically on the issues examined in this paper.[3]

Each row displayed in Table 1 represents a particular experiment.[4] The first column of the table indicates the study in which the experiment is reported. The second column indicates whether a statistically significant gap was reported as resulting from the particular experiment. Each remaining column of the table specifies a particular experimental design feature or gap measurement technique. There are five such columns: explanation of optimal responses, provision of practice rounds, incentive compatibility of the elicitation mechanism, the mechanism used to elicit valuations, and the method of gap measurement. We explain these in turn.

Explanation of Optimal Response.—A shaded cell in this column indicates that the experimenter provided subjects with some explanation of the optimal response corresponding to the elicitation mechanism. Explanation entails describing to the subjects the substantive features or purposes of the mechanism and the potential benefits of employing a particular strategy. It is important to note that the level of explanation varies substantially across experiments. For example, in Kahneman et al. (1990) and in our replication of this experiment, subjects simply were provided with the suggestion (sometimes incorrectly) that "[i]t is in [their] best interest to answer ... questions truthfully." This type of explanation is indicated as "suggested." In David S. Brookshire and Don L. Coursey (1987), the experimenters explained in some detail the elicitation mechanism and used numerical examples to illustrate the mechanics of the elicitation device. The experimenters, however, did not advise subjects on optimal responses.

Practice Rounds Provided.—In experiments in which practice rounds were provided, subjects gained experience with the elicitation mechanism while tutoring was available and were encouraged to ask questions. Blank cells indicate that no practice rounds were provided. In some cases, practice rounds were provided but not paid. These cases are indicated by cells marked as "hypothetical." If the practice rounds were paid, then the cells are shaded. In experiments providing paid practice, subjects made decisions and experienced the actual consequences of those decisions (i.e., gained or lost money or goods) under conditions similar to those researchers employed to elicit WTA and WTP responses used to measure the gap.

"Random" indicates that the study paid only a small subset of the subjects or a small subset of the rounds. For example, Kahneman et al. (1990) randomly selected a certain number of the subjects to be paid at the end of the experiment. In other treatments, they randomly selected one of many rounds to be binding at the end of the experiment.[5]

"Pooled" indicates studies that measured the gap using data aggregated across rounds. In such studies there is no clear distinction between practice rounds and gap measurement rounds. For instance, David W. Harless (1989) measures the gap by aggregating data from 12 rounds before which subjects had no paid practice rounds. Before the first round, subjects had no experience using the mechanism, but gained experience as they proceeded through the 12 rounds.

Valuations Elicited Using Incentive Compatible Mechanism.—When experimenters elicit valuations using incentive-compatible mecha-

[3] Note that the list of experiments using hypothetical elicitation methods reported in Table 1 is not all-inclusive. We limited the list to include a sufficient number of hypothetical experiments to demonstrate that gaps are almost always observed when subjects are not provided any incentive to reveal their valuations truthfully. Including more hypothetical experiments in the list would not teach us more than what we learn from the patterns revealed in Table 1.

[4] An on-line Appendix provides some additional information about the specific experiments cited: the issue under investigation, the good used, attributes of the subject pool, endowment to buyers, and the measurement instrument. See http://www.e-aer.org/data/june05_app_plott.pdf.

[5] Note Charles A. Holt (1986) demonstrated that if the experimenter elicits valuations of a number of lotteries over several rounds and then randomly selects one round for which subjects will be paid, mechanisms designed to produce truthful revelations of valuations will not necessarily produce such revelations.

nisms, care is taken to measure "true" valuations void of the influences of strategic behavior. Blank cells in this column indicate that the mechanism used to elicit valuations was not incentive compatible. Therefore, when evaluating these experiments one cannot assume that the responses elicited represent the "true" valuations of the subjects. Shaded cells indicate that the experimenters used a theoretically incentive-compatible mechanism to elicit valuations. The actual mechanisms used are described next.

Valuations Elicited Using Market Environment with Some Incentives.—This column reveals that experimenters have employed a wide variety of market environments to elicit valuations. Blank cells indicate that the experiment elicited valuations outside a market environment. Experiments employing market environments and/or some incentives have used Smith auctions, binary choice designs, sealed-bid one-price auctions, double auction call markets, Vickrey auctions, and the Becker, DeGroot, Marschak ("BDM") mechanism (ordered from least (theoretically) incentive compatible to most (theoretically) incentive compatible).

Gap Measured Directly Using Valuations.—The final column provides information about the method experimenters use to measure gaps. Blank cells indicate that the gap was measured using the number of trades relative to the predicted number of trades.[6] Shaded cells indicate that the gap is measured using the mean or median of actual WTP and WTA responses.

Table 1 does not include information on other experimental procedures that differ across experiments. These include: the good, attributes of the subject pool, bounds or restrictions on bids, cash endowments to buyers, whether the seller is physically endowed with the good, anonymity of decisions, and the statistical method used to measure the gap. Clearly such procedures possibly interact with unobserved variables such as subjects' understanding, attention, motivation, etc., but the exact nature of such interactions is unknown. Such deep issues are not explored directly in this study, but our procedures control for as many as possible.

Table 1 clearly answers any question about the consensus concerning the existence and nature of the gap. In short, no consensus exists. Furthermore, the table suggests that the likelihood an experiment results in a WTP-WTA gap appears to be related to the experimental procedures.[7]

II. KKT Replication Design and Results

This section discusses the experiment design reported in KKT (1990), which we attempt to replicate. The results from our replication indicate that our attempt was successful. When using the design reported by KKT, we observe a WTP-WTA gap.

We conducted two identical sessions with undergraduates from the California Institute of Technology. Each session consisted of two unpaid practice rounds using induced-value tokens and one binding round using mugs bearing a Caltech logo purchased at the bookstore for $7.00 each. The sessions lasted approximately 15 minutes and subjects earned less than $10. Payouts were not made anonymously.

The instructions replicated those used by KKT to test for misconceptions by subjects (referred to as "Experiment 5" in their paper). One-half of the subjects were given mugs and were referred to as "sellers." The remaining subjects were referred to as "buyers" and received no mugs. Buyers were allowed to inspect the mug of the seller sitting next to them. Each subject was assigned the same role in each of the three rounds (i.e., once a buyer, always a buyer). All subjects used the list method to reveal their values for the mug.[8] We used the

[6] Here we note an important point about methods used to measure the gap. If the goal of measuring the gap is to conclude whether WTA is significantly higher than WTP, then a distinction should be made between direct and indirect measurements. Specifically, comparing the number of actual trades to the predicted number of trades might not accurately reveal how WTA relates to WTP. Robert Franciosi et al. (1996) provide a clear example demonstrating that this measure might not accurately determine whether a significant gap between WTA and WTP exists in the data. The reader should keep this point in mind when considering Table 1.

[7] We did not attempt to include all studies that might have some bearing on the issue; therefore, the table should not be interpreted as a "meta analysis."

[8] See the on-line Appendix for an example of the list subjects used to report their valuations (http://www.e-aer.org/data/june05_app_plott.pdf.

TABLE 2—INDIVIDUAL SUBJECT DATA AND SUMMARY STATISTICS FROM KKT REPLICATION

Treatment	Individual responses (in U.S. dollars)	Mean	Median	Std. dev.
WTP (n = 29)	0, 0, 0, 0, 0.50, 0.50, 0.50, 0.50, 0.50, 1, 1, 1, 1, 1, 1.50 2, 2, 2, 2, 2, 2.50, 2.50, 2.50, 3, 3, 3.50, 4.50, 5, 5	1.74	1.50	1.46
WTA (n = 29)	0, 1.50, 2, 2, 2.50, 2.50, 3, 3.50, 3.50, 3.50, 3.50, 3.50, 4, 4.50 4.50, 5.50, 5.50, 5.50, 6, 6, 6, 6.50, 7, 7, 7, 7.50, 7.50, 7.50, 8.50	4.72	4.50	2.17

BDM mechanism to determine which subjects would transact and the price at which transactions would occur.[9] Buyers used their own money and were told that credit and change were available before the start of the experiment.

The data collected during the mug round of the KKT replication is displayed in Table 2.

The mean WTP response was \$1.74 (median = \$1.50) and the mean WTA response was \$4.72 (median = \$4.50). Statistical test results support the hypotheses that the two independent samples (WTA and WTP) were drawn from nonidentical distributions and the median WTA is significantly greater than the median WTP.[10] This result demonstrates a successful replication of the result obtained by KKT using the procedures employed in their study.[11]

[9] The BDM mechanism pits each seller and buyer against a random bid. All sellers stating bids lower than the random bid sell the good, but receive an amount of money equal to the random bid. All buyers stating bids higher than the random bid buy the good, but pay an amount of money equal to the random bid. Sellers who bid higher than the random bid, and buyers who bid lower than the random bid, do not transact.

[10] A Wilcoxon-Mann-Whitney test resulted in a z value of -4.8 with a corresponding p value of 0.00; similarly, a test of the equality of medians resulted in a Pearson χ^2 statistic of 20 (probability of equal medians equals 0). These statistical tests are discussed in Sections 6.4 and 6.3, respectively, of Sidney Siegel and N. John Castellan (1988).

[11] We note here that KKT conducted induced value token rounds and used those data as evidence that their subjects had no misconceptions about how responses affected outcomes (i.e., subjects almost always rationally responded with their induced values). Our lottery data allow us to draw similar conclusions about the level of subject understanding. When viewing our data for the lottery rounds with a certain dollar outcome (e.g., 30-percent chance of \$2 and 70-percent chance of \$2), we observe correct reporting of values similar to the accuracy for induced value rounds that were reported by KKT. That is, the proportion of subjects who gave the correct response was the same in our experiments as that reported by KKT. Yet, when the lotteries for certain outcomes are followed by additional procedures designed to remove misconceptions, as was done as part of our procedures, the behavior substantially changes. From that fact, we conclude that data from induced value rounds or data from rounds involving lotteries for certain outcomes should not be used to test for the existence (or absence) of misconceptions.

III. Plott-Zeiler Procedures and Results

This section discusses our experiments, which implement the controls for subject misconceptions found in the literature. If observed WTA-WTP gaps are explained by endowment effect theory, then we should observe a gap when we alter the procedures to control for subject misconceptions. As reported in detail below, however, we observed no gap.

A. *Plott-Zeiler Procedures*

We collected three sets of data. Two sessions of the experiment were conducted with law students at the University of Southern California Law School in Los Angeles. One session was conducted with undergraduate students at Pasadena City College in Pasadena. Each session consisted of a detailed training session, two unpaid practice rounds, 14 paid rounds using lotteries, and one paid round using mugs. The sessions lasted approximately 90 minutes, and subjects earned \$32 on average (including a \$5 show-up fee). Upon entering the room, each subject chose a laminated card at random indicating the subject's identification number. The subjects were told to keep the identification numbers private to facilitate anonymous payouts at the end of the experiment. Also, subjects were asked to avoid communicating with other subjects and verbally reacting to events that occurred during the experiment.

We designed the procedures explicitly to control for concerns identified in the literature. In particular, Table 1 suggests that a gap is ob-

served less often when an incentive-compatible mechanism is used to elicit valuations, and training and paid practice rounds are provided. Table 1 also reveals the absence of a particular and important set of procedures. Our analysis of the literature reveals that no one experiment designed to study WTP-WTA gaps implements a complete set of controls: an incentive-compatible elicitation device, training, paid practice, and anonymity. We fill this void with our experiment design.

First, using an *incentive-compatible elicitation device* (e.g., BDM) gives subjects an incentive to announce their actual valuations for the good with the goal of increasing the probability of earning the maximum amount possible. Lack of incentives can be associated with several features of arbitrary behavior. Although the specific reasons incentives might operate in such a manner is not well known, presumably incentives focus behavior in the sense that attention, thought, and care in understanding instructions depend on the nature of incentives. If earnings depend on subjects' decisions, subjects probably are more likely to allocate attention to instructions and decisions during experiments.

Second, *training* provides subjects with a basic understanding of the mechanism used to elicit valuations. Mechanisms used to elicit valuations might be unfamiliar to subjects or, more important, might be so similar to mechanisms with which subjects are familiar that subtle and important differences go unnoticed despite experimental controls. Many designs use incentive-compatible mechanisms such as the BDM mechanism to elicit nonstrategic valuations from subjects. These mechanisms, however, most likely are unfamiliar to subjects even though the task might appear to be a common buying or selling task. When confronted with an auction of any type, individuals might tend to operate under familiar auction rules (i.e., highest bidder takes the good and pays the amount he offered). Therefore, even if subjects are told it is in their best interest to bid their "true value," misconceptions about how the elicitation mechanism works might trigger subjects to default to the strategies associated with familiar auctions.

Our approach is based on a presumption that, to accurately measure preferences, misconceptions about the valuation elicitation mechanism must be eliminated. The presumption is that subjects must have a good operational understanding of procedures, including the available alternatives and the mapping of revealed valuations to consequences. Decision theorists might find the language used to describe procedures to be very clear because they are trained to give operational meaning to technical language, e.g., "true value."[12] To those not so schooled, however, the language can be unclear. In many cases, conducting paid practice rounds might be necessary to ensure that subjects understand the procedures and how revealed valuations map into consequences.

As part of the training process, numerical examples provide concrete illustrations, allowing subjects to see the mechanism in terms of its purpose. In addition, specific examples are used to illustrate why announcing valuations that are not actual valuations is a dominated strategy when the BDM mechanism is employed.

Third, *practice rounds* allow subjects to learn by gaining familiarly with the mechanism while still educating themselves about its properties. Encouraging questions during the practice rounds assists subjects in clearing up any misconceptions. In addition, the nonanonymous practice rounds give the experimenter an opportunity to check whether subjects were displaying behavior consistent with a clear understanding of the valuation task.

Providing *paid* practice rounds exposes subjects to the consequences of their decisions prior to the round during which subjects report valuations used to measure the WTP-WTA gap. It is well known that activities in people's daily lives automatically place them in situations of strategic interaction. Strategic reactions developed to engage in those interactions might seep into behavior exhibited in experiments in a manner that clouds gap measurement.[13] For example, the use of the word "sell" can automatically call forth a margin above the minimum that an individual might accept in exchange for a good.

[12] Consider another example. Economists might have a clear meaning of what a "preference" is, but subjects might not clearly recognize this property within themselves or associate it with other words such as likes, dislikes, wants, wishes, etc.

[13] For an evolutionary theory of this phenomenon, see Aviad Heifetz and Ella Segev (2001).

TABLE 3—SUMMARY OF EXPERIMENTS AND ORDER OF THE ROUNDS

Experiment 1: (USC students)	$n = 31$	Rounds 1–3 Small stake sellers	Rounds 4–6 Small stake buyers	Rounds 7–10 Large stake sellers	Rounds 11–14 Large stake buyers	Round 15 Mugs
Experiment 2: (USC students)	$n = 26$	Round 1 Mugs	Rounds 2–4 Small stake sellers	Rounds 5–7 Small stake buyers	Rounds 8–11 Large stake sellers	Rounds 12–15 Large stake buyers
Experiment 3: (PCC students)	$n = 17$	Rounds 1–3 Small stake sellers	Rounds 4–6 Small stake buyers	Rounds 7–10 Large stake sellers	Rounds 11–14 Large stake buyers	Round 15 Mugs

Notes: Experiments 1 and 3 used the BDM mechanism to elicit responses and employed paid practice, training, and anonymity. Experiment 2 used the BDM mechanism to elicit responses and employed training and anonymity (without paid practice rounds).

Even if the word "sell" is not used, simply being in a situation that calls for selling behavior might trigger an instinctive reaction (e.g., sell high and buy low). Despite this, many theories rely on the assumption that subjects in experiments understand their tasks and that observed behavior is not a result of strategic behaviors evoked by instructions. Interpreting data that might contain a mixing of motives layered over actual valuations can prove difficult, however.

Likewise, if subjects mistakenly believe that outcomes are manipulable, they might behave according to a strategy the mechanism does not reward. For example, if a subject is asked to provide a "selling price" that reflects his valuation for a good he owns, natural instincts might persuade him to announce an amount higher than his actual valuation. In fact, given bargaining instincts of sellers to inflate asks and buyers to deflate bids, those endowed with a good likely will bid more than their nonstrategic valuations. This behavior could be especially likely if subjects do not fully understand experimental procedures.

To control for this possibility, during the paid rounds subjects learn about the intricacies of the elicitation mechanism and are given an opportunity to adjust nonoptimal strategies to maximize their payouts. Most important, this learning and adjustment process takes place before the gap is measured, minimizing the possibility that the measurement of the gap includes strategic responses or responses that are clouded by misconceptions about the mechanism.

Finally, *anonymity* in decisions and payouts is ensured. Some commentators (e.g., Gertrud M. Fremling and Richard A. Posner, 2001) hypothesize that if decisions are not made anonymously, subjects might be concerned with how others view them. For example, talented and successful bargainers tend to sell high and buy low. Therefore, if a subject wishes to be known by other subjects or the experimenter as a talented bargainer, he might adjust his behavior accordingly even if the elicitation device does not reward that type of behavior. In addition, when making anonymous decisions, subjects might be less inclined to ponder the "correct" answer as viewed by others, and instead focus on choosing the amount that will reward them the most. While we attach no particular weight to any number of ideas about how subjects might want to represent themselves, we remove the opportunity and incentives for any such attempt.

After the two practice rounds, each subject participated in 15 paid rounds: 14 rounds conducted with lotteries and one round conducted with mugs. We used the data collected during the mug round to measure the gap. In the experiments including paid practice rounds, the mug round was conducted after the lottery rounds. In the experiment without paid practice rounds, the mug round was conducted before the lottery rounds. The first six lottery rounds involved lotteries with expected values of less than $1. The subjects were told that the lotteries would increase in magnitude, but the first few rounds allowed for additional (but paid) practice. All subjects acted as sellers in the first three lottery rounds and buyers in the second three lottery rounds. The first set of "large-stakes" lottery rounds (four in total) involved lotteries with expected values ranging from $2 to $8. All subjects acted as sellers during these rounds. During the second set of large-stakes lottery rounds (four in total), all subjects acted as buyers. Subjects were allowed to view only the lottery involved in the round being conducted. Table 3 summarizes the experiment design.

In the mug round, the item considered by the

TABLE 4—INDIVIDUAL SUBJECT DATA AND SUMMARY STATISTICS

Experiment	Treatment	Individual responses (in U.S. dollars)	Mean	Median	Std. dev.
Experiment 1: (USC/practice)	WTP (n = 15)	0, 1, 1.62, 3.50, 4, 4, 4.17, 5, 6, 6, 6.50, 8, 8.75, 9.50, 10	5.20	5.00	3.04
	WTA (n = 16)	0, 0.01, 3, 3.75, 3.75, 3.75, 5, 5, 5, 6, 6, 6, 7, 11, 12, 13.75	5.69	5.00	3.83
Experiment 2: (USC/no practice)	WTP (n = 12)	1, 2, 3.50, 5, 5, 5, 8, 8.50, 9, 11.50, 13, 23	7.88	6.50	6.00
	WTA (n = 14)	0.50, 1, 2, 2.50, 2.50, 4.50, 4.50, 5.70, 6.25, 8, 8, 8.95, 12, 13.50	5.71	5.10	4.00
Experiment 3: (PCC/practice)	WTP (n = 9)	2.50, 5.85, 6, 7.50, 8, 8.50, 8.50, 8.78. 10	7.29	8.00	2.23
	WTA (n = 8)	3, 3, 3.50, 3.50, 5, 5, 7.50, 10	5.06	4.25	2.50
Pooled data	WTP (n = 36)		6.62	6.00	4.20
	WTA (n = 38)		5.56	5.00	3.58

Notes: Experiments 1 and 3 used the BDM mechanism to elicit responses and employed paid practice, training, and anonymity. Experiment 2 used the BDM mechanism to elicit responses and employed training and anonymity (without paid practice rounds).

subjects was a plastic travel mug with a market price of approximately $8.50. The subjects were not informed of the market price. Approximately half the subjects acted as sellers and approximately half acted as buyers. All subjects were handed a mug before the start of the round. Sellers were told that they owned the mug. Buyers were told that they could inspect the mug but they did not own it. All subjects were prompted to record an offer (sellers offering the minimum amount they would accept to give up the mug, buyers offering the maximum amount they would pay in exchange for the mug). After offers were recorded, subjects were prompted to consider whether the offer chosen was the actual nonstrategic value and were allowed to change the offer before committing to it. After all committed offers were collected (i.e., slips placed into the boxes), the predetermined fixed offer was announced. The subjects recorded their round payoffs and accumulated payoffs for the experiment.[14]

B. *Plott-Zeiler Results*

Each subject revealed a personal valuation for a mug, either from the point of view of someone who owned the mug and is given an opportunity to sell it, or from the point of view of someone who has no mug but is given the opportunity to buy one. Thus, we perform a between-subject test to determine whether a significant WTP-WTA gap resulted.

In experiments 1 and 3, values for the mugs were collected after 14 rounds used to provide paid practice, during which subjects made decisions involving binding lotteries. In experiment 2, such practice rounds were absent prior to the mug round. Table 4 contains data on subjects' responses during the mug rounds. Data are reported for a total of 74 subjects.[15]

[14] See the online Appendix for detailed subject instructions and procedures (http://www.e-aer.org/data/june05_app_plott.pdf).

[15] Theoretically, the lottery rounds could be used to test for a WTP-WTA gap. The lottery round data, however, are contaminated by a design that was developed only for training and not for purposes of measuring a gap. All WTA rounds were conducted prior to the WTP rounds. Thus, subjects had extensive training on one mechanism (i.e., selling using the BDM mechanism), the meaning of lotteries, the instructions, the procedures, and other subtle features of the experiment design before being exposed to the second mechanism (i.e., buying using the BDM mechanism). Because the BDM mechanism would be more unfamiliar in a selling task than a buying task (subjects tend to have more experience buying than selling), the WTA rounds were used first. In addition to what appeared to us to be a logical approach to training, this procedure also ensured that subjects had money to spend in the WTP rounds, which eliminated complex explanations of collecting from losses, banking, and borrowing that would otherwise make learning more difficult. Mistake corrections, public answers to questions, and other procedures were also employed continuously, which confound the valuations provided in the lottery rounds and frustrate attempts to use these data to measure gaps.

Of course, the data from these rounds are not uninteresting

TABLE 5—STATISTICAL TEST RESULTS

	Wilcoxon-Mann-Whitney rank sum test (Null hypothesis: identical distributions)			Median test (Null hypothesis: populations have identical medians)		
	z	p-value	Conclusion (α = .05)	Pearson χ^2	p-value	Conclusion (α = .05)
Experiment 1 (USC)	−0.079	0.9368	Can't reject null	0.0392	0.843	Can't reject null
Experiment 2 (USC)	0.928	0.3536	Can't reject null	0.1548	0.694	Can't reject null
Experiment 3 (PCC)	1.738	0.0821	Can't reject null	1.5159	0.218	Can't reject null
Pooled data	1.267	0.2050	Can't reject null	1.3523	0.245	Can't reject null

Notes: The Kruskal-Wallis equality-of-populations rank test indicates that WTA responses from different experiments were drawn from the same population and WTP responses from different experiments were drawn from the same population; therefore, pooling the data is appropriate. The Pearson χ^2 statistics were corrected for continuity.

Table 4 also provides summary statistics for each experiment. In experiment 1, the mean WTP response was \$5.20 (median = \$5.00) and the mean WTA response was \$5.69 (median = \$5.00). In experiment 2, the mean WTP response was \$7.88 (median = \$6.50) and the mean WTA response was \$5.71 (median = \$5.10). Finally, in experiment 3, the mean WTP response was \$7.29 (median = \$8.00) and the mean WTA response was \$5.06 (median = \$4.25).

WTP-WTA Gap Results.—The main finding is striking and ubiquitous across experiments. No gap is observed. The following statements provide the details behind this finding.

Result 1. The data do not support the hypothesis that WTA is significantly greater than WTP in both experiments using the BDM mechanism to elicit valuations and employing training, anonymity, and paid practice rounds.

Support. In Table 5 we report the results of statistical tests to determine whether the data support the hypothesis that WTA is significantly greater than WTP. For all experiments, we perform Wilcoxon-Mann-Whitney tests, which test for whether the WTP and WTA samples were drawn from identical distributions, and median tests, which test for whether the WTP and WTA samples were drawn from distributions with identical medians.

The hypothesis that WTA is significantly greater than WTP, when the BDM mechanism is used to elicit responses and training, anonymity, and paid practice rounds are provided, is not

and contain some hints about the sources and nature of the misconceptions that the procedures help remove. We pass these along as mere speculations and conjectures with the hope that they will be useful in some way. First, subject misconceptions seemed to originate from three sources. The first source is the elicitation mechanism and the BDM procedure, in particular. Subjects seem to make two very different mistakes. Some subjects do not realize that overbidding (underbidding) in the buying (selling) task exposes them to a loss (if the price falls between the true valuation and the bid). Other subjects do not realize that underbidding (overbidding) in the buying (selling) task exposes them to an opportunity cost of a foregone profitable transaction. After instruction and, in some cases, experiencing these consequences, they seem to recognize and adjust to these features of the mechanism, sometimes only after repeated experiences.

The second source of misconceptions is the concept of randomization and the nature of probability. In some cases subjects do not understand statistical independence; they seem to believe that they can predict the future from past events. In other cases subjects seem to believe that a particular outcome will occur with probability zero or one. As a result, subjects think that they have the capacity to guess the outcome, depending in part on an impulse or an urge. Experience seems necessary for subjects unfamiliar with random devices to incorporate true notions of randomization and the nature of probability.

The third source is the assignment of value to a lottery where random devices are employed to determine monetary payoffs. Valuing lotteries is not a common activity for most subjects, and they do not perform this task with the immediate and automatic instincts of a decision theorist. Basically, the concept of an expected value is foreign to some subjects, and they struggle in various ways to quantify their preferences over lotteries. Experiencing the consequences of their choices might help subjects familiarize themselves with the random nature of the outcomes.

substantiated by the data in either experiment 1 or experiment 3. With respect to experiment 1, the Wilcoxon-Mann-Whitney rank sum test statistic resulted in a z value of -0.079 ($p = 0.9368$); therefore, we cannot reject the null hypothesis that the two independent samples were drawn from identical population distributions. In addition, a median test resulted in a Pearson χ^2 test statistic of 0.0392 ($p = 0.843$); therefore, we cannot reject the null hypothesis that the two independent samples were drawn from populations that have identical medians. These two statistical tests were also performed using the data collected during experiment 3. Similar results obtained: the Wilcoxon-Mann-Whitney rank sum test statistic produced a z value of 1.738 ($p = 0.0821$) and the median test resulted in a Pearson χ^2 test statistic of 1.5159 ($p = 0.218$).

Result 2. The data do not support the hypothesis that WTA is significantly greater than WTP in the experiment using the BDM mechanism to elicit valuations and employing training and anonymity (with no paid practice rounds).

Support. The hypothesis that WTA is significantly greater than WTP, when the BDM mechanism is used to elicit responses and training and anonymity are provided (without paid practice rounds), is not substantiated by the data produced in experiment 2. The Wilcoxon-Mann-Whitney rank sum test statistic produced a z value of 0.928 ($p = 0.3536$) and the median test resulted in a Pearson χ^2 test statistic of 0.1548 ($p = 0.694$).[16]

Clearly the extensive instruction and training might have removed misconceptions without a need for paid practice rounds to do so. It should be noted that other researchers conducted experiments without paid practice rounds and observed a gap. Other aspects of the instructions differed from ours, however. This fact might help reconcile our results with those obtained in other studies. The main point, however, is that paid practice rounds seem unnecessary in the presence of other procedures thought to control for subject misconceptions.

[16] The individual data from experiment 2 suggest that the buyer who offered $23.00 for the mug might be driving the result. Evaluating the data without this high offer, however, produces identical results.

Given the high variance in the data, we performed a check on the power of our statistical tests by testing the null hypothesis of WTA $= 2 \cdot$ WTP. Many claim that WTA seems to be twice WTP (see e.g., W. R. Dubourg et al. (1994) and Jack L. Knetsch et al. (2001)). A t test assuming unequal variances led to a rejection of the null in favor of the alternative, WTA $< 2 \cdot$ WTP ($t = -5.06$, $p = 0.0000$). A two-sample Wilcoxon-Mann-Whitney rank-sum test also rejects this null ($z = 4.64$, $p = 0.0000$), as does a test of equal medians (Pearson $\chi^2 = 19.53$, $p = 0.0000$). It also should be noted that, while we observe a WTP that is on average greater than a WTA, the difference is not statistically significant.

House Money Effect Conjecture.—The dramatic difference between measurements taken under a full set of controls for misconceptions and the measurements taken under the KKT procedures motivates questions about how particular procedural features might contribute to the differences. One question focuses on whether "house money effects" might explain our results. Specifically, it could be that subjects might be more willing to spend money earned during the experiment than money taken from their own pockets. A house money effect might elevate WTP in a manner that eliminates the gap. More precisely, the hypothesis is that a house money effect acts asymmetrically to increase WTP and reduce the difference between WTP and WTA.[17] The following result addresses this conjecture.

Result 3. The data do not support the house money effect hypothesis. That is, there is no support for the hypothesis that money earned during the practice rounds accounts for the fact that WTA does not exceed WTP in our experiments.

Support. The support for this result originates from two sources. First, in experiment 2 mug valuations are revealed before money is earned during the practice rounds. In that experiment, house money effects could not have played a

[17] We are indebted to Colin Camerer, Richard Thaler, and Leeat Yariv for drawing our attention to this hypothesis.

role because subjects did not earn money before they participated in the mug round. Yet, in that experiment, WTA did not exceed WTP. Thus, this experiment incorporated a direct control and provides no support for the house money effect conjecture.

The second source of support is the revealed mug valuations themselves. The house money effect conjecture implies that there is some relationship between money earned in the practice rounds and the revealed mug valuations. To test for the existence of such a relationship we regressed individual revealed mug valuations against the amounts earned during the practice rounds. The regression analysis produced the following equations:

$$\mathrm{WTP}_i = 5.77 + 0.0108Y_i$$

$$(t = 3.816) \quad (t = 0.155)$$

$$\mathrm{WTA}_i = 5.37 + 0.005Y_i$$

$$(t = 3.55) \quad (t = 0.08)$$

where Y_i represents subject i's income prior to the mug round. The regression results indicate that none of the variation in mug valuations is explained by variation in income earned during the practice rounds. Not only are the coefficients close to zero, but also they are not statistically significantly different from zero. These results allow us to reject strongly the hypothesis that income earned during the practice rounds had a substantial effect on either WTP or WTA.

IV. Discussion and Conclusions

The issue explored here is not whether a WTP-WTA gap can be observed. Clearly, the experiments of KKT and others show not only that gaps can be observed, but also that they are replicable. Instead, our interest lies in the interpretation of observed gaps. The primary conclusion derived from the data reported here 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. In addition, our results suggest that observed gaps should not be interpreted as support for prospect theory.

A review of the literature reveals that WTP-WTA gaps are not reliably observed across experimental designs. Given the nature of reported experimental designs, we posited that differences in experimental procedures might account for the differences across reported results. This conjecture prompted us to develop procedures to test for the robustness of the phenomenon. We conducted comparative experiments using procedures commonly used in studies that report observed gaps (i.e., KKT). We also employed a "revealed theory" methodology to identify procedures reported in the literature that provide clues about experimenter notions regarding subject misconceptions. We then conducted experiments that implemented the union of procedures used by experimentalists to control for subject misconceptions. The comparative experiments demonstrate that WTP-WTA gaps are indeed sensitive to experimental procedures. By implementing different procedures, the phenomenon can be turned on and off. When procedures used in studies that report the gap are employed, the gap is readily observed. When a full set of controls is implemented, the gap is not observed.

The fact that the gap can be turned on and off demonstrates that interpreting gaps as support for endowment effect theory is problematic. The mere observation of the phenomenon does not support loss aversion—a very special form of preferences in which gains are valued less than losses. That the phenomenon can be turned on and off while holding the good constant supports a strong rejection of the claim that WTP-WTA gaps support a particular theory of preferences posited by prospect theory. Loss aversion might in some sense characterize preferences, but such a theory most likely does not explain observed WTP-WTA gaps.

Exactly what accounts for observed WTP-WTA gaps? The thesis of this paper is that observed gaps are symptomatic of subjects' misconceptions about the nature of the experimental task. The differences reported in the literature reflect differences in experimental controls for misconceptions as opposed to differences in the nature of the commodity (e.g., candy, money, mugs, lotteries, etc.) under study.

That said, we hasten to add that our thesis is not especially satisfying because we have nei-

ther 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, as they might depend on such subtle features as the speed with which experimental instructions are delivered, the distance of subjects from the chalkboard if it is used, the size of writing on the board, how loud the instructions are read, and the nature of pauses or emphasis. Understanding appears to us to be a delicate matter and to control fully for it represents a daunting task. In fact, we have no direct evidence that our procedures actually eliminate all misconceptions about how revealed valuations map into payoffs. What we have shown, however, is that when an experiment simultaneously implements all known controls for misconceptions, a gap is not observed.

Several possible interpretations avail themselves. Each is a matter of speculation, but we list them in order to facilitate discussion. One interpretation is that WTP-WTA gaps are observed when revealed valuations are confounded by ill-conceived motivations to announce something other than a "true" valuation. Under this interpretation the lack of robustness of the gap is due to differences in levels of understanding by the subjects. When the procedures and method for measuring the gap carefully control for such motivations, gaps are not observed. Under this interpretation, use of the label "endowment effect" to describe observed gaps reflects an inappropriate application of prospect theory.

A second interpretation is that the procedures themselves remove attitudes that would foster any difference between WTA and WTP. In particular, according to this conjecture, by allowing the subjects to participate in both the buying side and the selling side of the lottery rounds, objects are translated into commodities for which neither ownership nor loss plays a particular part in the preference formation process. Then, in the subsequent mug round, the attitude toward lotteries somehow is transferred to the mug. Under this interpretation, the procedures (and not the actual measurement of preferences) play a role in the transformation of preferences influenced by loss aversion to preferences not influenced by loss aversion. Of course, prospect theory says nothing about such a dynamic development of preferences. No evidence exists to support the conjecture, and there is evidence that works against it. Specifically, in experiment 2 the mug round was conducted before the subjects participated in the lottery round, and no gap was observed. Therefore, experience with the lotteries could not have played a crucial role in the disappearance of the gap in that instance.

A third interpretation is that the procedures themselves involve a type of demand effect in which the subjects perceive that the experimenter wants to strip from responses any special value of ownership. The conjecture is that by responding to a demand that the answers be "thoughtful," the subjects remove from the response a preference related to "ownership" that would otherwise be reflected in choice. While the mechanism through which this transformation is supposed to take place is not clear, the conjecture itself cannot be rejected using the data from our experiments.

A fourth interpretation is that the procedures suggested some particular value as the "correct" response and that our measurements recorded the suggested value as opposed to preferences. As the same procedures were used for all subjects, the conjecture implies that the valuations elicited from the subjects should all be similar. Our data, however, do not support this conjecture. We observed significant variance in responses for each experiment. We mention here one interesting aside. The test for understanding that KKT employed and that we use as well (in conjunction with other procedures) could be subject to criticism under this conjecture. The fact that subjects reveal "correct" valuations might simply reflect their tendency to report the suggested value (i.e., the certain lottery value in our design or the induced token value in KKT's design). Thus, correct answers under these conditions might not demonstrate that subjects understood the mechanism.

A fifth interpretation is that the WTP-WTA gap reflects features of a decision process, as opposed to a preference. Plott (1996) advances a "discovered preference hypothesis," positing that responses in experiments reflect a type of internal search process in which subjects use paid practice rounds along with trial and error to "discover" what their preferences are. As the subjects gain experience, they begin to discover their preferences, which are then reflected in

their behavior. The hypothesis is that stages of the process can be identified and during the initial stages, when the situation is least familiar to the subject, the effects of framing are most important. Under this interpretation prospect theory itself emerges as a stage of the process. Rather than describing a feature of preferences, the theory describes the features of an early stage of the preference discovery process. Under this interpretation prospect theory becomes part of a theory of how the process of cognition interacts with preference formation and decision making.

We 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. We do agree, however, that sorting out the conditions under which we observe a gap is a necessary precursor to understanding the nature of the gap. We also agree that endowment effect theory and prospect theory most likely do not explain observed WTP-WTA gaps. Finally, claims that WTP-WTA gaps are unrelated to experimental procedures are clearly misleading.[18]

REFERENCES

Arlen, Jennifer; Spitzer, Matthew and Talley, Eric. "Endowment Effects within Corporate Agency Relationships." *Journal of Legal Studies*, 2002, *31*(1), pp. 1–37.

Bateman, Ian J.; Kahneman, Daniel; Munro, Alistair; Starmer, Chris and Sugden, Robert. "Testing Competing Models of Loss Aversion: An Adversarial Collaboration." *Journal of Public Economics* (forthcoming).

Bateman, Ian J.; Munro, Alistair; Rhodes, Bruce; Starmer, Chris and Sugden, Robert. "A Test of the Theory of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 1997, *112*(2), pp. 479–505.

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

Boyce, Rebecca R.; Brown, Thomas C.; McClelland, Gary H.; Peterson, George L. and Schulze, William D. "An Experimental Examination of Intrinsic Values as a Source of the WTA-WTP Disparity." *American Economic Review*, 1992, *82*(5), pp. 1366–73.

Brookshire, David S. and Coursey, Don L. "Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures." *American Economic Review*, 1987, *77*(4), pp. 554–66.

Coursey, Don L.; Hovis, John L. and Schulze, William D. "The Disparity between Willingness to Accept and Willingness to Pay Measures of Value." *Quarterly Journal of Economics*, 1987, *102*(3), pp. 679–90.

Dubourg, W. R.; Jones-Lee, M. W. and Loomes, Graham. "Imprecise Preferences and the WTP-WTA Disparity." *Journal of Risk and Uncertainty*, 1994, *9*(2), pp. 115–33.

Franciosi, Robert; Kujal, Praven; Michelitsch, Roland; Smith, Vernon L. and Gang, Deng. "Experimental Tests of the Endowment Effect." *Journal of Economic Behavior and Organization*, 1996, *30*(2), pp. 213–26.

Fremling, Gertrud M. and Posner, Richard A. "Market Signaling of Personal Characteristics." University of Chicago, John M. Olin Law & Economics Working Paper: No. 87 (2nd series), 1999.

Harless, David W. "More Laboratory Evidence on the Disparity between Willingness to Pay and Compensation Demanded." *Journal of Economic Behavior and Organization*, 1989, *11*(3), pp. 359–79.

Heifetz, Aviad and Segev, Ella. "The Evolutionary Role of Toughness in Bargaining." *Games and Economic Behavior*, 2004, *49*(1), pp. 117–34.

Holt, Charles A. "Preference Reversals and the Independence Axiom." *American Economic Review*, 1986, *76*(3), pp. 508–15.

Horowitz, John K. and McConnell, Kenneth E. "A Review of WTA/WTP Studies." *Journal of Environmental Economics and Management*, 2002, *44*(3), pp. 426–47.

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

[18] See, e.g., Russell Korobkin (2003), who states that "[a]lthough experimental conditions probably have some explanatory power in some cases, the weight of the evidence suggests that it is extremely unlikely that the effect is merely an artifact of the experimental methods that demonstrate it."

Knetsch, Jack L. "The Endowment Effect and Evidence of Nonreversible Indifference Curves." *American Economic Review*, 1989, *79*(5), pp. 1277–84.

Knetsch, Jack L. and Sinden, Jack A. "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value." *Quarterly Journal of Economics*, 1984, *99*(3), pp. 507–21.

Knetsch, Jack L.; Tang, Fang-Fang and Thaler, Richard H. "The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing?" *Experimental Economics*, 2001, *4*(3), pp. 257–69.

Knez, Peter; Smith, Vernon L. and Williams, Arlington W. "Individual Rationality, Market Rationality, and Value Estimation." *American Economic Review*, 1985 (Papers and Proceedings), *75*(2), pp. 397–402.

Korobkin, Russell. "The Endowment Effect and Legal Analysis." *Northwestern University Law Review*, 2003, *97*(3), pp. 1227–93.

Lehmann, Erich L. *Nonparametrics: Statistical methods based on ranks*. San Francisco: Holden-Day, Inc., 1975.

List, John A. and Shogren, Jason F. "Price Information and Bidding Behavior in Repeated Second-Price Auctions." *American Journal of Agricultural Economics*, 1999, *81*(4), pp. 942–49.

Loewenstein, George and Issacharoff, Samuel. "Source Dependence in the Valuation of Objects." *Journal of Behavioral Decision Making*, 1994, *7*(3), pp. 157–68.

Mann, Henry B. and Whitney, Roy D. "On a Test of Whether One of Two Random Variables Stochastically Larger Than the Other." *Annals of Mathematical Statistics*, 1947, *18*(1), pp. 50–60.

Morrison, Gwendolyn C. "Willingness to Pay and Willingness to Accept: Some Evidence of an Endowment Effect." *Applied Economics*, 1997, *29*(4), pp. 411–17.

Ortona, Guido and Scacciati, Francesco. "New Experiments on the Endowment Effect." *Journal of Economic Psychology*, 1992, *13*(2), pp. 277–96.

Plott, Charles R. "Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis," in Kenneth J. Arrow, Enrico Colombatto, Mark Perlman, and Christian Schmidt, eds., *The rational foundations of economic behavior*. IEA Conference Vol. 114. London: Macmillan, 1996, pp. 225–50.

Rachlinski, Jeffrey J. and Jourden, Forest. "Remedies and the Psychology of Ownership." *Vanderbilt Law Review*, 1998, *51*(6), pp. 1541–82.

Shogren, Jason F.; Cho, Sungwon; Koo, Cannon; List, John A.; Park, Changwon; Polo, Pablo and Wilhelmi, Robert. "Auction Mechanisms and the Measurement of WTP and WTA." *Resource and Energy Economics*, 2001, *23*(2), pp. 97–109.

Shogren, Jason F.; Shin, Seung Y.; Hayes, Dermot J. and Kliebenstein, James B. "Resolving Differences in Willingness to Pay and Willingness to Accept." *American Economic Review*, 1994, *84*(1), pp. 255–70.

Siegel, Sidney and Castellan, N. John. *Nonparametric statistics for the behavioral sciences*. 2nd Ed. New York: McGraw-Hill Book Co., 1988.

Singh, Harinder. "The Disparity between Willingness to Pay and Compensation Demanded: Another Look at Laboratory Evidence." *Economics Letters*, 1991, *35*(3), pp. 263–66.