# Substitutability, experience, and the value disparity: evidence from the marketplace ☆

## John A. List[a,b,*]

[a] AREC, University of Maryland, 2200 Symons Hall, College Park, MD 20742-5535, USA
[b] National Bureau of Economic Research, Inc., 1050 Massachusetts Avenue, Cambridge, Massachusetts 02138-5398, USA

**Abstract**

Recent empirical evidence suggests that important disparities exist between willingness to pay and compensation demanded for the same good. These results, which clearly contradict closely held economic doctrines, have led some influential commentators to call for an entirely new economic paradigm to displace conventional neoclassical theory. This study examines the generality of these experimental findings by going to a well-functioning marketplace and examining more than 350 individual decisions across two incentive-compatible elicitation mechanisms. The data suggest that behavior of individuals with intense experience approaches neoclassical predictions: any observed WTA/WTP disparity amongst this group is negligible.

## 1. Introduction

A fundamental postulate of neoclassical theory is that preferences between two goods are independent of the consumer's reference point. This basic independence assumption naturally leads to the prediction that with small income effects and many available substitutes, the willingness to accept (WTA) and willingness to pay (WTP) measures of value for a commodity should be roughly equivalent. While most theoretical and applied economic models invoke this assumption, it has been directly refuted in several experimental settings (e.g., [9]).[1] Some commentators have argued, however, that these findings represent merely a mistake made by inexperienced consumers and that through time the behavior of those consumers will be more in line with predictions from neoclassical models (e.g., [3,5,15]). While this argument has intuitive appeal, critics contend that the available evidence is mixed at best, and that overall the data do not conclusively support the learning premise [13,14].

In this study, I take a different approach to understanding the market learning process by observing actual market behavior. This approach provides me with an opportunity to observe behavior of agents that have endogenously selected into certain roles within the market, such as being an intense or casual consumer, while simultaneously making use of certain controls afforded in an experiment. This methodology may lead to different results than in an experiment where the roles are exogenously induced by the experimenter (i.e., some subjects are given experience while others are not), but a rigorous examination of behavior in an actual environment is an important step in testing the validity of economic theory.

The specific goals of this study are twofold. First, I compare WTA statements of value with WTP and "endowment-adjusted" WTP values gathered via two incentive-compatible institutions: a random $n$th price auction and the Becker et al. [2] discrete choice auction.[2] Standard neoclassical arguments suggest that the results of several previous experimental tests that found a disparity between WTA and WTP values (e.g., [5,13]) may be consonant with neoclassical theory after all, since these studies compared value estimates from different indifference curves and thus different marginal rates of substitution applied.[3] By comparing WTA and endowment-adjusted WTP, I control for income and substitution effects and therefore provide a more appropriate comparison of value measures. This distinction allows a cleaner test of alternative theories. Second, I push the substitutability argument a bit harder by examining whether exogenous changes in substitutability influence the WTA/WTP disparity.

The empirical results provide some interesting insights. First, I find that intense consumers (i.e., those consumers that have a history of active market participation) display a negligible value disparity in both elicitation institutions. This result is in accord with List [17,18], and provides a robustness check of his main findings in much different institutional settings. Second, I find that neoclassical theory does an adequate job of organizing the data from experienced consumers, as

---

[1] Early evidence of the anomaly was found in contingent valuation studies (see, e.g., [4]). The interested reader should see Horowitz and McConnell [8] for a literature review of WTA/WTP studies. Throughout the paper I will refer to the difference in compensation demanded and willingness to pay as the WTA/WTP disparity.

[2] Since impacts of alternative market institutions have been found to be non-trivial in other settings (see, e.g., [21,24]), a comparison of value statements across institutions seems worthwhile.

[3] A similar argument can be made concerning studies, such as [3], which compare movements along different portions of the same indifference curve.

the comparative static predictions are generally met. For example, (i) endowment-adjusted WTP better approximates WTA than does WTP, and (ii) the estimate of the effect of income changes on WTP provided by the observed WTA/WTP disparity is more reasonable for intense consumers than for casual consumers. Furthermore, if one takes into account informational asymmetries, neoclassical theory provides an adequate explanation for patterns observed in the pooled data. Third, I find some evidence that exogenous changes in substitutability influence the degree of the WTA/WTP disparity.

The remainder of this study proceeds as follows. Section 2 provides theoretical structure to the experimental design. Section 3 summarizes the empirical results. Section 4 discusses limitations of the study. Section 5 concludes with a summary and directions for future research.

## 2. Background and experimental design

Ever since the initial experimental findings suggesting that mere ownership of a good may induce a reference-dependent preference structure, neoclassical models have been under heavy scrutiny, as their basic independence assumption is clearly violated in such scenarios. In addition to casting doubt on closely held economic doctrines, influential commentators have used the disparity findings to question the entire basis of welfare economics and common interpretations of indifference curves. From a policy perspective, Knetsch [12] notes the importance of the disparity findings as follows: "it is likely that, among other implications, losses are understated, standards are set at inappropriate levels, policy selections are biased, too many environmentally degrading activities are encouraged, and too few mitigation efforts are undertaken." More narrowly, the value disparity has been viewed implicitly as a failure of survey methods [26].

Before becoming immersed in the field experimental design, it is worthwhile to summarize intuitively the major theoretical predictions of neoclassical and endowment theory.[4] Assume that an individual derives utility, $u = u(g, x)$, from consuming two goods, $X$ and $G$, where $X$ could be viewed as a composite commodity, quantities of which could be measured in money units, and $G$ is a particular consumption good. Consider Fig. 1, where perfect substitution exists between the two goods. In this case, neoclassical theory predicts WTA = WTP: minimum WTA (or Hicksian equivalent surplus) of $x_1 - x_0$ is equivalent to maximum WTP (or Hicksian compensating surplus) of $x_0 - x_2$. Upon relaxing the perfect substitutability assumption, and assuming quasi-concave utility functions, these linear indifference curves become strictly convex to the origin, as presented in Fig. 2. In this case, Randall and Stoll [22] demonstrate that the WTA/WTP disparity depends on the "price flexibility of income": the elasticity of the marginal valuation of $g$ with respect to $x$. Hanemann [6] later demonstrated that the "price flexibility of income" is analytically equivalent to the ratio of the ordinary income elasticity of demand for the good to the Allen-Uzawa elasticity of substitution between the good and the numeraire.

Hanemann [6] argues that the large WTA/WTP disparities that have been observed in the literature can, therefore, be reconciled with neoclassical theory. Examining Fig. 2, and measuring

---

[4] Space limitations preclude me from doing this topic justice. I point the interested reader to Hanemann [7], who provides an excellent elucidation of the economic theory of WTA and WTP.
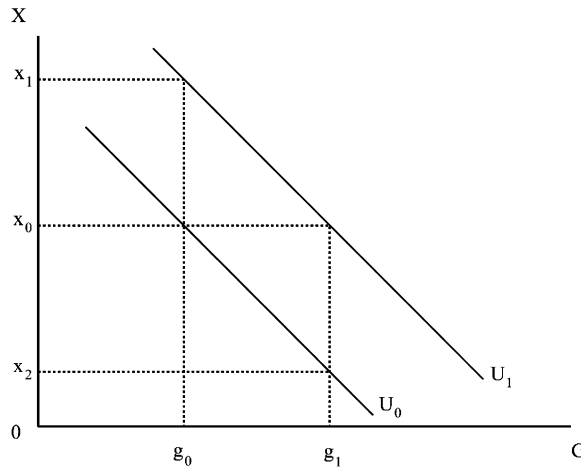
Fig. 1. Linear indifference curves and WTA/WTP.
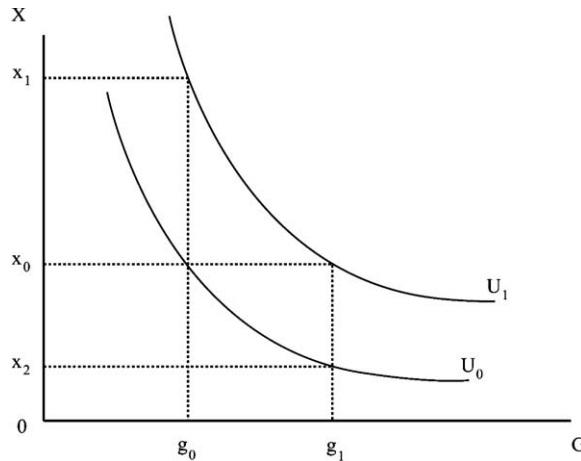
Fig. 2. Convex to the origin indifference curves and WTA/WTP.

WTA (WTP) for changes in $G$ from $g_1$ to $g_0$ ($g_0$ to $g_1$), illustrates this result.[5] In this case, WTA $= x_1 - x_0 > x_0 - x_2 =$ WTP; due to income (movement to a higher indifference curve) and substitution (curvature of the indifference curves) effects, neoclassical theory predicts value divergences.[6] In practice, it would appear that the substitution effect would hold more prominence in shaping the value disparity than would the income effect. The intuition behind this thought is contained in the Engel aggregation condition, which requires that the income elasticities of demand for $g$ and $x$, weighted by their respective budget shares, sum to one. While

---

[5] Note that the curvature of the indifference curves could be thought of as merely a reflection of convexity of the expenditure function in $g$. And, if one assumes utility is quasi-convex in $g$, then a corner solution emerges rather than an interior solution. In this case, indifference curves are concave and the substitution effect becomes positive, rather than negative.

[6] Recall that for price changes, the difference between the compensating and equivalent variation depends simply on the income effect.

this certainly limits the size of the income effect associated with $g$, the substitution effect can range from zero to infinity.[7]

Insights on the disparity are identical if one considers movements along *different* segments of the *same* indifference curve. One can gain this intuition by choosing any point along either convex indifference curve in Fig. 2 and moving equidistant (in $g$) along the indifference curve in opposite directions. Instead of both income and substitution effects causing the divergence, in this case the divergence is caused purely by convexity of the indifference curve. Depending on the structure of the utility function, the divergence observed along different segments of the same indifference curve can be greater or smaller than differences observed when moving along different indifference curves. A majority of the WTA/WTP experimental studies cited in Horowitz and McConnell [8] fall into one of these two categories, having subjects moving along either different indifference curves or different portions of the same indifference curve.

The simple intuition behind Hanneman's [6] neoclassical arguments can be tested by appropriately compensating Group 2 subjects so that they state their maximum WTP to move from $g_0$ to $g_1$ along the *same* indifference curve as Group 1 subjects. In Fig. 2, this can be achieved simply by endowing Group 2 subjects at point $(g_0, x_1)$ and asking those subjects to state their maximum WTP to move from $g_0$ to $g_1$. Comparing these endowment-adjusted WTP values with WTA statements yields a relatively clean test of whether the indifference curve is reversible, as neoclassical theory predicts. Such an exercise naturally controls for both income (different indifference curves) and substitution (curvature of the indifference curves) effects.

This demonstrates the intuition of List [17], who examined trading patterns of sports memorabilia on the floor of a sportscard show, and trading patterns of collector pins in a market constructed by *Walt Disney World* at the *Epcot Center* in Orlando, Florida. In one treatment a subject is endowed with good $A$ and has the option to trade it for good $B$. In a second treatment, a different subject is endowed with good $B$ and has the option to trade it for good $A$. Since subjects are allocated to one of the two treatments randomly, for preferences to be consistent, the proportion of subjects who choose $B$ over $A$ should be equal to one minus the proportion who choose $A$ over $B$.

The experimental test constructed below represents the value elicitation analogs of these trading tests in the marketplace. If preferences are defined over changes in consumption, then one could envision a kink for Group 1 subjects at point $(g_1, x_0)$; this kink would cause the indifference curve for Group 1 subjects to pivot from $U_1$ to $U_1^E$, as displayed in Fig. 3. Under neoclassical arguments, WTA should be measured as $x_1 - x_0$, but assuming the endowment effect induces the indifference curve to pivot, WTA becomes $x_3 - x_0 > x_1 - x_0$, which illustrates the agent's aversion to sacrifice a good once it becomes part of her endowment. The WTA/WTP disparity would be exacerbated if a similar kink formed for Group 2 subjects, as it would pivot their indifference curves counterclockwise from point $(g_0, x_1)$. This seems unlikely if $X$ represents money;

---

[7] Care should be taken not to dismiss income effects out of hand, however. Under certain scenarios, the spacing of the indifference curves is quite important. While in Fig. 2, I have implicitly assumed a unitary income elasticity by making the indifference curves for $U_0$ and $U_1$ parallel displacements, if I assumed an income elasticity greater than 1, or that the marginal rate of substitution increased along any ray from the origin, the WTA/WTP disparity would increase. The opposite would occur, of course, if $G$ were an inferior good—the marginal rate of substitution would decrease along any ray from the origin.
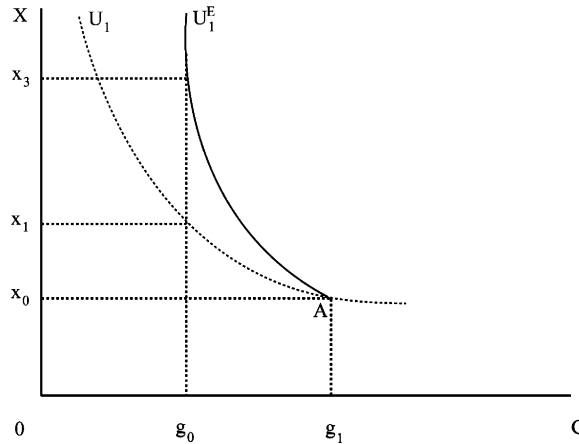
Fig. 3. Endowment effects and the WTA/WTP disparity.

nevertheless, this is the general intuition behind the endowment effect, which implies that a good's value increases once it becomes part of an individual's endowment.

Following Sugden [25], one can push the neoclassical explanation a bit further by exploiting the relationship between the WTA/WTP disparity and the predicted effect of income on WTP. One advantage of such an exercise is that it provides a further use of the observed WTA/WTP ratio and permits an examination of whether the observed ratio is consistent with economic intuition. To add structure to the argument, I continue with Fig. 2 and assume that the initial endowment is $(g_0, x_0)$. As previously suggested, the individual's WTP to move from $g_0$ to $g_1$ is defined as follows:[8]

$$u(g_1, x_0 - \text{WTP}(x_0)) = u(g_0, x_0),$$ (1)

where $\text{WTP}(x_0) = x_0 - x_2$. Similarly, endowment-adjusted WTP is

$$u(g_1, x_1 - \text{WTP}(x_1)) = u(g_0, x_1),$$ (2)

where $\text{WTP}(x_1) = x_1 - x_0$. WTA is defined analogously, thus

$$u(g_0, x_0 + \text{WTA}) = u(g_1, x_0).$$ (3)

Using the intuition that $x_1 - x_0 = \text{WTA}$, and these simple definitions of WTA and WTP, the following relationship is derived:

$$u(g_1, x_0 + \text{WTA} - \text{WTP}(x_0 + \text{WTA})) = u(g_0, x_0 + \text{WTA}) = u(g_1, x_0).$$ (4)

Equality of the first and third expressions implies $x_0 + \text{WTA} - \text{WTP}(x_0 + \text{WTA}) = x_0$, or $\text{WTP}(x_0 + \text{WTA}) = \text{WTA}$. A first-order approximation of $\text{WTP}(x_0 + \text{WTA})$ yields $\text{WTA} \approx \text{WTP} + \text{WTA} \frac{\partial WTP}{\partial y}$ (where $y$ is income). Rewriting this leads to the basic relationship

$$\frac{\partial WTP}{\partial y} \approx 1 - \frac{WTP}{WTA}.$$ (5)

---

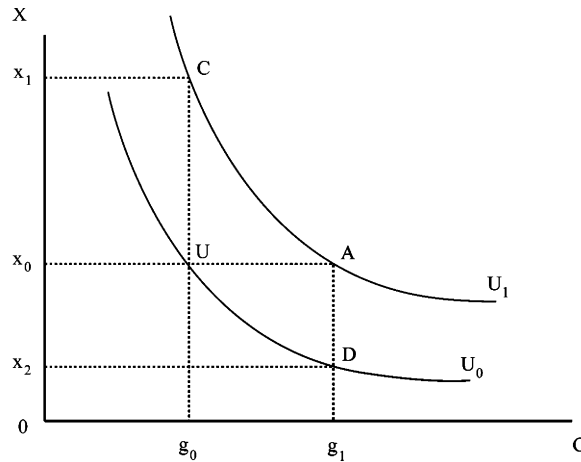[8] The following formulation closely follows Sugden [25].

Fig. 4. Experimental design.

Eq. (5) provides a convenient test of neoclassical theory, as $[1-\text{WTP}/\text{WTA}]$ values can be compared to data on $\frac{\partial WTP}{\partial y}$.[9]

## 2.1. Experimental design

The construct laid out above provides a rich set of predictions associated with the endowment effect and neoclassical theory. As aforementioned, Fig. 3 provides a framework with which to understand the intuition behind the theory of the endowment effect. If subjects instantaneously place an "attachment value" on a good once it becomes part of their endowment, then a kink for Group 1 subjects at point $(g_1, x_0)$ causes the indifference curve for Group 1 subjects to pivot and rotate clockwise. This leads to point $(g_0, x_1)$ being on a lower indifference curve, which induces WTA > WTP. Thus, if the endowment effect is important, it should create a wedge between WTA and WTP values even if endowment-adjusted WTP values are elicited from Group 2 subjects. Alternatively, whether indifference curves are linear or convex to the origin, neoclassical theory predicts equivalence of WTA and endowment-adjusted WTP, but provides an explanation for deviations between WTP and WTA values.

The field experimental treatments attempt to create an environment that allows an investigation into these theories while simultaneously providing a test of whether market experience induces behavior more in line with neoclassical predictions. Fig. 4 illustrates the general experimental design. $U$, $C$, and $A$ are the initial endowment points. Subjects endowed at Point $U$ are asked to state their maximum willingness to pay to acquire $g_1 - g_0$ units of $G$; this is termed WTP. Subjects endowed at Point $C$ are also asked to state their maximum willingness to pay to acquire $g_1 - g_0$ units of $G$; this dollar value represents endowment-adjusted WTP. Finally, subjects endowed at

---

[9] Given that many experimental studies in the literature report WTA values that exceed WTP by factors greater than two (for everyday consumables), using Eq. (5), neoclassical theory predicts that if an agent were instantly endowed with \$100, then she would spend at least \$50 more on the good in question. Most scholars probably would say that this value is implausibly large.

Point $A$ are asked to state their minimum willingness to accept to relinquish $g_1 - g_0$ units of $G$, which represents WTA.

This simple experimental design is used for several different treatment types. Before presenting the various treatments, I should note that I was careful in choosing goods the individual would actually consume rather than put up for trade or sale immediately after the transaction. During the exit interview, more than 90 percent of the subjects stated that they planned to consume the good in question (e.g., keep it for their own collection). The good I used is a variety of Kansas City Royals game ticket stubs dated June 14, 1996. The stubs were issued for admission to the baseball game in which Cal Ripken Jr. broke the world record for consecutive games played. I was able to obtain numerous copies of the ticket stubs because I attended the event. Given that I rarely see this particular piece of memorabilia on the market, market value is difficult to determine, but I have seen a quality piece fetch as much as $40.

The sportscard marketplace is a natural venue where the good's attributes play an integral role in its valuation. This link is perhaps best illustrated by an example. Making use of the grading system of a well-known third-party, Professional Sports Authenticators (PSA), consider the value difference between a PSA-graded (Gem Mint 10) Ken Griffey Jr. 1989 *Upper Deck* baseball card and an identical Griffey card, but one which has slight fraying on one corner—a Mint 9 card. The Gem Mint version of the Griffey Jr. card readily sells for $2700, whereas the identical Griffey Jr. card with slight corner fraying will garner only between $125 and $350. The price difference is mainly driven by perceived scarcity and available substitutes: as of April 14, 2002, approximately two percent of all Griffey cards were graded PSA 10 (44,358 have been graded), while about 30 percent received grades of PSA 9.

To provide variation in quality across the various Ripken ticket stubs, I hand-graded more than 500 ticket stubs to provide enough goods that varied across important attributes. In trying to maintain consistency with PSA guidelines and standards, I graded 180 of the stubs as "near mint" (corresponding to a PSA grade of 7) and 150 as "mint" (corresponding to a PSA grade of 9), with the remaining stubs generally grading below near mint. I then placed them in protective stubholders with the grade clearly marked.

The various treatments, which are depicted in Fig. 4, can be summarized as follows:

### 2.1.1. Treatment A subjects

*WTA1*: Subjects are endowed with a "near mint" (grade 7) ticket stub and asked to state their minimum WTA to sell the stub. Subjects are reminded that the stub is graded near mint and that an examination of typical grading distributions reveals that only about 10 percent of *graded* items actually grade *below* near mint.

*WTA2*: Subjects are endowed with a mint (grade 9) ticket stub and asked to state their minimum WTA to sell the stub. Subjects are reminded that the stub is graded mint and that typical distributions of grading reveal that only about 2 percent of *graded* items actually grade *above* mint.

### 2.1.2. Treatment C subjects

*WTPC1*: Subjects are endowed with the average WTA value from treatment WTA1 ($13.50) and asked to state their maximum WTP for a near mint ticket stub. Subjects are reminded that the

stub is graded near mint and that an examination of typical grading distributions reveals that only about 10 percent of *graded* items actually grade *below* near mint.

*WTPC2*: Subjects are endowed with the average WTA value from treatment WTA2 ($16.75) and asked to state their maximum WTP for a mint ticket stub. Subjects are reminded that the stub is graded mint and that typical distributions of grading reveal that only about 2 percent of *graded* items actually grade *above* mint.

### 2.1.3. Treatment U subjects

*WTPU1*: Subjects are asked to state their maximum WTP for a near mint ticket stub. Subjects are reminded that the stub is graded near mint and that an examination of typical grading distributions reveals that only about 10 percent of *graded* items actually grade *below* near mint.

*WTU2*: Subjects are asked to state their maximum WTP for a mint ticket stub. Subjects are reminded that the stub is graded mint and that typical distributions of grading reveal that only about 2 percent of *graded* items actually grade *above* mint.

By crossing these six treatments with two distinct elicitation institutions (more fully discussed below), I have an experimental design comprised of 12 treatments. These treatments are summarized in Table 1. In Table 1, I also denote the number of subjects included in each treatment—30 subjects—for a total of 360 subjects.

Prior to discussing the experimental procedures, it is worthwhile to discuss the various treatments and hypotheses related to these treatments. Comparing values from WTA1 (WTA2) with WTPU1 (WTPU2) represents the typical experimental examination in the extant literature. Yet a comparison of these WTA and WTP measures of value will not yield a clear-cut delineation between major theories: both neoclassical and alternative theories predict that value divergences may occur across these measures. Many commentators have conceded this point but have argued that the observed disparities are too large to be explained by standard neoclassical arguments. Whether the size of the divergence makes intuitive sense is up for debate, however. As Bateman et al. [1, p. 482] quip: "In the absence of a direct measure of the price flexibility of income, or of

Table 1
Experimental design

| Stub quality | WTA | WTP | WTP (endowment-adjusted) |
| --- | --- | --- | --- |
| Near mint | Denoted WTA1<br>Random $n$th: $n = 30$<br>BDM: $n = 30$ | Denoted WTPU1<br>Random $n$th: $n = 30$<br>BDM: $n = 30$ | Denoted WTPC1<br>Random $n$th: $n = 30$<br>BDM: $n = 30$ |
| Mint | Denoted WTA2<br>Random $n$th: $n = 30$<br>BDM: $n = 30$ | Denoted WTPU2<br>Random $n$th: $n = 30$<br>BDM: $n = 30$ | Denoted WTPC2<br>Random $n$th: $n = 30$<br>BDM: $n = 30$ |

*Notes*: Each cell represents two unique treatments. For example, "Random $n$th: $n = 30$" in row 1, column 1 denotes that one treatment had 30 individuals offering, in a random $n$th price auction, to sell their endowed good—a "near mint" Ripken ticket stub. The treatment is labeled WTA1, for WTA for the "near mint" stub. "BDM" denotes Becker–DeGroot–Marschak [2] discrete auction mechanism. No subject participated in more than one treatment.

the elasticity of substitution, it is difficult to decide whether an observed divergence between WTA and WTP is too large to be compatible with Hicksian theory''.

This represents the first point of departure in this study—gathering WTPC1 and WTPC2 values to delineate between the major theoretical explanations of the value disparity. A comparison of WTPC1 and WTA1 (WTPC2 and WTA2) measures of value provides a relatively clean test of the neoclassical model. As outlined above, by appropriately compensating subjects who are randomly allocated into the WTP treatments, one can roughly place them on the same indifference curve that WTA subjects occupy. While neoclassical theory allows value divergences along different indifference curves, movements along (up and down) identical segments of the same indifference curve should yield equivalent value estimates. Hence, I endow subjects in treatments WTPC1 and WTPC2 with money before they state their WTP values.

The amount of money is determined by the average realized values in the WTA1 and WTA2 treatments—WTPC1 (WTPC2) subjects are endowed with the average WTA value in treatment WTA1 (WTA2). In this case, since income and substitution effects are controlled, neoclassical theory predicts that WTPC1 = WTA1 and WTPC2 = WTA2. Endowment theory predicts that pivoting of the indifference curve renders WTPC1 < WTA1 and WTPC2 < WTA2. These disparate predictions form the basis of the first test of neoclassical versus endowment theory.[10]

This leads to a second test of the value disparity that relates to the substitutability of the two goods with money. As Randall and Stoll [22] note, elasticity of substitution values critically depend on friction within the marketplace: in markets with zero transactions costs the goods possess the same traits as money. In my particular case, the key question relates to the amount of "friction" one confronts in attempting to buy or sell the stub at the market price. Since it is not easy to have strong intuitions about elasticity of substitution values, I used my personal insights gained from the marketplace as a dealer/consumer for more than two decades. Given that even when an economic agent is willing to pay (accept) the going market rate, dealing in thin markets, such as scarce, high-quality sports memorabilia (mint stubs), presents much more friction than dealing in thicker markets (near mint stubs); therefore, I determined that the mint treatments would be associated with more convex indifference curves (I return to this issue in Section 4).

Thus, comparison of data across stubs represents a rough attempt at manipulating the substitutability of the good in question. Economic intuition implies that a movement to mint stubs (WTA2 and WTPC2) from near mint stubs (WTA1 and WTPC2) will result in movement to a higher indifference curve (income effect) that is more convex to the origin (substitution effect). In this case, regardless of the shape of the indifference curves, neoclassical theory predicts that WTA2/WTPC2 = WTA1/WTPC1 = 1.

Alternatively, under popular variants of reference-dependent theory, individuals recognize decision problems as gains or losses relative to a reference point. Thus the individual has a preference ordering over the goods that depends on the current asset position, implying that preferences conditional on one reference point may be different from preferences conditional on another reference point. Individuals that are "loss averse" are expected to have kinked indifference curves at the reference point. Morrison [20] represents such a relationship by rotating

---

[10] I am not the first to compare WTA with endowment-adjusted WTP statements of value (see, for example, [9,11]). Yet to my knowledge this line of research is the first to examine endowment-adjusted and WTP value statements in an actual marketplace.

the indifference curve around a reference point. Assuming that the degree of the pivot is identical at the reference point, Morrison [20] shows that the endowment effect is positively correlated with indifference curve convexity. The intuition behind this argument can be seen in Fig. 3: the kink at point $(g_1, x_0)$ causes the indifference curve to pivot from $U_1$ to $U_1^{E}$, yielding a WTA value of $x_3 - x_0$, which increases as indifference curve convexity increases conditional on similar degrees of pivot at the reference point. Thus, following Morrison's [20] line of argument, endowment theory predicts that WTA2/WTPC2 > WTA1/WTPC1 > 1.[11]

## 2.2. Experimental procedure

The experiment was run on the floor of various sportscard shows in a large southern city. Each participant's experience typically followed four steps: (1) completing a survey, (2) learning the auction rules; (3) considering the bid (offer), and (4) conclusion of the transaction and exit interview. In Step 1, the monitor approached potential subjects entering the trading card show and inquired about their interest in baseball collectibles. If the subject was interested in baseball memorabilia, then the monitor asked the potential subject if he/she would like to fill out a survey that would take about 5 min (the survey is available at http://www.aere. org/journal/index.html). If the individual agreed, the monitor briefly explained that in return for completing the survey the subject would receive their endowment, if applicable. After physically being given the appropriate endowment (when applicable), the subject proceeded to fill out the survey. The monitor worked one-on-one with the participant and no time limit was imposed.

In Step 2, the monitor informed subjects in Treatment types $U$ and $C$ ($A$) that they now had the opportunity to bid (offer) in an auction for the good on the table (their good). After inspecting the good, the monitor thoroughly explained the rules of the applicable allocation institution: either a random $n$th price auction or a Becker–DeGroot–Marschak (BDM) discrete-choice auction.

Following List and Shogren [19], the random $n$th-price auction can be explained in four simple steps: (1) each bidder submits a bid (offer); (2) each bid (offer) is rank-ordered from lowest to highest; (3) the monitor selects a random number ($n$) uniformly distributed between 2 and $Z$ ($Z$ bidders); and (4) in the WTP case, the monitor sells one unit of the good to each of the $(n-1)$ highest bidders at the $n$th-price; in the WTA case, the monitor buys one unit each from the $(n-1)$ lowest bidders and pays the $n$th-lowest bid (http://www.aere.org/journal/index.html contains the experimental instructions for the WTP random $n$th price auction).[12]

As described in Shogren et al. [23], the BDM random pricing mechanism works as follows. If the agent is in the WTP (WTA) treatments, she is deemed a buyer (seller) and asked to determine

---

[11] A final, much more qualified and less revealing test of the substitutability effect revolves around a comparison of WTA and WTP values. Since lower elasticity of substitution values map into larger "price flexibility of income" values, WTA2/WTPU2 *may* be larger than WTA1/WTPU1. This hypothesis must be qualified because under neoclassical arguments the disparity depends on both the ordinary income elasticity of demand for the good and the Allen-Uzawa elasticity of substitution. If the ordinary income elasticity of demand is roughly equivalent for the mint and near mint stubs, then neoclassical theory predicts WTA2/WTPU2 > WTA1/WTPU1. Since, theoretically, substitutability is related to indifference curve convexity (see [20]), endowment theory makes a similar prediction.

[12] The BDM instructions are identical except for the necessary rule changes (see below).

her maximum WTP (minimum WTA) by marking an "X" on a recording sheet listing a discrete price schedule:

|                        | I will buy (sell) | I will not buy (sell) |
|------------------------|-------------------|-----------------------|
| If the price is $ 50.00 | ––––––––––––      | ––––––––––––          |
| .                      | .                 | .                     |
| .                      | .                 | .                     |
| If the price is $ 20.00 | ––––––––––––      | ––––––––––––          |
| If the price is $ 19.50 | ––––––––––––      | ––––––––––––          |
| If the price is $ 19.00 | ––––––––––––      | ––––––––––––          |
| .                      | .                 | .                     |
| .                      | .                 | .                     |
| If the price is $ 0.00 | ––––––––––––      | ––––––––––––          |

The price range for the goods was $0 to $50.00, in steps of $0.50. This range was chosen because it encompassed all of the offers (bids) in the WTA (WTP) random $n$th price auctions for both good types and because it included the highest price I had seen the good bring ($40).

In the WTP random $n$th price auction treatments, after completing the survey, learning the auction rules, and examining the good, each participant privately wrote a bid on the bidding sheet and placed it in an opaque box. The monitor informed the participant that his or her bid would not be opened until after the show and that all bids would be destroyed when the research project was completed. In the WTA random $n$th price auction treatments, after physically giving the subject the good, the subject filled out the survey. He or she proceeded to learn the auction rules and then completed the recording sheet by stating his or her minimum WTA to sell the stub. Each subject worked one-on-one with the monitor and no time limit was imposed on his or her inspection of the stub. In Step 4, the monitor explained that if the participant won the auction, she would be contacted by email or telephone within 3 days. Upon receipt of payment, I would send her the stub (or money). I paid all postage.

In the BDM treatments, everything was identical except the timing of the transaction. After collecting the bidding sheet, one price from the sheet was selected randomly via a draw out of a box that included the 101 slips of potential prices. If a buyer was willing to pay at least the random price, she immediately purchased the stub. If the seller was willing to accept less than or equal to the random price, she immediately sold the stub.

Given that learning might influence whether agents conform to neoclassical theoretic predictions, an added advantage of experimenting within the sportscard marketplace is that individuals bring varying levels of experience to the experiment. For example, while some agents are intense consumers, engaging in numerous trades and transactions in a typical month, others are more casual consumers who execute a transaction quite infrequently. To ensure that sub-populations are similar across treatments, I *randomize* participants into *one* of the 12 treatment types—at the top of each hour I changed treatments, and continued to change treatments until each cell had 30 observations (note that I completed the WTA and the random $n$th price auction treatments first because I needed these values for completion of the WTP endowment-adjusted and BDM treatments).

A few noteworthy aspects of the experimental design merit further consideration. First, no subjects participated in more than one treatment. Second, note that subjects received the stub as payment for completing the survey, and had the stub in their possession while filling out the survey. These two attributes have been found to strengthen significantly the endowment effect. Third, since I am not testing the incentive-compatibility of the two institutions, and I want to avoid any excess noise, I inform the subjects that it is in their best interest to bid (offer) their true value in the auctions. I reinforce this notion via several examples that illustrate the optimal strategy of truthtelling in the two mechanisms.

## 3. Results

To put my findings into perspective, I first review the results of List [17], who controlled for Hicksian income and substitution effects by asking subjects to choose between bundles—in one treatment the subject is endowed with good $A$ and has the option to trade it for good $B$; in a second treatment, a different subject is endowed with good $B$ and has the option to trade it for good $A$. Pooling his data across more than 300 subjects yields Fig. 5, which makes the trade probability a function of trading intensity (trades in a typical month—see question 2 in the survey contained in Appendix A). In support of the received literature, List's [17] field evidence suggests that an inefficiently low number of trades occur for naïve traders, consistent with endowment theory. In contrast, this anomaly is not evident for intense consumers. Fig. 5 reveals that individual behavior converges to the neoclassical prediction as trading rates intensify. This study follows List [17] and focuses on trading intensity and its association with the disparity.

Table 2 presents descriptive statistics for the new WTA/WTP data. The top panel of Table 2 provides data from the random $n$th price auction treatments, whereas the bottom panel summarizes the BDM data. Columns 1–6 display the six distinct treatments used for each auction type. Central tendencies of the variables reported in Table 2 suggest that the subgroups are relatively similar in important demographic characteristics. For example, while there is substantial variation between certain cells, such as average trading experience across WTPU1 (3.8) and
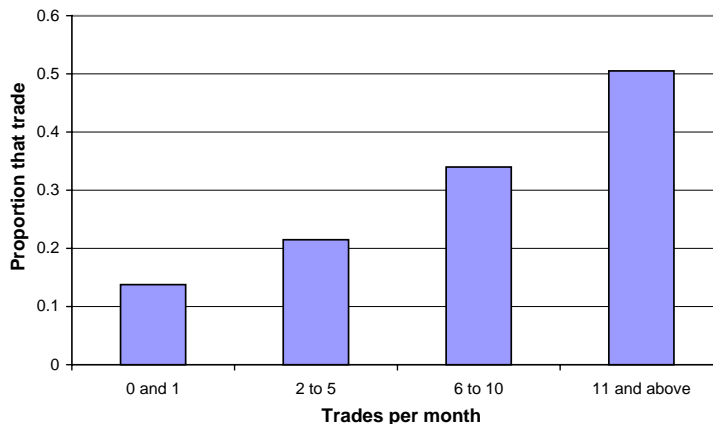


Fig. 5. Summary of List's (2001b) trading results.

Table 2
Sample statistics

| | WTA1 Mean (Std. Dev.) | WTA2 Mean (Std. Dev.) | WTPU1 Mean (Std. Dev.) | WTPU2 Mean (Std. Dev.) | WTPC1 Mean (Std. Dev.) | WTPC2 Mean (Std. Dev.) |
|---|---|---|---|---|---|---|
| *Random nth* | | | | | | |
| Bid or offer | 13.53 | 16.78 | 4.32 | 4.90 | 4.25 | 5.98 |
| | (9.2) | (12.0) | (4.85) | (4.57) | (4.61) | (6.12) |
| Trading experience | 4.3 | 5.9 | 3.8 | 4.9 | 4.3 | 6.0 |
| | (3.4) | (3.9) | (3.7) | (3.2) | (4.6) | (6.1) |
| Years of market | 6.9 | 6.9 | 8.7 | 14.3 | 10.1 | 12.6 |
| experience | (9.2) | (6.4) | (8.3) | (14.3) | (13.5) | (13.1) |
| Income | 3.3 | 3.4 | 3.4 | 4.6 | 4.0 | 3.7 |
| | (1.9) | (1.9) | (1.9) | (1.9) | (1.9) | (2.2) |
| Age | 31.9 | 30.9 | 31.1 | 35.6 | 35.7 | 28.8 |
| | (14.9) | (10.6) | (12.7) | (12.7) | (15.4) | (11.6) |
| Gender | 0.89 | 0.90 | 0.88 | 0.86 | 0.86 | 0.72 |
| | (0.32) | (0.31) | (0.32) | (0.35) | (0.35) | (0.45) |
| Education | 2.8 | 3.4 | 3.2 | 3.9 | 3.5 | 3.5 |
| | (1.7) | (1.7) | (1.5) | (1.4) | (1.8) | (1.7) |
| | | | | | | |
| *BDM* | | | | | | |
| Bid or offer | 11.98 | 13.47 | 5.85 | 6.67 | 5.53 | 7.78 |
| | (11.6) | (12.1) | (5.1) | (8.7) | (5.9) | (6.7) |
| Trading experience | 4.4 | 4.8 | 5.6 | 5.2 | 3.5 | 4.1 |
| | (4.5) | (5.3) | (4.8) | (5.1) | (3.3) | (3.7) |
| Years of market | 10.7 | 9.5 | 8.4 | 9.5 | 8.5 | 7.9 |
| experience | (11.1) | (10.5) | (6.3) | (12.7) | (8.6) | (9.2) |
| Income | 3.9 | 3.5 | 3.8 | 3.5 | 3.5 | 3.4 |
| | (2.4) | (2.1) | (2.3) | (2.1) | (2.0) | (1.9) |
| Age | 33.8 | 30.4 | 32.1 | 30.3 | 29.9 | 34.0 |
| | (14.1) | (13.1) | (14.6) | (14.0) | (11.8) | (14.9) |
| Gender (male) | 0.86 | 0.78 | 0.79 | 0.88 | 0.86 | 0.87 |
| | (0.35) | (0.42) | (0.42) | (0.33) | (0.35) | (0.35) |
| Education | 3.8 | 3.4 | 3.7 | 3.1 | 2.6 | 3.0 |
| | (1.8) | (1.7) | (1.7) | (1.6) | (1.5) | (1.5) |

*Notes*:
1. Thirty subjects are included in each of the 12 treatments, for a total of 360 subjects.
2. *Trading experience* represents the number of trades made in a typical month.
3. *Years of market experience* denotes years that the subject has been active in the market.
4. *Income* denotes categorical variable (1–8): (1) Less than $10,000, (2) $10,000 to $19,999, (3) $20,000 to $29,999, (4) $30,000 to $39,999, (5) $40,000 to $49,999, (6) $50,000 to $74,999, (7) $75,000 to $99,999, (8) $100,000 or over.
5. *Age* denotes actual age in years.
6. *Gender* denotes categorical variable: 0, if female; 1, if male.
7. *Education* denotes categorical variable. (1) Eighth grade or less, (2) high school, (3) 2-year college, (4) other post-high school, (5) 4-year college, (6) graduate school education.

WTPC2 (6.0) and average years of market experience across treatments WTA2 (6.9) and WTPU2 (14.3) in the random $n$th price auctions, in general the data suggest that the randomization procedure was reasonably effective in allocating the 360 subjects randomly into the various treatment cells. This casual observation is verified by a one-way multivariate analysis of variance (MANOVA) test, which compares the multivariate means of the six individual-specific variables. Results from the MANOVA test yield a "$d$" value equal to 0, which suggests that the null hypothesis that the means of each group are the same $n$-dimensional multivariate vector cannot be rejected at the $p < 0.05$ level.

Concerning valuation estimates, I find that, consistent with expectations, sellers have higher asking prices for the mint ticket stub than for the near mint stub: in the random $n$th price auctions the difference is \$3.25 (23.9 percent), while in the BDM auctions the difference is \$1.49 (12.4 percent). Data from the WTP treatments exhibit similar patterns—differences range from 13.8 percent to over 40 percent, with larger differences being associated with the endowment-adjusted WTP treatments. Yet I should note that not all of the data are perfectly organized. For example, economic intuition suggests that WTP values in the endowment-adjusted treatments should be larger than WTP statements in the WTPU treatments.[13] This expectation is realized in the mint stub auctions, but it is not true in the near mint treatments, where average values in WTPU1 across both auction types are larger than average values in WTPC1. While not too much should be made of statistically insignificant differences, the astute reader will notice some similar data anomalies below.

More importantly for our purposes, across both allocation institutions there is a large discrepancy between WTA and each WTP measure of value. Whereas WTA values are around \$13, WTP values tend to be in the \$5 range. Even when subjects are endowment-adjusted, the differences are large. And, all of these value differences are statistically significant at the $p < 0.05$ level using a large-sample $t$-test. In each case, this finding is substantiated when one considers results from a Mann–Whitney rank-sum test of treatment differences. The rank-sum test is a non-parametric test that has a null hypothesis of no treatment effect, or that the two samples are derived from identical populations. These preliminary results suggest that endowment theory does an adequate job of organizing the pooled data.[14]

The data can be pushed a bit further by using the insights gained from Eq. (5): $\partial \text{WTP}/\partial y = 1 - \text{WTP}/\text{WTA}$. Using results from the BDM mint auctions (BDM WTA2 versus WTPU2), I find that $\partial \text{WTP}/\partial Y = 0.505$, which suggests that if an agent was instantly endowed with \$100, he would spend about \$50 on Ripken ticket stubs. Even if this comparative static represented increased spending on sports memorabilia in general, it seems much too large for the average consumer. In sum, therefore, the data indicate that there is a large value disparity in the field.

Another interesting trend in the data is that WTA/WTP discrepancies are somewhat smaller in the BDM auctions than in the random $n$th price auctions. This result is unexpected, as previous laboratory data reported in Shogren et al. [23] suggest that the disparity is larger in BDM auctions.

---

[13] Subjects randomly allocated into the WTP endowment-adjusted treatments received the average offer from the appropriate random $n$th price auction: each WTPC1 subject received \$13.50 and each WTPC2 subject received \$16.75. For consistency, I endowed subjects with these same dollar values whether they were in the BDM or random $n$th auction treatments.

[14] Note that if one assumes preference/value uncertainty, neoclassical theory also provides a viable explanation for these findings. I will return to this issue in the concluding section.

Table 3
Bid/offer sample statistics across consumer types

|  | WTA1 Mean (Std. Dev.) | WTA2 Mean (Std. Dev.) | WTPU1 Mean (Std. Dev.) | WTPU2 Mean (Std. Dev.) | WTPC1 Mean (Std. Dev.) | WTPC2 Mean (Std. Dev.) |
|---|---|---|---|---|---|---|
| **Pooled** | | | | | | |
| Bid or offer | 12.76 (10.4) ($n = 60$) | 15.12 (12.0) ($n = 60$) | 5.08 (4.9) ($n = 60$) | 5.78 (7.0) ($n = 60$) | 4.89 (5.3) ($n = 60$) | 6.88 (6.4) ($n = 60$) |
| **"Light" consumers** | | | | | | |
| Bid or offer | 23.40 (14.7) ($n = 10$) | 22.21 (13.7) ($n = 7$) | 3.34 (5.6) ($n = 13$) | 2.75 (3.8) ($n = 6$) | 4.75 (7.2) ($n = 12$) | 5.06 (9.9) ($n = 12$) |
| **"Casual" consumers** | | | | | | |
| Bid or offer | 16.31 (11.6) ($n = 35$) | 18.11 (14.0) ($n = 28$) | 4.48 (5.7) ($n = 30$) | 6.21 (9.7) ($n = 24$) | 3.50 (4.7) ($n = 34$) | 4.51 (6.6) ($n = 32$) |
| **"Intense" consumers** | | | | | | |
| Bid or offer | 7.78 (5.7) ($n = 25$) | 12.51 (9.4) ($n = 32$) | 5.68 (3.5) ($n = 30$) | 5.50 (4.4) ($n = 36$) | 6.71 (5.6) ($n = 26$) | 9.59 (5.2) ($n = 28$) |
| **"Super-intense" consumers** | | | | | | |
| Bid or offer | 6.71 (5.7) ($n = 12$) | 8.33 (8.1) ($n = 15$) | 5.67 (4.1) ($n = 12$) | 7.17 (5.3) ($n = 12$) | 6.17 (4.1) ($n = 9$) | 7.65 (4.9) ($n = 10$) |

*Notes*: Data are pooled across BDM and random $n$th price auctions in each major treatment cell. "Light" consumers are those that trade 0 times in a typical month; "Casual" consumers are those that trade less than 5 times in a typical month; "Intense" consumers are those that trade 5 or more times in a typical month; "Super-Intense" consumers are those that trade 9 or more times per month. In choosing these thresholds, I considered both the mean number of trades in a typical month (4.7) and its standard deviation (4.2). Thus, Casual consumers are those subjects below the average trading rate; Intense consumers are those subjects above the average trading rate, and Super-Intense consumers are those consumers at least one standard deviation above the average trading rate.

Even though persistent differences exist across the value elicitation cells, the BDM data and the random $n$th price auction data are not significantly different statistically, thus I pool the data below.

Table 3 presents WTA and WTP measures of value across five distinct subsamples of the data. Figures in the first row are derived from the pooled sample, and highlight the value disparity that is evident in the data. In the auctions for the near mint stub, the WTA/WTP ratios are 2.5 (WTA1 versus WTPU1) and 2.6 (WTA1 versus WTPC1). The ratios are comparable across auctions for the mint stub—2.6 (WTA2 versus WTPU2) and 2.1 (WTA2 versus WTPC2).[15] While these ratios

---

[15] Recall that both neoclassical and endowment theory predict WTA2/WTPU2 > WTA1/WTPU1. While that result is observed here and in spots below, it is not consistently significant at conventional levels.

are significantly different from one using parametric and non-parametric statistical tests, suggesting that there is a value disparity evident in the field data, there is no evidence that WTA2/WTPC2 > WTA1/WTPC1. As we will see, this null result will continue to arise for finer classifications of the data, and suggests that the value disparity is not larger for goods that have fewer available substitutes.

The remaining rows in Table 3 split the data by consumer type. I define "light" consumers as those subjects that stated they trade zero times in a typical month; "casual" consumers execute fewer than five trades in a typical month; "intense" consumers execute five or more trades per month; and "super-intense" consumers trade at a rate of nine or more times per month. While these thresholds may appear somewhat unscientific I have attempted to delineate by first and second moments of the trading variable. Because the average number of trades in a typical month is 4.7, I label subjects below five trades per month as casual and those above five trades per month as intense. And given that the standard deviation of the number of monthly trades is 4.2, I label those subjects that execute nine or more trades ($\approx$ mean + one standard deviation) as super-intense. The general pattern of results does not rely on these thresholds, as will be illustrated in the regression analysis summarized below.

Even though the subsamples of light consumers are small, the data strongly suggest that WTP is *much* less than compensation demanded. Results of both a small sample *t*-test and a Mann–Whitney rank-sum test of treatment differences suggest that WTA is greater than WTP for every applicable treatment comparison at the $p < 0.05$ level. For light consumers, the ratio of WTA/WTP ranges from 4.4 to 8, which is consistent with previous experimental studies, but at odds with conventional economic theory. Using Eq. (5), these ratios suggest that if an agent were instantly endowed with $100, he/she would spend approximately $80 on Ripken ticket stubs. I suspect that even the most ardent of neoclassical supporters would find this figure implausibly high. Yet note again that these data provide no evidence that WTA2/WTPC2 > WTA1/WTPC1.

Overall, these results are mirrored in the casual consumer subsample. While data from the casual consumers indicate a lesser value disparity across the various treatment cells, in each case the discrepancies are significantly different from zero at conventional levels. And, use of Eq. (5) again provides far-fetched $\partial$WTP/$\partial y$ values.

Data contained in the lower rows of Table 3 paint a much different picture, however. In the intense consumer group data, using small-sample *t*-tests I find that WTA2 and WTPU2, as well as WTA1 and WTPU1, are significantly different at conventional levels using a one-sided alternative (mint stubs: $t = 4.0$ and near mint stubs: $t = 1.68$). However, comparing WTA with *endowment-adjusted* WTP values yields value ratios of 1.16 and 1.30, which are never significantly different from one using small-sample *t*-tests (mint stubs: $t = 1.46$ and near mint stubs: $t = 0.68$).

These findings are in accord with neoclassical theory. Rejection of the homogeneous null in the first set of results—WTA versus WTP—may well be due to a comparison between value estimates drawn from different indifference curves. The latter set of null results, which arise from a comparison of WTA and endowment-adjusted WTP, suggests that indifference curves are reversible. This finding is consonant with neoclassical theory, which predicts that no disparity should exist in such a setting. Yet it is important to note that if one takes the observed point estimates literally, WTA/WTP ratios still provide seemingly large $\partial$WTP/$\partial y$ values.

Even though these results begin to suggest that neoclassical theory holds promise in properly organizing the data, empirical results for the super-intense consumer group make an even stronger
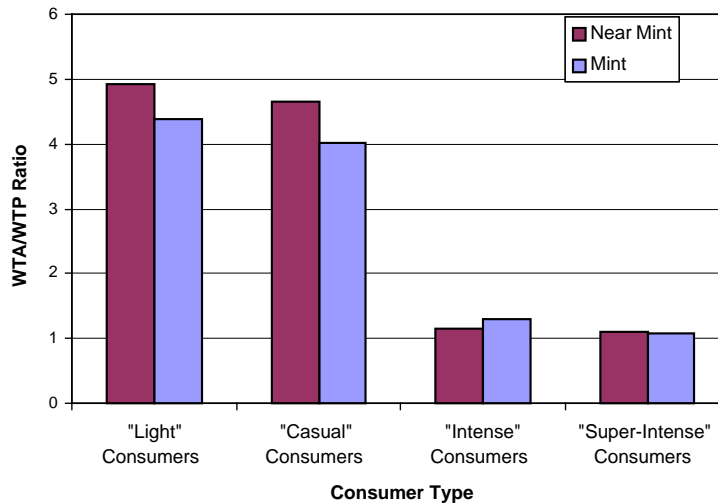
Fig. 6. Summary of WTA/WTP ratios.

case. Here, I find that neither parametric nor non-parametric statistical tests can reject the equality null in *any* of the treatment comparisons; therefore I also cannot reject the null that WTA2/WTPC2 = WTA1/WTPC1 = 1. Furthermore, computed WTA/WTP ratios of approximately 1.15 provide more plausible $\partial WTP/\partial y$ values: a \$100 increase in income leads to approximately \$14 more being spent on Ripken ticket stubs.

But how plausible is this estimated income effect? Running the risk of making too much of a few point estimates rather than relying on inference gained from the statistical tests, I will conduct some back-of-the-envelope calculations to shed light on this issue. Average income levels in the sample are roughly \$30,000 ($\approx 3.5$ on the survey) and average WTP values for near mint and mint stubs are about \$5 and \$6. Assume now that the average person's income increases by \$100 to \$30,100. If WTP increased by the same proportion to income, it would increase by about \$0.02 to \$5.02 or \$6.02. To get the expected increase in WTP (\$14), one would need to assume that WTP increases with extremely large powers of income. This illustrates that when WTP is only a small fraction of income, neoclassical theory predicts very small divergences, even smaller than what the point estimates herein suggest.

To provide a sense of the relationship between the value disparity and market intensity, I provide Fig. 6. Similar to Fig. 5, Fig. 6 reveals that individual behavior converges to the neoclassical prediction as trading experience intensifies. These empirical results supplement List's [17] trading results summarized in Fig. 5. Yet, with these new auction data one can more meaningfully understand what is behind the data patterns observed in Fig. 6.

The statistical summary presented in Table 3 implies that less casual subjects state much lower WTA figures than casual subjects, which induces the value divergence to lessen: super-intense consumers' WTA is approximately \$17 less than light consumers' WTA, while their WTP is generally only \$2–\$4 greater. To provide further insights into the convergence process, I estimate the following bid/offer model:

$$BO = \alpha + \beta' X + \varepsilon, \tag{6}$$

where *BO* is subject *i*'s bid or offer, *X* includes subject-specific variables that may affect the bid or offer, *ε* is the well-behaved error component, and *α* and *β* are parameters to be estimated. Variables in *X* are listed in Table 2, and include market intensity, number of years in the market, annual income, age, gender, and two dichotomous treatment variables: (i) denoting whether the subject was bidding/offering for the mint versus near mint stub and (ii) whether the subject was in the endowment-adjusted versus WTP treatment. Estimation of Eq. (6), therefore, does not represent a means to pit neoclassical theory against endowment theory; rather, it provides insights into the factors that influence the bid/offer level.

Before discussing the empirical estimates, a few important points should be mentioned. First, to account for censoring (zero bids in the WTP treatments), in the bid models I present estimates from both ordinary least squares and Tobit models. Second, since there were signs of non-spherical disturbances, I use White's robust standard errors. Third, because there is evidence that learning may be non-linear, I include models that have market intensity entering non-linearly. Fourth, since income is used as a control variable in the main specifications I do not allow interactions or introduce it as seven income dummies. Fifth, due to multicollinearity I exclude education from the main specifications but note that results are robust to its inclusion.

Summary estimates of Eq. (6) are presented in Table 4. Columns 1 and 2 in Table 4 provide insights into whether market intensity is correlated with WTA statements of value. Empirical results strongly support the unconditional findings in Table 3. For example, in the linear learning model, the coefficient estimate of intensity is significantly different from zero at the $p < 0.05$ level and implies that a one standard deviation increase in intensity ($\approx 4.2$) is correlated with a $4.28 decrease in WTA. Coefficients in the non-linear learning specification also suggest that market intensity is negatively correlated with compensation demanded, yet here the coefficients suggest that there are diminishing effects of intensity. The majority of other coefficient estimates are not significantly different from zero at conventional levels, except for the dummy variable *Mint*. In this case, both coefficient estimates are significant at the $p < 0.05$ level and indicate that subjects demanded about $4 more for mint stubs than they demanded for near mint stubs.

Columns 3–6 contain empirical estimates from the WTP specifications. In the Tobit models, I present marginal effects computed at the sample means. While market intensity again enters in a statistically significant manner (with opposite sign compared to WTA), its marginal influence in absolute value is much lower than comparable estimates in the WTA specifications. For example, in the OLS linear model presented in column 3, an estimate of $0.21 suggests that WTP values are $0.84 larger for each one standard deviation increase in market intensity. This estimate represents about 20 percent of the marginal effect estimate in the WTA specifications. This result corresponds well with findings in Table 3, and further suggests that the convergence of values tends to occur for more intense subjects because of lower WTA values rather than because of significant increases in WTP. Consistent with the WTA specifications, there is some evidence that learning dissipates for more intense consumers. In this case, the learning curve reaches its peak at approximately 20 trades in a typical month, which is on the boundary of the sample (only one WTP subject stated he executed more than 20 trades in a typical month).[16]

---

[16] While the results herein do not discern between treatment and selection effects, List [17] presents evidence that suggests that both play a role.

Table 4
Estimation results

| | WTA (OLS) | WTA (OLS) | WTP (OLS) | WTP (Tobit) | WTP (OLS) | WTP (Tobit) |
|---|---|---|---|---|---|---|
| Intensity | −1.07** | −2.04** | 0.21** | 0.17** | 0.39** | 0.49** |
| | (0.28) | (0.65) | (0.08) | (0.07) | (0.20) | (0.17) |
| Intensity$^2$ | — | 0.07* | — | — | −0.01 | −0.01** |
| | | (0.04) | | | (0.09) | (0.007) |
| Mint | 4.1** | 3.9** | 1.08 | 0.59 | 1.09 | 0.75 |
| | (2.0) | (2.0) | (0.83) | (0.76) | (0.83) | (0.75) |
| Endowment-adjusted | — | — | 0.24 | 0.50 | 0.25 | 0.37 |
| | | | (0.81) | (0.76) | (0.31) | (0.73) |
| Years of market experience | 0.12 | 0.10 | −0.07* | −0.08** | −0.07* | −0.09** |
| | (0.09) | (0.08) | (0.04) | (0.03) | (0.04) | (0.03) |
| Income | −0.33 | −0.35 | −0.23 | −0.03 | −0.24 | −0.03 |
| | (0.50) | (0.49) | (0.21) | (0.22) | (0.22) | (0.21) |
| Gender (male) | −0.67 | −0.52 | 1.43 | 1.62* | 1.44 | 1.83** |
| | (3.38) | (3.27) | (1.02) | (0.87) | (1.0) | (0.81) |
| Age | 0.05 | 0.08 | 0.03 | 0.01 | 0.03 | 0.01 |
| | (0.09) | (0.09) | (0.04) | (0.03) | (0.04) | (0.03) |
| Constant | 12.2** | 13.65** | 2.32 | 1.21 | 1.96 | −0.25 |
| | (5.0) | (4.1) | (2.02) | (1.83) | (2.01) | (1.8) |
| N | 112 | 112 | 215 | 215 | 215 | 215 |
| $R^2$ | 0.17 | 0.19 | 0.04 | — | 0.05 | — |

*Notes*:
1. Dependent variable is subject's offer or bid.
2. Mint equals 1 if subject was in the treatment with "mint" stubs; 0, otherwise. Endowment-adjusted equals 1 if subject was in an endowment-adjusted treatment; 0, otherwise. Gender = 1, if male; 0, otherwise.
3. Tobit coefficients are marginal effects computed at the sample means. Standard errors are in parentheses beside coefficient estimates.
4. Sample sizes may not match with sample sizes in Table 1 due to some respondents not filling in the entire questionnaire (e.g., not including income responses). Reported results omit these observations. If means are used to fill in the missing observations, results are not quantitatively or qualitatively different from the results reported.
5. **(*) Denotes coefficient estimate is significant at the $p < 0.05$ (0.10) level.

Other WTP coefficient estimates are generally insignificant, but there are spots of significance.[17] Interestingly, years of market experience gains statistical significance at the $p < 0.10$ level in each specification, but appears to be *attenuating* the bidding level. And, I should note that this effect has economic significance: one extra year of experience maps into roughly a \$0.10 decrease in WTP, and a one standard deviation increase in years of market experience yields a \$1.00 decrease

---

[17] Of some interest is the fact that the income response coefficient is found to be negative and insignificant in every specification. This is curious given that the raw data provided very large empirical estimates of $\partial WTP/\partial y$. To explore this issue further, I allowed the income effect to vary across inexperienced and experienced subjects by augmenting Eq. (6) with an interaction term—intensity $*$ income. In these estimations, there is evidence that the income effect is positive among inexperienced consumers and that the interaction term is negative and significant at the $p < 0.10$ level.

in WTP. This result represents a potentially important anomaly since, *ceteris paribus*, lower WTP values map into larger WTA/WTP value differences. As an exploratory probe, I ran a new set of empirical specifications that replaced the intensity and years of market experience variables with an interaction term—intensity ∗ years of market experience. The interaction term could be thought of as a measure of the stock of market experience. In these models, the coefficient of the interaction term is generally positive and statistically significant at conventional levels, suggesting that the overall stock of market experience increases WTP values, leading to smaller WTA/WTP disparities.[18]

## 4. Potential limitations

While the above results provide insights that may be interesting to economists and policymakers, it is important to recognize the inherent limitations in the experimental design. First, recall that *all* subjects randomly allocated into the WTP endowment-adjusted treatments received the average offer from the appropriate random $n$th price auction: each WTPC1 subject received $13.50 and each WTPC2 subject received $16.75. Considering the mean WTA values across subject types (see Table 3), some subjects were likely over-endowed (intense and super-intense consumers by nearly $7 in some cases) while others were likely under-endowed (light consumers by nearly $10 in some cases). This design decision, therefore, may have compromising effects if subjects' WTP is highly responsive to this change in wealth.

To address this potential shortcoming, I returned to the same marketplace and ran three ancillary WTP treatments (denoted WTP$6.71, WTP$13.50, and WTP$23.40) with 30 subjects each to examine if this particular design choice unduly compromises my results. The three WTP treatments are identical in every way possible to BDM treatment WTPC1, except that in WTP$X I endow every subject with $X, where $X is $6.71, $13.50, or $23.40. Insights from these three new treatments provide a sense of whether the over- and under-endowment influenced outcomes in the experiment.

Rather than prolong the discussion, I only summarize the data and make a broader set of results available at http://www.aere.org/journal/index.html. I am able to report several insights from these ancillary data. First, in treatment WTP$13.50, I find that the average bid of $5.11 is in the ballpark of the average bid in BDM treatment WTPC1 ($5.53; see Table 2). This provides a nice validity check. Second, in the overall data, the endowment difference does have an influence on WTP statements of value, but not a statistically significant one at conventional levels. For example, in treatment WTP$6.71 (WTP$23.40) the average bid was $4.90 ($6.31) for the stub. It is important to highlight that the average bid in treatment WTP$23.40 is roughly 25 percent greater than the average bid in treatments WTP$6.31 and WTP$13.50; yet it is not statistically different from either of these average bid estimates using a parametric *t*-test or a Wilcoxon non-parametric

---

[18] There is also evidence that men bid higher than women. The OLS estimates are relatively noisy, but in the Tobit specifications these effects are significant at conventional levels and imply that men bid about $1.75 more than women. This result, coupled with the fact that men provide lower WTA values, suggests that men have lower WTA/WTP disparities than women. This finding is potentially of interest, but due to small sample sizes of women, I find the evidence premature to make a statistical statement concerning the role of gender on the WTA/WTP disparity. This is an interesting area for future research.

test of distribution location. Nevertheless, Type 2 error is a distinct possibility in this case so it is important to recognize that there are differences in value statements across treatments WTP$6.71 and WTP$23.40.

Third, and most importantly, when I allocate these data into ''individual experience cells'' in a fashion similar to Table 3 and compare them to WTA data, the qualitative results from Table 3 do not change: inexperienced consumers exhibit a large value disparity, while the disparity is significantly attenuated among experienced consumers. This holds even when I limit the empirical examination to the WTP$6.71 (WTP$23.40) data for experienced (inexperienced) consumers. It appears that the $7 or $10 difference in endowed resources does not unduly influence the major results, which makes sense given that the convergence of WTA/WTP values occurs for more intense subjects because of lower WTA values rather than because of significant increases in WTP.

A second potential limitation that should be discussed concerns the measure of substitution across the two good types. Since it is not easy to have strong intuitions about elasticity of substitution values, I used my knowledge of the marketplace to generate two goods that would vary along levels of friction in exchange in the marketplace. In theory, one can use Eq. (5) directly to measure substitution effects. Empirical results in Table 4 provide initial insights concerning this issue and suggest that if one restricts income to have an isomorphic effect across good types, a statistically insignificant responsiveness of WTP to income changes is a general outcome. Indeed, as discussed above, the point estimate of the income coefficient is found to be negative and insignificant in every specification. To further explore this issue, I augmented the WTP empirical specifications in Table 4 to include an interaction term that allowed mint and near mint stubs to have different income effects. In each specification I find that the estimated income effects are not different from one another, or different from zero, at conventional levels. Thus the empirical evidence does not support the assumption that mint stubs are characterized by more convex indifference curves.

## 5. Concluding remarks

Substantial-experimental evidence indicates that a non-trivial gap exists between WTA and WTP statements of value. These findings have induced even the most ardent supporters of neoclassical theory to doubt the validity of certain neoclassical modeling assumptions. Now that experimental economists have obtained several unique insights on the value disparity, however, it is important to examine whether these experimental findings generalize to the types of environments that theorists' models purport to explain. Field experiments are a natural next step, as they provide a useful intermediary between the sterile lab environment and the vagaries of completely uncontrolled field data.

In this study, I examined data gathered in a naturally occurring marketplace. Examining data from 12 experimental treatments yields several unique insights. Most importantly, I find evidence that intensity of market experience and the WTA/WTP disparity are negatively related. In the limit, intense consumers exhibit a negligible value disparity.

A few clear next steps in this research agenda are to bolster sample sizes of the experienced consumer subset and examine whether these results spill over to the domain of public goods. This latter extension is necessary since the disparity has such profound impacts on applied welfare

issues such as non-market valuation and, more broadly, cost/benefit analysis. Besides these obvious extensions, it makes sense to investigate the underlying learning mechanism at work. For example, why do casual consumers display a large value disparity while intense consumers exhibit an insignificant one? Is it simply that intense consumers are experienced arbitrageurs and via learning they have become rational neoclassical utility maximizing agents? If so, what have they learned? Is it to treat the good leaving their endowment as an opportunity cost rather than a loss? Or are they more certain of their preferences because they are intense consumers? Could lower transactions costs play a role?

While I find answers to these provocative questions quite important, I defer a more extensive discussion until another occasion. Yet I would be remiss not to discuss a few of the possible explanations briefly. In terms of neoclassical explanations, preference/value uncertainty seems to be the logical place to start. Indeed, the overall pattern of observed results herein bears some consistency with Kolstad and Guzman's [16] theoretic bidding model in which subjects have private values, but know them imperfectly. In their model, an increase in the level of individual information decreases the gap between WTP and WTA. In addition, as the cost of information increases, the expected gap between WTP and WTA increases. In my auction treatments, if casual consumers have little idea about the market price of the ticket stub, they may be reluctant to sell it quickly, and therefore might place a large offer price in the auction. Likewise, casual consumers may not attach a high value to the good when bidding in the auction because they do not want to overpay for the good. In this sense, if intense consumers have a better knowledge of values, then the *overall* pattern of results observed in the auction data is consistent with neoclassical theory. If more recent trading experience provides important market information needed to overcome the value disparity (and years of experience are not as valuable), then the econometric estimates also provide support for the information asymmetry conjecture.[19]

The spirit of the auction results is also consistent with Kling et al. [10], who have recently obtained promising experimental results suggesting that a neoclassical option-based theory may have some predictive power. Using a good that is commonly traded in the sportscard market, they find that subjects state higher WTP values when they believe it would be difficult to purchase the good at a later date. And, subjects state lower WTP bid values if they believe that it would be difficult to reverse the purchasing transaction at a later date (i.e., sell the good later if they purchase it in the experiment). Consistent with these insights, their data also suggest that WTA values decrease in the difficulty of delay and increase in the difficulty of reversing the transaction. With certain reasonable assumptions, their comparative static findings are consistent with the results herein.

Besides more fully exploring these neoclassical-based explanations, I believe psychological explanations can also provide insights into the effect of marketlike experience on the value disparity. Empirical results from a recent set of market experiments reported in List [18] suggest that agents have learned to treat goods leaving their endowment as an opportunity cost rather than a loss. Thus, while psychological effects have been extremely popular in explaining the endowment effect anomaly, they may also have some explanatory power regarding the attenuation of the anomaly.

---

[19] Thanks to Don Fullerton for urging me to explore this issue.

# References

[1] I. Bateman, A. Munro, B. Rhodes, C. Starmer, R. Sugden, A theory of reference-dependent preferences, Quart. J. Econ. 112 (1997) 479–505.

[2] G. Becker, M. DeGroot, J. Marschak, Measuring utility by a single response sequential method, Behavioral Sci. 9 (1964) 226–236.

[3] D. Brookshire, D. Coursey, Measuring the value of a public good: an empirical comparison of elicitation procedures, Amer. Econ. Rev. 77 (1987) 554–566.

[4] D. Brookshire, A. Randall, J.R. Stoll, Value increments and decrements in natural resource service flows, Amer. J. Agr. Econ. 62 (1980) 185–193.

[5] D. Coursey, J. Hovis, W. Schulze, The disparity between willingness to accept and willingness to pay measures of value, Quart. J. Econ. 102 (1987) 679–690.

[6] W.M. Hanemann, Willingness to pay and willingness to accept: how much can they differ, Amer. Econ. Rev. 81 (1991) 635–647.

[7] W.M. Hanemann, The economic theory of WTP and WTA, in: I.J. Bateman, K.G. Willis (Eds.), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries, Oxford University Press, Oxford, 1999.

[8] J.K. Horowitz, K.E. McConnell, A review of WTA/WTP studies, J. Environ. Econ. Manage. 44 (2002) 426–447.

[9] D. Kahneman, J.L. Knetsch, R.H. Thaler, Experimental tests of the endowment effect and the coase theorem, J. Polit. Economy 98 (1990) 1325–1348.

[10] C. Kling, J.A. List, J. Zhao, The WTP/WTA disparity: have we been observing dynamic values but interpreting them as static?, Working Paper, Iowa State University, 2002.

[11] J.L. Knetsch, The endowment effect and evidence of nonreversible indifference curves, Amer. Econ. Rev. 79 (1989) 1277–1284.

[12] J.L. Knetsch, Environmental-policy implications of disparities between willingness to pay and compensation demanded measures of values, J. Environ. Econ. Manage. 18 (1990) 227–237.

[13] J.L. Knetsch, J.A. Sinden, The persistence of evaluation disparities, Quart. J. Econ. 102 (1987) 691–695.

[14] J.L. Knetsch, F.F. Tang, R.H. Thaler, The endowment effect and repeated market trials: is the Vickrey auction demand revealing?, Exper. Econ. 4 (2001) 257–269.

[15] P. Knez, V.L. Smith, A. Williams, Individual rationality, market rationality, and value estimation, Amer. Econ. Rev. 75 (1985) 397–402.

[16] C. Kolstad, R. Guzman, Information and the divergence between willingness to accept and willingness to pay, J. Environ. Econ. Manage. 38 (1999) 66–80.

[17] J.A. List, Does market experience eliminate market anomalies?, Quart. J. Econ. 118 (2003) 41–71.

[18] J.A. List, Neoclassical theory versus prospect theory: evidence from the marketplace, NBER Working Paper ♯9763, 2003.

[19] J.A. List, J.F. Shogren, The deadweight loss of Christmas: comment, Amer. Econ. Rev. 88 (1998) 1350–1355.

[20] G.C. Morrison, Resolving differences in willingness to pay and willingness to accept: comment, Amer. Econ. Rev. 87 (1997) 236–240.

[21] C. Plott, Laboratory experiments in economics: the implication of posted-price institutions, Science 232 (1986) 732–738.

[22] A. Randall, J.R. Stoll, Consumer's surplus in commodity space, Amer. Econ. Rev. 71 (1980) 449–457.

[23] J. Shogren, S. Cho, C. Koo, J. List, C. Park, P. Polo, R. Wilhelmi, Auction institutions and the measurement of WTP and WTA, Resource Energy Econ. 23 (2001) 97–109.

[24] V.L. Smith, Microeconomic systems as an experimental science, Amer. Econ. Rev. 72 (1982) 589–597.

[25] R. Sugden, Alternatives to the neoclassical theory of choice, in: I.J. Bateman, K.G. Willis (Eds.), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries, Oxford University Press, Oxford, 1999.

[26] J. Zhao, C. Kling, Real options and the WTA/WTP disparity, Working Paper, Iowa State University, 1999.