

USING FIELD EXPERIMENTS IN ENVIRONMENTAL AND RESOURCE ECONOMICS[†]

Does Market Experience Eliminate Market Anomalies? The Case of Exogenous Market Experience

By JOHN A. LIST*

I. Introduction

Nearly three decades ago, a substantial body of empirical evidence began to develop that provided evidence that willingness to pay (WTP) and willingness to accept (WTA) measures of value were quite different. Typically, these studies found that WTA measures were substantially greater than WTP measures for the same commodity (see, e.g., Judd Hammack and Gardner M. Brown 1974). The initial reaction of many economists was to argue that the results were a survey artifact and that WTA estimates were unreliable and should not be treated seriously.

The difference between WTP and WTA value measures, however, has proven to be robust across a wide variety of experimental protocols and goods—from neoteric goods, such as irradiated sandwiches, to common goods, such as chocolate bars and coffee mugs. Whether preferences are defined over consumption levels or changes in consumption has serious implications for the discipline of economics. In a normative sense, the basic independence assumption (that preferences are orthogonal to current entitlements), which is used in most theoretical and applied economic models to assess the operation of markets, is directly refuted. In a positive sense, the disparity has considerable relevance. For example, these findings call into question commonly held interpretations of indifference curves, make cost-benefit analysis illegitimate,

and change the procedure necessary to resolve damage disputes.

While the major theory to explain such differences in measures of value is prospect theory, where gains are more heavily weighted than losses, one can potentially reconcile the value disparity with neoclassical theory. Consider Alan Randall and John R. Stoll's (1980) "price flexibility of income," which 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. Thus, for low elasticity of substitution values, which may adequately describe many non-marketed goods and services, the "price flexibility of income" is large, suggesting that the WTA/WTP ratio is also large. Many commentators concur with this point but have argued that the observed disparities are too large to be explained by standard neoclassical arguments.

This uncertainty about the applicability of Hicksian theory has motivated the experimental work of, for example, Jack L. Knetsch (1989), Ian Bateman et al. (1997), and John A. List (2003, 2004a), who ask subjects to (implicitly) rank two goods. The ranking, by definition, controls for all Hicksian income and substitution effects. For my purposes, the results in List (2003, 2004a) are the most relevant, since they show that market experience has an important influence on the shape and magnitude of the disparity. In particular, in his pair of studies List (2003, 2004a) reports that market experience measured by activity in naturally occurring trading markets for pins, sports cards and sports memorabilia lead to behavior that is not significantly different from neoclassical expectations.

Even though the thrust of List's experience results have been broadly replicated in both the lab and the field (see, e.g., Lei Feng and Mark S. Seasholes 2005; Deborah A. Kermer et al. 2006;

[†]*Discussants:* Lint Barrage, Yale University; Hunt Allcott, Massachusetts Institute of Technology and New York University; Margaret McConnell, Harvard University; Jeffrey Carpenter, Middlebury College.

*Department of Economics, The University of Chicago and NBER, 126 East 59th Street, Chicago, IL 60637 (e-mail: jlist@uchicago.edu). Thanks to Lint Barrage for excellent comments. Alec Brandon provided able research assistance.

Ravi Dhar and Ning Zhu 2006; Alistair Munro and Yannick Ferreira De Sousa 2008; Robin Greenwood and Stefan Nagel 2009; Simon Gächter et al. 2009; Hyuk Choe and Yunsong Eom 2009; Dirk Engelmann and Guillaume Hollard 2010; Amit Seru, Tyler Shumway, and Noah Stoffman 2010), one nagging issue is that individual market experience is endogenous; that is while List attempts to econometrically parse treatment (market experience) from selection, his results rely on his modeling assumptions. This paper attempts to rectify this issue by making market experience exogenous. I do so by returning to the sports card market and exogenously inducing a random sample of subjects to enter the sports card market and make trades (i.e., gain experience). I then follow the subjects over a six month period and at the end of this period test for market experience effects via a field experiment.

The data for this study are gathered in three distinct experiments within a six month period—September of 2007 and February of 2008. By using several measures to ensure that subjects will return for the follow-up experiments, I am able to avoid the attrition problems present in other longitudinal studies. In this way, I am able to induce experience in a natural way in a natural environment.

My results can be simply stated: there is evidence that market experience eliminates the value disparity. For example, subjects who were induced to enter the market after the September experiment were significantly more willing to trade their wares in the second and third experiments than those who were not induced into the market. This result suggests that market experience alone can attenuate an important market anomaly. Interestingly, although the results are broadly consistent, I find that more market experience is needed to overcome the market anomaly than, for example, the lab results in Engelmann and Hollard (2010) suggest. This is an important area for future research, as it might be the case that when one focuses on a specific task or treatment in the lab the attention paid to that one dimension takes on a greater meaning and import than if one does the same in environments with multiple stimuli.

II. Experimental Design

The initial experiment was run at a sports card show in a suburb of Chicago, Illinois

in September 2007. The design of the actual experiment followed List (2003) closely. 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*. Since subjects are allocated to one of the two treatments randomly, if p percent trade good *A* for good *B*, then $(1 - p)$ percent should trade good *B* for good *A*. That is, for equal valued goods, approximately 50 percent of the subjects should trade their good. Alternatively, if an endowment effect exists, considerably fewer than 50 percent of the subjects should trade their good. In Knetsch (1989), the evidence in favor of the endowment effect is sharp: 89 percent of those originally endowed with a mug chose to keep the mug, and 90 percent of those endowed with a chocolate bar decided to keep the chocolate bar. List (2003) found similar results for his highly inexperienced subjects.

What differed from List (2003) in the September experiment is the recruiting process. In Step 1, I approached potential subjects entering the sports card show and inquired about their interest in filling out a survey that would take about five minutes (the survey was similar to List 2003 and obtained information on market experience). If the individual agreed, I briefly explained that in return for completing the survey the subject would receive good *A* (or good *B*), where the goods were sports memorabilia. After physically receiving either good *A* or *B*, the subject proceeded to fill out the survey (in this way, in the parlance of Glenn W. Harrison and List 2004, I am executing a framed field experiment).

In Step 2, I first ascertained whether the subject could be considered very inexperienced, that is, had zero executed trades in a typical month. If the subject fell into this category, I asked the subject if he would be available for at most three follow-up experiments between September and February. If the individual agreed, I asked him to sign an affidavit in which he promised to participate again in future experiments. If the individual had executed previous trades in the market, disagreed with being available for a follow-up experiment, or would not sign the affidavit, then he was placed in an unrelated experiment.

For those who signed the affidavit to participate in a future experiment, I retrieved the other good from under the table and informed the subject that he had the opportunity to trade good

A for good B, or vice versa. I allowed the subject to inspect both goods, after which the subject either consummated a trade or kept the original good. In Step 3 I randomly allocated subjects into one of two treatment cells: *experience* and *noexperience*. The delineation was changed at the top of each hour, so subjects' treatment type was determined exogenously based on the time they visited my dealer's table. For the *experience* subjects, I provided a "parting gift" of approximately \$25 worth of sports cards and memorabilia in Step 4 of the experiment.

In each subject's gift bundle, I included several items designed to engage the subject in the marketplace; in theory, the items could yield several dozen trades. The monitor informed subjects in this treatment group that the gifts were theirs to keep, and they could sell or trade the gifts in the marketplace or take them home. The monitor stressed that dealers at the show were interested in the goods and that the goods had a book value of approximately \$25. Furthermore, in an effort to ensure that the local dealers would have an interest in buying and trading with these subjects, prior to the show the monitor discussed various items with the dealers attending the show to determine the appropriate composition of the gift packages. In practice, to motivate these agents to engage in buying, selling, and trading in the live marketplace, I gave them a lottery ticket for every trade that they executed (local dealers gave receipts for verification). The lottery took place after the final experiment in February, and the winner received \$500.

Treatment *noexperience* was identical to the *experience* treatment except that in Step 4 the monitor closed the experiment by thanking the subject for his or her participation, and no "parting gift" was provided.

To gather the experimental evidence necessary to analyze the learning issue, I returned to the same sports card show in early December and ran an identical experiment using the same subject pool, identical experimental procedures, but different goods. To recruit the same subjects, I telephoned and/or e-mailed all of the subjects who participated in the first September sports card show experiment. Again, to close the experiment I gave the *experience* subjects a parting gift of approximately \$25 worth of sports cards and memorabilia, and they were incented to trade via lottery tickets. In total, all 29 people earned at least one lottery ticket.

To complete the experimental design, I returned to the same sports card show in February and ran an identical experiment using the same subject pool, identical experimental procedures, but different goods. To recruit the same subjects, I again telephoned and/or e-mailed all of the subjects who participated in the first September sports card show experiment.

A few noteworthy aspects of the design should be mentioned before I move to the results summary. First, the type of good (A or B) was changed at the top of each hour, so subjects' treatment type was determined based on the time they visited the table at the card show. Second, since I ask subjects to (implicitly) rank the two goods, by definition I am controlling for all Hicksian income and substitution effects. Third, I was careful in choosing goods the individual would actually consume, rather than put up for trade or sale immediately after the transaction.

Fourth, note that I use a selection rule to attenuate attrition: I choose only people who agreed to come back again and again to be part of this experiment. This design choice does not present a problem provided such people are not more likely to show experience effects: that is, provided the people who signed the affidavit are not more or less likely to learn to overcome the endowment effect, this approach is fine. Finally, I still worried about attrition given the problems List and Daniel L. Millimet (2008) faced using a similar recruiting procedure. I therefore informed participants that I would draw 20 names in February. All 20 people who were drawn, and showed up for all experiments, were paid \$500. This lottery was in addition to the lottery that I ran for those incentivized to trade in the six months.

III. Experimental Results

Table 1 provides an overview of the results. The table places the 60 subjects in their bins: 30 subjects in the *noexperience* treatment and 30 in the *experience* treatment. A first consideration is that the experimental design was able to balance subjects across treatment and control: 13.3 percent of subjects in the *noexperience* treatment traded in the September experiment, whereas 10 percent in the *experience* treatment traded. These trading rates are not different at conventional levels using a test of proportions.

Equally important, these trading rates suggest that undertrading occurred. Given that

TABLE 1—SUMMARY TRADING STATISTICS

Treatment	Percent traded	z-value for test of proportions
<i>noexperience</i>		
September	13.3 (4 of 30)	0.40
December	10.7 (3 of 28)	-2.13
February	20.7 (6 of 29)	-2.71
<i>experience</i>		
September	10.0 (3 of 30)	
December	34.5 (10 of 29)	
February	55.2 (16 of 29)	

Notes: Percent traded provides the percentage of subjects who traded their endowed good for the alternative in the experiment. Test of proportions has a null hypothesis of no treatment effect. For example, the first test measures the *noexperience* September trading rate against the *experience* September trading rate. A z-value of 0.40 suggests that the trading rates are not different at conventional levels.

subjects were randomly allocated either good *A* or good *B*, equivalence of WTA and WTP would imply that approximately half of the goods were improperly allocated and should be traded. The actual percentages of subjects who chose to trade was roughly 10–13 percent, suggesting that $WTA > WTP$. These figures suggest that once endowed with one of the goods the subjects were roughly 10 times more likely to select that good (computed as $\frac{1}{2}((P_{A|A}/P_{A|B}) + (P_{B|B}/P_{B|A}))$).

Second, the experimental design was able to attenuate attrition: overall, only two people dropped out completely after the September experiment and only one other person missed the December experiment. The results summarized in Table 1 are robust to allowing these subjects to behave in a manner that refutes the learning hypothesis (i.e., those agents in the *noexperience* treatment traded in the latter experiments and those in the *experience* treatment did not).

Third, there is considerable evidence in favor of market experience influencing the trading rate. When comparing the trading rates across the *noexperience* and *experience* treatments, there is a clear departure in trading patterns: whereas the September rates are not significantly different, differences in the February rates are statistically significant at conventional levels ($z = -2.71$). In addition, when I examine within person trading rates I find that significantly more people who did not trade in the September experiment opted to trade in the February experiment in the *experience* treatment.

My finding that individual undertrading can dissipate after six months of trading experience strikes me as a large treatment effect, though it is smaller than the recent laboratory results of Engelmann and Hollard (2010). Combining this large effect with the fact that subjects overtraded in the February experiment deserves further explanation. At least three factors might be driving my results: i) market experience is indeed strong enough to leads to considerably higher trading rates, ii) the manner in which the design was executed lead to an experimenter-demand effect that yielded higher trading rates over time (subjects in the *experience* treatment, who were incited to trade over the six month period, felt it apropos to trade in the experiment), and iii) the sample sizes are small, which can yield such results simply due to chance.

IV. Conclusions

Previous field experiments show that market experience and the endowment effect are negatively associated. Their designs typically capture the distinction between consumers who had previous market experience and those who have had no rigorous market experience. While performance differentials across these groups lend insights into the rational choice behavior of individuals in and out of the sports card market, self-selection of individuals into trading in the sports card trading market prohibits one from drawing strong conclusions concerning the causal effect of market experience on rational behavior.

This study amends this potential shortcoming by making market experience exogenous. In doing so, I find that the major result of previous work is confirmed: as market experience intensifies, behavior approaches neoclassical expectations. Beyond its deep implications for economic theory, the lessening of the value disparity has important implications for damage assessment and benefit cost analysis. For example, how might one wish to think about using such findings to assess the value of ecosystem damages, such as oil spills?

Using experiments that measure statements of value rather than trading rates, List (2004b) finds that less experienced subjects state much higher WTA figures than more experienced subjects, which induces the value divergence to lessen for more experienced agents: WTA is approximately \$17 less, while their WTP is

generally only \$2–\$4 greater. This is consonant with the notion that people overestimate the pain of giving something up because of an affective forecasting error (see Kermer et al. 2006, for similar results in the lab).

Such results give one hope that it is possible to calibrate the results of contingent valuation survey data. Using the List (2004b) results as an example, true valuation would be a weighted average of WTP and WTA, with the weights being derived from the experiment: true valuation = $(17 \times \text{WTP} + 3 \times \text{WTA})/20$ in this case. This approach highlights that by using a real-market methodology for private goods, one can discover general principles (e.g., for inexperienced agents, stated WTP is a more reliable guide to “true” preferences than stated WTA) that can be applied to stated-preference methods for public goods. Much more research is necessary, clearly, before adopting such an approach.

REFERENCES

- Bateman, Ian, Alistair Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden.** 1997. “A Test of the Theory of Reference-Dependent Preferences.” *Quarterly Journal of Economics*, 112(2): 479–505.
- Choe, Hyuk, and Yunsung Eom.** 2009. “The Disposition Effect and Investment Performance in the Futures Market.” *Journal of Futures Markets*, 29(6): 496–522.
- Dhar, Ravi, and Ning Zhu.** 2006. “Up Close and Personal: Investor Sophistication and the Disposition Effect.” *Management Science*, 52(5): 726–40.
- Engelmann, Dirk, and Guillaume Hollard.** 2010. “Reconsidering the Effect of Market Experience on the ‘Endowment Effect.’” *Econometrica*, 78(6): 2005–19.
- Feng, Lei, and Mark S. Seasholes.** 2005. “Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Financial Markets?” *Review of Finance*, 9(3): 305–51.
- Gächter, Simon, Henrik Orzen, Elke Renner, and Chris Starmer.** 2009. “Are Experimental Economists Prone to Framing Effects? A Natural Field Experiment.” *Journal of Economic Behavior and Organization*, 70(3): 443–46.
- Greenwood, Robin, and Stefan Nagel.** 2009. “Inexperienced Investors and Bubbles.” *Journal of Financial Economics*, 93(2): 239–58.
- Hammack, Judd, and Gardner M. Brown.** 1974. *Waterfowl and Wetlands: Toward Bioeconomic Analysis*. Baltimore, MD: Johns Hopkins University Press.
- Harrison, Glenn W., and John A. List.** 2004. “Field Experiments.” *Journal of Economic Literature*, 42(4): 1009–55.
- Kermer, Deborah A., Erin Driver-Linn, Timothy D. Wilson, and Daniel T. Gilbert.** 2006. “Loss Aversion Is an Affective Forecasting Error.” *Psychological Science*, 17(8): 649–53.
- Knetsch, Jack L.** 1989. “The Endowment Effect and Evidence of Nonreversible Indifference Curves.” *American Economic Review*, 79(5): 1277–84.
- List, John A.** 2003. “Does Market Experience Eliminate Market Anomalies?” *Quarterly Journal of Economics*, 118(1): 41–71.
- List, John A.** 2004a. “Neoclassical Theory versus Prospect Theory: Evidence from the Marketplace.” *Econometrica*, 72(2): 615–25.
- List, John A.** 2004b. “Substitutability, Experience, and the Value Disparity: Evidence from the Marketplace.” *Journal of Environmental Economics and Management*, 47(3): 486–509.
- List, John A., and Daniel L. Millimet.** 2008. “The Market: Catalyst for Rationality and Filter of Irrationality.” *The B.E. Journal of Economic Analysis & Policy*, 8(1): Article 47.
- Munro, Alistair, and Yannick Ferreira De Sousa.** 2008. “Truck, Barter and Exchange versus the Endowment Effect: Virtual Field Experiments in an Online Game Environment.” Munich Personal RePec Archive (MPRA) Paper 8977.
- Randall, Alan, and John R. Stoll.** 1980. “Consumer’s Surplus in Commodity Space.” *American Economic Review*, 70(3): 449–55.
- Seru, Amit, Tyler Shumway, and Noah Stoffman.** 2010. “Learning by Trading.” *Review of Financial Studies*, 23(2): 705–39.