Environmental & Resource Economics (2005) 32: 35–53
DOI 10.1007/s10640-005-6027-1

Scientific Numerology, Preference Anomalies, and Environmental Policymaking

JOHN A. LIST
The University of Maryland, 2200 Symons Hall, College Park, MD 20742-5535, USA (e-mail: Jlist@arec.umd.edu)

Accepted 11 March 2005

Abstract. Recently an abundance of experimental evidence has been gathered that is consonant with the notion that individual preferences are inconsistent and unstable. These empirical results potentially undermine the theoretical foundation of welfare economics, as the degree of preference lability claimed suggests that perhaps no optimization principles underlie even the most straightforward of choices. Yet policymakers in the environmental arena continue to prescribe policies based on economics-based methods that are constructed on the very principles that have been directly refuted. Are policymakers creatures of habit that move at glacial speed or is there something deeper behind their inertness? In this study, I explore this issue within the U.S. context and argue that there is some rationality behind current public policy decisionmaking. I then explore whether the empirical evidence supports the view that policymakers should take preference anomalies seriously. As a case study, I focus on some of my recent findings on preference inconsistencies in the marketplace.

Key words: anomalies, government policy

JEL classifications: Q26, C93

The crux of modern day welfare economics can be found in the theory of riskless choice, as pioneered by Jeremy Bentham and James Mill, which characterizes utility maximization as an individual process whereby decisionmakers' preferences are consistent and stable. This paradigm lays the groundwork of decisionmaking theory today, as a hallmark of proposed public policies around the globe is an account and estimation of their benefits and costs. Yet numerous experimental studies over the past few decades have argued that the theoretical foundation of welfare economics may not be as stable as first surmised. In an influential experimental study using a discrete-choice auction to buy and sell commodities with close substitutes (pens and coffee mugs), Kahneman et al. (1990) provide evidence that preferences between two goods are not independent of the consumer's current entitlements. In a related line of research, theoretically equivalent decision modes, such as juxtaposed versus isolated choice, have led to systematically different choices (Hsee 1998).

While the pattern of results in these two lines of inquiry certainly questions whether individual preferences are consistent and stable, an important open issue is whether these findings invalidate the standard methods of environmental policymakers. For example, if willingness to pay (WTP) and willingness to accept (WTA) measures of value are not roughly equivalent (taking into account income and substitution effects), the approach to cost-benefit analysis needs to be re-thought because its theoretical foundation is illegitimate. Perhaps policymakers are unaware of the apparent behavioral anomalies. Alternatively, policymakers may be cognizant of the empirical findings and are currently charting the correct course of action to take account of these results.

In this study, I explore these issues in a purely descriptive manner. I report that in my (albeit limited and unscientific) discussions with agency officials in the U.S. who perform/oversee benefit-cost analyses, many are aware of these empirical findings, and realize that they have been robust across unfamiliar goods, such as irradiated sandwiches, and common goods, such as chocolate bars, but many remain skeptical of the received results. Most importantly for our purposes, some policymakers view experimental laboratory results with a degree of suspicion, one noting that the methods are akin to "scientific numerology." When pressed on this issue, some suggest that their previous experience with stated preference surveys leads them to discount experimental results, especially those with student samples, and they conclude that the empirical findings do not merit policy changes *yet*. A few policy officials openly wondered if the anomalous findings would occur in experiments with "real" people.

Given that experimental economics is a relatively new area of study, it is not surprising that some "mainstream" economists share a few of these same concerns. In the literature concerning preference inconsistency, for example, some economists have conjectured that the WTA/WTP disparity is merely the result of a mistake made by inexperienced consumers and through time these consumers will learn, and their behavior will more closely match predictions from neoclassical models (e.g., Knez et al. 1985; Coursey et al. 1987; Brookshire and Coursey 1987; Myagkov and Plott 1997). This line of thought could be considered an appropriate summary of the beliefs held by the modal policymaker whom I interviewed.

In this study I review several of my own market-based data sets to provide some insights into whether actual market participants exhibit preference structures that resemble previous patterns observed from undergraduate students in the lab. Overall, I find that with respect to the value disparity (i.e., WTA > WTP), experienced agents in the marketplace behave largely in accordance with neoclassical expectations. This result, in and of itself, is quite amazing considering the robustness of the extant literature. Yet within my data sets when one examines behavior of agents who are less experienced,

anomalous behavior is observed with high frequency. This same general pattern remains when I examine value statements across theoretically equivalent decision modes (juxtaposed versus isolated choice); but here even experienced agents exhibit a degree of anomalous behavior.

The balance of this study is organized into three sections. Section 1 provides a brief background and a summary of my discussions with agency officials. Section 2 summarizes the various experimental market designs and empirical results. Section 3 concludes with an interpretation of the data and a discussion of how policymakers can potentially make use of these findings.

1. Background and Agency Beliefs

1.1. WTA/WTP DISPARITY

Nearly three decades ago, a substantial body of empirical evidence began to develop that provided evidence that WTP and WTA measures of value were quite different (e.g., Hammack and Brown 1974; Gordon and Knetsch 1979; Brookshire et al. 1980; Rowe et al. 1980; Schulze et al. 1981). Typically these studies found that WTA measures were substantially greater than WTP measures for the same commodity.

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 (for a good discussion of the latter see Kahneman 1986). 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.[1] Of course, whether preferences are defined over consumption levels or changes in consumption has serious implications for the discipline of economics in general. 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.

For environmental economists, in particular, the disparity has considerable relevance. For example, it is well known that 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. From a practical viewpoint, the decision on whether to use compensating or equivalent variation measures is important because the losses associated with changes in the status quo would weigh much more heavily than corresponding gains (Knetsch 1990).[2]

One could argue that the large WTA/WTP disparities that have been observed in the literature can be reconciled with neoclassical theory. Hanemann (1991) summarized this line of reasoning by demonstrating that

Randall and Stoll's (1980) "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. 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 the Hicksian theory has motivated the experimental work of, for example, Knetsch (1989), Bateman et al. (1997), and List (2003, 2004a), who ask subjects to (implicitly) rank two goods. The ranking, by definition, controls for all Hicksian income and substitution effects. For our purposes, I will focus on the work of List (2003, 2004a) below, since it examines directly whether market experience has a degree of influence on the shape and magnitude of the disparity.

1.2. "MORE IS LESS" PREFERENCE REVERSAL

The intersection of psychology and economics has provided several fundamental insights. Amongst the seminal contributions is the oft-cited preference reversal literature: theoretically equivalent measures of preference, such as choices and prices, can lead to systematically different preference orderings (see, e.g., Slovic and Lichtenstein 1968). A cousin to this line of inquiry is the recent documentation of another type of preference reversal: the "more is less" preference reversal, which relates to decisionmaking between two different modes of valuation.

Bazerman et al. (1992) provide the original demonstration of changes in preferences across elicitation formats. They focused on absolute versus relative hypothetical payoffs in potential resolutions of a dispute. They found that preferences reversed across scenarios based on whether the choices were evaluated in isolation or juxtaposed. In a related experimental study, Hsee (1998) had undergraduate students indicate their hypothetical WTP for one or both sets of dinnerware. Set A contained 24 high quality pieces, whereas Set B contained 40 pieces – 24 high quality pieces and 16 low quality pieces. Hsee finds that in separate evaluations Set B is valued less than Set A, but when juxtaposed Set B dominates Set A. Hsee (1998) also performs various related thought experiments using private consumable goods such as scarves, coats, and ice cream cones and finds similar insights.[3]

List (2002) provides the first test of the "more is less" preference reversal in an actual marketplace with real transactions. In List's (2002) experiment, subjects endogenously enter the marketplace and self-select into their roles as experienced or inexperienced consumers. Even in the marketplace, with the rigors of competition in full force and arbitrage occurring all around, List

(2002) finds a ''more is less'' result in his data.[4] Alevy et al. (2002) further the work of List (2002) by examining hypothetical valuation decisions over public goods. They too find a ''more is less'' result. I will return to a more thorough discussion of these latter two studies in Section 3.

As List (2002) points out, even if one overlooks the fact that these results represent serious problems for economics-based methods of decision support in general, from a nuts-and-bolts policy perspective these results merit serious consideration in several circles. One particularly important area concerns benefit-cost analysis – ever since President Reagan's 1981 Executive Order 12291, federal agencies have been required to consider both the benefits and costs of regulations prior to their implementation. Two very distinct methodologies are currently quite popular in the estimation of the total value of non-market goods and services: (i) dichotomous choice questions, wherein the good or service under consideration is valued in isolation, and (ii) choice-based methods, wherein the economic agent selects the most preferred alternative from a *set* of choices. The ''more is less'' empirical findings suggest that with the proper presentation of attributes, these two institutions could very well yield *opposite* policy recommendations. Since benefit-cost analysis remains the central paradigm used throughout the public sector, these results indicate that much more attention should be paid to the development of consistent approaches for estimating the benefits and costs of public programs.

1.3. AGENCY BELIEFS

To the rationalist, the empirical results summarized above on the WTA/WTP disparity and the ''more is less'' preference reversal are a weed in the Utopian garden that awaits eradication by the market. To the behavioral scientist, findings such as these are not startling, and a behavioralist might argue that these types of results should be harvested since eradication is simply implausible. When I began this investigation I suspected that many policymakers lay somewhere in between these two extremes.

I approached several U.S. environmental policymakers to obtain their insights into these matters. Appendix A provides a verbatim copy of the questionnaire that was sent via e-mail to the group of potential respondents. To provide an incentive for timely and thorough responses, I also telephoned several agency officials and offered to have lunch with them at their earliest convenience. In total, I e-mailed, telephoned, or visited with more than 30 individuals who have the power to influence environmental policymaking in the U.S.

Each of the respondents had previously worked on a cost-benefit analysis or had overseen an analysis that weighted benefits and costs. Concerning the impetus for doing such studies, most respondents noted President Clinton′s

Executive Order 12866, which reaffirmed the earlier executive order of the Reagan Administration requiring that federal agencies consider costs, benefits, and economic impacts of regulations prior to their implementation.[5] In terms of computing benefit estimates, respondents had a keen understanding of the tradeoffs between revealed and stated preference approaches. Most noted that they "trusted" empirical estimates from revealed methods more than estimates from stated methods, but understood that contingent-valuation provides the only analytical approach currently available for estimating total values.[6] In this sense, policymakers understand that the flexible and holistic contingent valuation approach makes it the "only game in town" in a wide variety of situations.

Yet they remain cautious in their use of stated preference methods. Many fear that the hypothetical nature of the exercise as well as what one respondent described as the ability of agencies to "get whatever number they want by cute usage of survey techniques" may undermine the nature of the value estimate. While I believe it would not be difficult to standardize survey approaches to avoid these "cute" methods, the message about the status of the various estimation techniques should nevertheless be clear: stated values are used, but often begrudgingly. This strongly points to a need for future research in this area. This research should aim to provide a consistent and flexible approach to value estimation. I believe one path that is worth pursuing is an exploration into the effects of consequentialism in value elicitation (see, e.g., Carson et al. 2002).

Moving to anomalies, agency officials understand that certain empirical regularities have been discovered, and realize that they have been robust across many types of goods – one noting that "yes, I have seen those results for candy bars and mugs and believe they occur for a lot of other types of goods too." While at the same time, many remain somewhat skeptical of the results, another noting that "even though these results appear prevalent, they are suspiciously drawn... by methods similar to scientific numerology." When pressed on this very issue (in interviews), some suggest that their previous experience with stated preference surveys leads them to discount experimental results, especially those with "students," and conclude that they do not believe the empirical findings merit policy changes *yet*. A few policy officials openly wondered if these preferences would occur when "real" people were used as subjects. I find this a useful avenue for future research, some of which is expanded on below.

2. Experimental Setup and Results

A striking feature among the responses by agency officials is their agreement on the *potential* importance of the experimental results (e.g., WTA/WTP

disparity), but some policymakers view these experimental laboratory results with a degree of suspicion. Many desire external validation, in that they wish to know if there is evidence of such an effect outside of the laboratory, or with non-student subject pools. Given that these are the same sorts of questions that many economists have brought to bear on certain experimental findings, I believe that their prudence in the face of such uncertainty could well be considered rational.

Yet, it is now possible to shed light on whether anomalies are evident in the marketplace. This represents the genesis of some of my recent work, which aims to explore behavior in a natural setting with real market players. I do so by making a marketplace my experimental laboratory: subjects would be engaging in similar activities whether I were present in the marketplace (and running an experiment) or were a passive observer. In this sense, I am gathering data in the least obtrusive way possible while still maintaining the necessary control to execute a clean comparison between treatments.

Furthermore, it is important to recognize that in my field experiments agents endogenously select into the market and they are likely to have previous experience trading related goods. This experimental strategy may lead to different results compared to an experiment where the roles are exogenously induced by the experimenter (e.g., some subjects are given experience while others are not), but it is my belief that a rigorous examination of behavior in an actual marketplace is an important next step in understanding the nature of preference-based anomalies.

2.1. WTA/WTP EVIDENCE

This section provides a summary of data in List (2003; 2004a; 2004b), who examined behavior across consumers with intense market experience (dealers) and those that typically had less market experience (non-dealers). List's studies can be split into four categories and are a mix of ''framed'' and ''natural'' field experiments (see Harrison and List 2004): (i) examining *trading* patterns of ''familiar'' goods, (ii) examining *trading* patterns of ''unfamiliar'' goods, (iii) examining *bidding* patterns for ''familiar'' goods, and (iv) examining *bidding* patterns for ''unfamiliar'' goods. All of these data are gathered on the floor of a sportscard show or in a collector pin market organized by Walt Disney World at the Epcot Center in Orlando, Florida. As elaborated upon more fully in the original studies, with the rise in popularity of memorabilia in the past two decades, markets that organize buyers and sellers have naturally arisen. Temporal assignment of the physical marketplace is typically done by a professional association or local sportscard dealer who rents a large space, such as a gymnasium or hotel conference center, and allocates tables to dealers for a nominal fee. When the market opens, consumers mill around the marketplace, higgling and bargaining with dealers,

who have their merchandise prominently displayed on their table. The duration of a typical show is a weekend, and a lucrative show may provide any given dealer hundreds of exchange opportunities (buying, selling, and trading of goods).

The trading pattern data are gathered in the spirit of Knetsch (1989) and Kahneman et al. (1990), who use a straightforward random allocation design with two treatments. 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% trade when endowed with one good, (100-p)% should trade when endowed with the other good for neoclassical predictions to be satisfied. Given that the goods in these experiments are roughly of equal value, researchers have simply stated that fewer than 50% of the subjects should swap their good if an endowment effect exists.[7] Alternatively, if an endowment effect does not exist, approximately 50% of the subjects should trade their good. In both Knetsch (1989) and Kahneman et al. (1990), the evidence is sharp and suggests that an endowment effect exists.

In my "familiar" goods trading experiments run at sportscard (pin trading) shows, I randomly endow subjects with unique sports (pin) memorabilia and examine individual trading rates.[8] I call these "familiar" good trading exercises because subjects commonly buy, sell, and trade related goods in this particular marketplace. In these trading data, a substantial amount of evidence suggests that individual behavior converges to the neoclassical prediction as trading experience intensifies. This major insight is perhaps best illustrated in Figure 1, which pools the data across the field trading treatments – a total of 281 subjects. Figure 1, which makes the trade probability a function of previous trading experience, clearly illustrates that individual behavior converges to the neoclassical prediction as consumers gain experience. This effect was found in both the pin and sportscard trading markets.

The experimental design to gather data on trading patterns of "unfamiliar" goods is identical to that of "familiar" goods with one exception: rather than endowing subjects with a unique item that they typically buy/sell/trade in the marketplace, I follow Knetsch (1989) and Kahneman et al. (1990) and endow subjects with either a mug or a candy bar (see List 2004a). This exercise represents a particularly strict test of the role of market experience on shaping choices since psychological research suggests that transfer of learning across situations is quite weak (Loewenstein 1999).

The data again show a distinct sign: individual behavior converges upon the neoclassical prediction as trading experience intensifies. This major insight is illustrated in Figure 2, which includes data across 191 subjects. The empirical results are consistent with the trading patterns observed for "familiar" goods. This is quite surprising, and suggests that market partici-

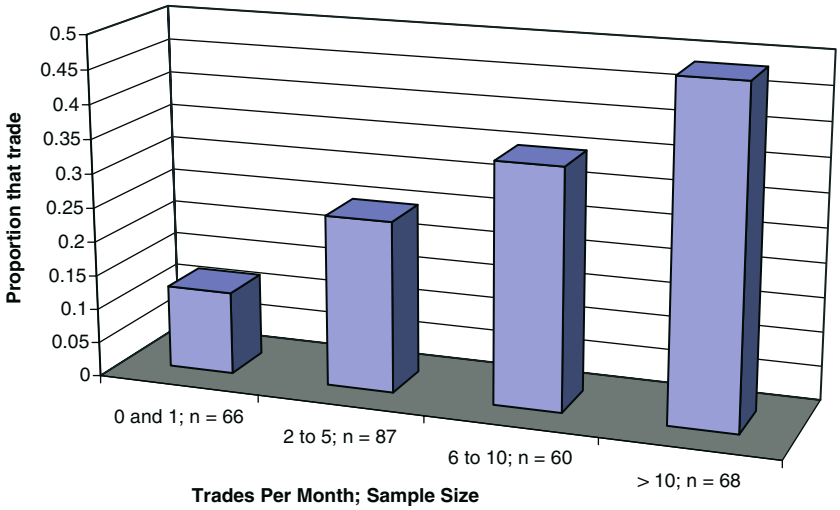

Figure 1. Summary of "familiar" good trading results.

pants are treating the mugs and candy bars similarly to goods they commonly buy/sell/trade in this particular marketplace. Note that in considering these data there is nothing magical about 11 trades per month. I make use of this figure as the high cutoff-point in both Figure 1 and 2 because it represents the mean monthly trading rate amongst dealers in the marketplace and is about one standard deviation above the average trading rate of non-dealers in the marketplace.

Data for the latter two categories – bidding patterns for "familiar" goods and "unfamiliar" goods – were gathered in the same manner as the trading data, but in these treatments I run actual auctions on the marketplace floor. The allocation institution used was either a random *n*th price auction or a Becker-DeGroot-Marschak discrete-choice auction. The distinction between "familiar" and "unfamiliar" goods is the same as that used in the trading treatments.

To provide a sense of the relationship between the WTA/WTP value disparity and market intensity, I provide Figure 3 for the "familiar" goods treatments (see List 2004b, for further details). Figure 3 plots the WTA/WTP ratio as a function of market experience. In choosing these thresholds, I again considered both the mean number of trades in a typical month (4.7) and its standard deviation (4.2). Figure 3 provides summary estimates for a relatively low-valued good (denoted "near mint") and a relatively high-valued good (denoted "mint").

Similar to the previous two figures, for both types of goods the data in Figure 3 suggest that individual behavior converges to the neoclassical prediction as trading experience intensifies. Yet with these auction data one

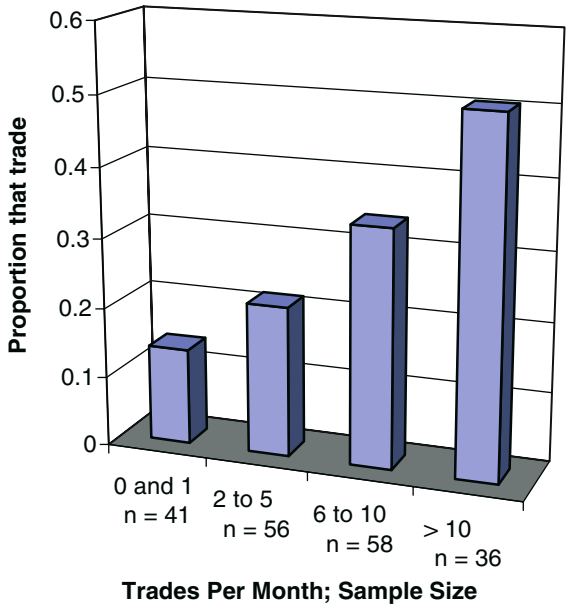

Figure 2. Summary of "unfamiliar" trading results.

can understand more meaningfully what is behind the data pattern observed. A closer examination of the two components of the value disparity (WTA and WTP) reveals that less casual subjects state much lower WTA figures than casual subjects, which induces the value divergence to lessen: super-

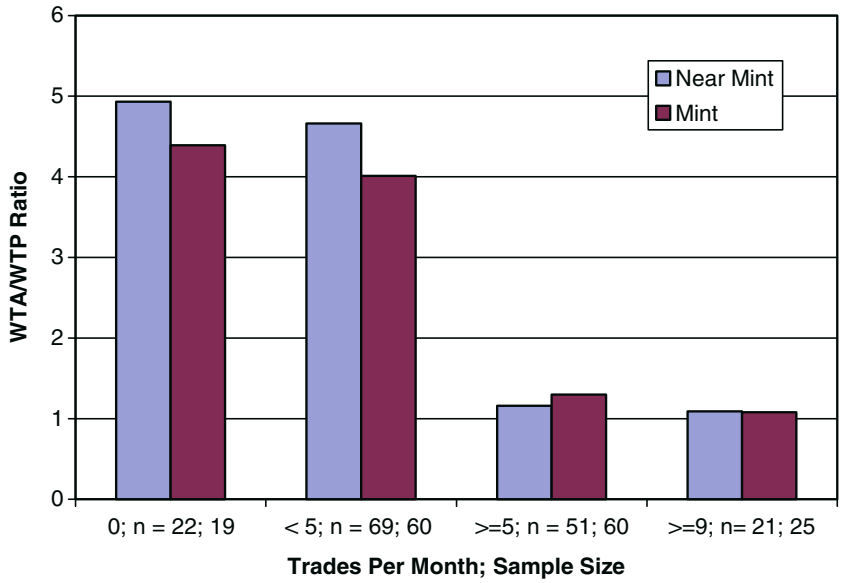

Figure 3. Summary of WTA/WTP ratios for "familiar" goods.

intense consumers' WTA is approximately \$17 less than light consumers' WTA, while their WTP is generally only \$2–\$4 greater.

Data for the "unfamiliar" goods treatments are displayed in Figure 4 (see List 2004c). Figure 4 provides a summary of individual behavior across the WTA and WTP treatments for candy bars and coffee mugs. The figure includes data across 120 subjects and makes the WTA/WTP ratio a function of trading intensity. Akin to Figure 3, Figure 4 clearly shows that individual behavior converges upon the neoclassical prediction as trading experience intensifies. In light of the extant body of psychological evidence that reports limited transfer of learning across tasks, much like the trading results, these findings are quite surprising. Akin to the results for "familiar" goods, in this case market learning appears to be on the sell-side of the market: inexperienced agents state WTA values more than three times larger than experienced agents, whereas WTP measures of value are roughly equivalent.

In the "unfamiliar" goods case, the data are broadly consistent with at least one potential explanation: the main effect of endowment is not to enhance the appeal of the good one owns, but rather the "pain" of giving it up (Loewenstein and Kahneman 1991). Thus, via market interaction and numerous arbitrage opportunities, practiced agents may have learned to overcome this "pain" and treat the good leaving their endowment as an opportunity cost rather than a loss. This learning process may proceed in several ways, but one avenue is that agents actually over-estimate the pain of giving something up and it is not until after doing so several times that the agent realizes that the duration and extent of the pain were not as severe as they predicted *ex ante*.[9] In this sense, while psychological effects have been extremely popular in explaining the endowment effect, data from the current

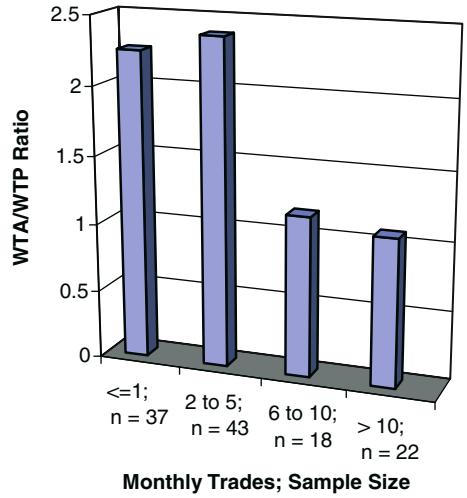

Figure 4. Summary of WTA/WTP ratios for "unfamiliar" goods.

set of treatments suggest that psychological effects may also help to explain attenuation of the anomaly.

2.2. "MORE IS LESS" EVIDENCE

The WTA/WTP market data certainly paint a picture that market experience has eradication power. Another potential anomaly that should be of great interest to environmental policymakers is the "more is less" phenomenon. While none of the environmental policymakers cited this anomaly as potentially frustrating the underpinnings of cost-benefit analysis, a number alluded to the fact that framing and context may influence stated preferences considerably. And, the policymaker who noted that surveyers can get whatever number they want by usage of survey techniques certainly understood the power of framing and context. In this section, I review data from one such framing approach – the "more is less" preference reversal (Alevy et al. 2002; List 2002).

Briefly, akin to the WTA/WTP studies, List (2002) was also carried out in an actual marketplace and included consumers that had intense market experience (dealers) and those that had less market experience (non-dealers). Amongst other tests, my data included Treatment IS (IS denotes inferior, separate), which auctioned off 10 1982 *Topps* professionally graded baseball cards. Each of the 10 cards was graded near mint/mint, and the 10-card bundle had a book value of approximately \$15. In a separate treatment SS (SS denotes superior, separate), I auctioned off a bundle of 13 cards: 10 *identical Topps* baseball cards *and* an additional 3 different 1982 baseball cards that were professionally graded to be in "poor" condition – the worst grade possible. While the three additional cards are of much lower quality than the original 10 cards, they do have economic value: in aggregate, the 13-card bundle has a book value of approximately \$18.

In a third treatment, denoted Treatment J, I sold in an auction the exact same two bundles side by side. Accordingly, each subject submits two bids, one for each commodity bundle. To provide comparable budget sets across the three treatments, I informed subjects in Treatment J that if they were deemed winners of both auctions, a random coin toss would determine which auction was binding.

Figure 5 contains a summary of the experimental data. In total, the data include decisions of 241 subjects: 130 non-dealers and 111 dealers. The data paint a compelling picture: non-dealers bid, on average, \$4.86 for the 10-card bundle and only \$3.06 for the 13-card bundle, a difference of approximately 37%. This difference is statistically significant at the $P < .05$ level using a large-sample t-test ($t = 2.03$). In this case, "more" is certainly mapping into "less". Moving rightward, the non-dealer data indicate a preferences reversal: the 10-card bundle's mean bid is \$3.72 (denoted JI) whereas the 13-card

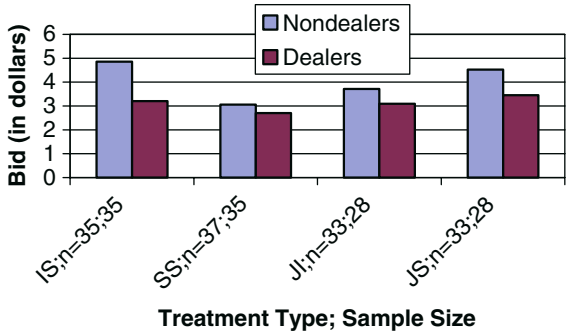

Figure 5. "More is less" private goods data.

bundle's mean bid is \$4.52 (denoted JS). This difference of approximately 20% is statistically significant. While these results are stark, and quite surprising to observe in a mature market setting, an observer could contend that if a fundamental "more is less" preference reversal exists, it should prevail regardless of the level of subject experience.

Figure 5 shows that this preference reversal, while somewhat attenuated, is alive and well, even amongst dealers, who could reasonably be considered "super-experienced" consumers. The mean dealer bid in the 10-card auction is \$3.20, whereas the mean bid decreases to \$2.70 for the 13-card bundle. While these bidding distributions are not statistically different from one another at conventional levels, the observed difference of more than 15 percent is noteworthy. Yet when the bundles are juxtaposed, statistically significant differences are found. I observe mean bids of \$3.09 (JI) and \$3.45 (JS) for the 10- and 13-card bundles. Using both a matched pairs t-test and a Wilcoxon signed-rank test for matched pairs, I can reject the null hypothesis that bids in the 10-card auction are derived from the same parental population as bids in the 13-card auction at the $P < .01$ level ($t = 3.73$). Unlike the WTA/WTP disparity, here the anomaly does not disappear when moving from non-dealers to dealers. Certainly, however, more needs to be done in this area because there is attenuation observed, suggesting that experience might matter.

It is important to examine if this anomalous behavior over private goods extends to the realm of public good valuation. To this end, Alevy et al. (2002) examine stated preferences for goods and services that are not traded in markets. Amongst other treatments, in their study they investigated stated values for a public good: wetlands restoration. The questions offered a dichotomous "yes" or "no" choice regarding willingness to contribute to wetlands restoration. The dichotomous choice evaluations were each made in isolation and so are analogous to the "separate" treatment in the sportscard

market. In one of these, question IS (''inferior separate'') asks about willingness to contribute to *entirely cleaning up 500 acres of wetlands* in the Chesapeake Bay; question SS (denoting ''superior separate'') asks about willingness to contribute to *entirely cleaning up 500 acres of wetlands* in the Chesapeake Bay and *partially cleaning up 50 acres*. The third question offers a choice between the two clean-up programs and the status quo (no purchase). Alevy et al. (2002) varied the required contribution amount for the public good from $50 to $100.

Figure 6 summarizes their data gathered from 166 individuals who attended Maryland Day in April 2002, an event that took place on the University of Maryland's College Park campus. Unlike List's (2002) data, in these data there is no working measure of individual experience; hence I present these results noting only that these data were gathered from individuals who cared enough about agricultural and environmental issues to visit the College of Agriculture's site at Maryland Day.

Data in Figure 6 show that the ''more is less'' preference reversal arises in both the $50 and $100 treatments. In the $50 case, 68% of IS respondents voted ''yes'' to the contribution question, whereas 63% of SS respondents voted ''yes''. Yet when juxtaposed, 50% voted for JS while only 30% voted for JI. This pattern is also evident in the $100 treatments, where slightly fewer respondents voted ''yes'' in the SS treatment than in the IS treatment, and 45% voted for JS while only 31% voted for JI when the goods were juxtaposed. This evidence is consonant with List (2002) and the previous ''more is less'' preference reversal literature, and suggests that even amongst individuals who, by revealed preference (showing up at the College of Agriculture site), have knowledge about environmental matters, the framing effect is important.

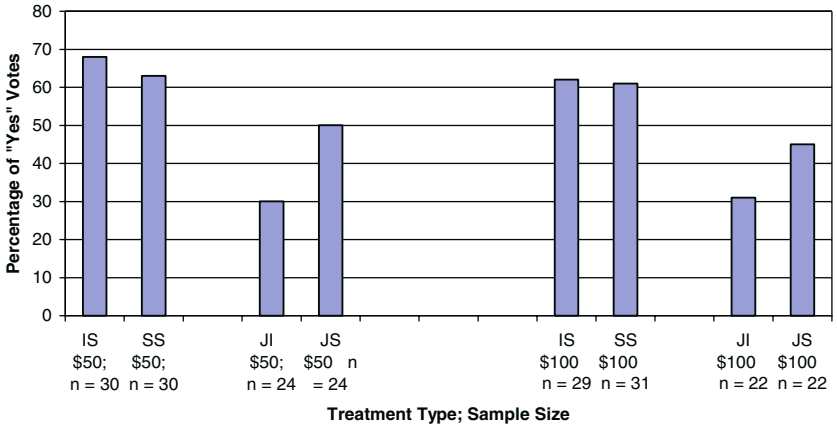

Figure 6. ''More is less'' public goods data.

3. Discussion

Whether behavioral research will be the dead limb on the evolutionary environmental policy tree or a vibrant branch remains undetermined. I find it quite clear that an important step in mapping behavioral results into environmental public policy decisionmaking is to not only refine and understand the nature of the extant laboratory results, but also to take these inquiries into the field. This observation may appear self-serving from a researcher who has made a living out of executing experiments in the marketplace, but it also represents the reality of the current landscape. While experimental studies in the lab might be a logical first step in experimental exploration – since they provide the ability to control for subtleties and examine comparative static effects while simultaneously controlling for numerous other important factors – once the lab provides an important insight, a next important step is to subject this finding to real world settings: real world settings where actors endogenously choose their preferred market in which to interact, their roles within that market, and the level of time and effort they spend in the market.

This paper summarizes field data that attempt to take this next step into the realm of two important behavioral anomalies – the WTA/WTP disparity and the "more is less" preference reversal. Rather than revert to a summary of what I have described above, I use my remaining space in this special issue to discuss what environmental policymakers, particularly those interested in practical cost-benefit analysis and stated preference methods, can take from this set of results.[10] In doing so, I focus on the value disparity findings. The crux of the issue pertains to how one interprets my findings. If, on the one hand, the market itself *creates* preferences that are "loss aversion free," then it is not clear that policymakers will find much to be excited about here – indeed, in such cases policymakers may well not want to *screen* out loss aversion from valuations they receive from stated preference methods.

Alternatively, if one interprets my findings as suggesting that individuals have "true" preferences that do not exhibit loss aversion, and market experience allows those true preferences to be "discovered," then the valuation problem becomes a task of finding methods of preference elicitation for nonmarketed goods that simulate the relevant experience before elicitation takes place. Tentatively, this is how I am interpreting my findings.[11] Note that this interpretation is consistent with empirical results reported in List and Shogren (1999) and the lottery auction results of Loomes et al. (2003).[12]

In practice, how the policymaker provides the level of experience necessary for agents to come upon their "true preferences" most likely revolves around the good in question and the resources available. One generic starting point is to use the intuition in Plott and Zeiler (2003): practice with money, explain the optimum strategy, answer questions, and use examples exten-

sively. All of these exercises have been found to facilitate getting subjects through the initial reaction/strategic responses and reflecting in the ''right'' directions.

Alternatively, if the policymaker found this approach infeasible, there is potential scope for calibration.[13] For example, List (2004b) finds 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. Of course, research has suggested that the value disparity varies along certain observable characteristics of the good (see Horowitz and McConnell 2002), yet if one replicates my convergence findings in the lab across several types of goods and finds ''conditional'' disparities and ''conditional'' convergence estimates for the various classes of goods, then it is distinctly possible to calibrate the results of contingent valuation survey data. Using my 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*WTP + 3*WTA)/20. 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.

Acknowledgements

Robert Sugden and an anonymous reviewer provided excellent comments that improved the study considerably. Participants at the ''Coping with preference anomalies in environmental decision-making'' conference at East Anglia University also provided useful comments, as did Wiktor Adamowicz, Jonathan Alevy, and Liesl Koch. Melonie Williams helped in the distribution of the agency survey. E-mail exchanges with Charles Plott also sharpened the messages contained in the study.

Notes

1. See Horowitz and McConnell (2002) for a good survey of the literature.
2. The impact of the value disparity on optimal policymaking has been crafted much more cogently than I can produce; hence I direct the reader to Knetsch (2002) for a recent treatment.
3. Note the similarities between this line of research and the asymmetric dominance effects found by Huber et al. (1982) and Simonson and Tversky (1992); in addition, this research is related to the ''cycling'' findings reported in Starmer (1999). The link between these three lines of work is that choices from a set of options are influenced by dominance relations

between those options in a way that does not carry over to other choice problems in which the options do not appear side-by-side. Thanks to Robert Sugden for pointing this out.

4. Of course, this finding also has important implications for non-expected utility resolutions.
5. The more than 100 federal agencies issue approximately 4500 new rulemaking notices each year. About 25% of those 4500 are significant enough to warrant Office of Management and Budget (OMB) review. Of those, about 50–100 per year meet the necessary condition of being "economically significant" (more than $100 million in *either* yearly benefits or costs). Every economically significant proposal receives a formal analysis of the benefits and costs by the agency. The OMB establishes guidelines for the agencies on how to perform benefit-cost analysis. Every so often the OMB revisits these guidelines. During the time I was a Senior Economist at the Council of Economic Advisers (2002–2003), the OMB and the Council of Economic Advisers jointly revised these guidelines.
6. See Sugden (2005) for intuitive reasons why revealed preference methods are preferred to stated preference methods when both are feasible.
7. Thaler (1980) coined the term "endowment effect," which implies that a good's value increases once it becomes part of an individual's endowment. The literature has loosely interchanged "endowment effect" and "WTA/WTP disparity." This is unfortunate, as the endowment effect is but one reason why one may observe a value disparity.
8. Consistent with previous studies, subjects trade with the experimenter.
9. This line of thought is consonant with recent findings in health and behavioral economics, where studies oftentimes report that individuals are better at adapting to tragic loss of a limb or divorce, for example, than they predicted *ex ante*.
10. Many thanks to Robert Sugden for urging me to discuss these issues and for his advice on where to explore.
11. See Plott (1996) for a good discussion of the discovered preference hypothesis.
12. I should stress "consistent with" since Loomes et al. (2003, p. C165) state that "our results suggest that market experience *does* tend to erode whatever casual factors generate the tendency for WTA to be systematically greater than WTP." But, they later discuss why they cannot pinpoint whether the mechanism at work is "refining" or "market discipline."
13. This line of reasoning is consonant with Sugden's (2005) short run/long run discussion.

Appendix A. Policymaker questionnaire

1. Have you ever taken part in a project that involved cost/benefit analysis?
Yes No
2. What determines whether your agency undertakes a cost/benefit analysis?
3. Which of the following benefit estimation techniques do you, generally, believe provides a more reliable signal of economic value? (Please circle)
Stated Preference (e.g., contingent valuation) Revealed Preference (e.g., travel cost, hedonics)
4. Do you believe that stated preference techniques have any important drawbacks?
5. Some recent laboratory studies have found that certain anomalies are persistent in value statements—e.g., willingness to pay is significantly less than willingness to accept. Do you believe that these findings create a major problem for cost/benefit analysis?
Yes No
Can you please explain?
6. Are you aware of any other anomalies that may influence cost/benefit analysis?

References

Alevy, J., J. A. List and W. Adamowicz (2002), "'More Is Less' Preference Reversals and Non-Market Valuation", working paper, University of Maryland.

Bateman, I., A. Munro, B. Rhodes, C. Starmer and R. Sugden (1997), 'A Test of the Theory of Reference-Dependent Preferences', *Quarterly Journal of Economics* **112**(2, May), 479–505.

Bazerman, M. H., G. F. Loewenstein and S. Blount White (1992) 'Reversals of Preference in Allocation Decisions: Judging an Alternative Versus Choosing Among Alternatives', *Administrative Science Quarterly* **23**, 220–240.

Brookshire, D. and D. Coursey (1987), 'Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures', *American Economic Review* **77**(4, September), 554–566.

Brookshire, D., A. Randall and J. R. Stoll (1980), 'Value Increments and Decrements in Natural Resource Service Flows', *American Journal of Agricultural Economics* **62**(3), 185–193.

Carson, R., T. Groves, J. A. List and M. Machina (2002) 'Probabilistic Influence and Supplemental Benefits: A Field Test of the Two Key Assumptions Underlying Stated Preferences'. Working Paper: University of California, San Diego.

Coursey, D., J. Hovis and W. Schulze (1987), 'The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value', *Quarterly Journal of Economics* **102**(3), 679–690.

Gordon, I. M. and J. L. Knetsch (1979), 'Consumer′s Surplus Measures and the Evaluation of Resources', *Land Economics* **55**(1, February), 1–10.

Hammack, J. and G. Brown (1974), *Waterfowl and Wetlands: Toward Bioeconomic Analysis*, Baltimore: The Johns Hopkins University Press for Resources for the Future.

Hanemann, W. M. (1991), 'Willingness to Pay and Willingness to Accept: How Much Can They Differ?', *American Economic Review* **81**(3, June), 635–647.

Harrison, G. W. and J.A. List (2004), 'Field Experiments', *Journal of Economic Literature* ***XLII***, 1009–1055.

Horowitz, J. K. and K. E. McConnell (2002), 'A Review of WTA/WTP Studies', *Journal of Environmental Economics and Management* **44**, 426–447.

Hsee, C. K. (1998), 'Less is Better: When Low-Value Options Are Valued More Highly Than High-Valued Options', *Journal of Behavioral Decision Making* **11**, 107–121.

Huber, J., P. John and P. Christopher (1982), 'Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis', *Journal of Consumer Research* **9**(June), 90–98.

Kahneman, D. (1986), Comments by Professor Daniel Kahneman, in R. G. Cummings, D. S. Brookshire and W. D. Schulze, eds., *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman and Allanheld.

Kahneman, D., J. L. Knetsch and R. H. Thaler (1990), 'Experimental Tests of the Endowment Effect and the Coase Theorem', *Journal of Political Economy* **98**(6, December), 1325–1348.

Knetsch, J. L. (1989), 'The Endowment Effect and Evidence of Nonreversible Indifference Curves', *American Economic Review* **79**(5, December), 1277–1284.

Knetsch, J. L. (1990), 'Environmental Policy Implications of Disparities between Willingness to Pay and Compensation Demanded Measures of Values', *Journal of Environmental Economics and Management* **18**, 227–237.

Knetsch, J. L. (2002) 'Policy Analysis and Design with Losses Valued More than Gains and Varying Rates of Time Preference', in R. Gowda and J.C. Fox, eds., *Judgements, Decisions, and Public Policy*. Cambridge: Cambridge University Press.

Knez, P., V. L. Smith and A. Williams (1985), 'Individual Rationality, Market Rationality, and Value Estimation', *American Economic Review* **75**(2, May), 397–402.

List, J. A. (2002), 'Preference Reversals of a Different Kind: The More is Less Phenomenon', *American Economic Review* **92**(5), 1636–1643.
List, J. A. (2003), 'Does Market Experience Eliminate Market Anomalies?', *Quarterly Journal of Economics* **118**(1), 41–71.
List, J. A. (2004a) 'Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace', *Econometrica* **72**(2), 615–625.
List, J. A. (2004b) 'Substitutability, Experience, and the Value Disparity: Evidence from the Marketplace', *Journal of Environmental Economics and Management* **47**(3), 486–509.
List, J. A. (2004c) 'Using Hicksian Surplus Measures to Examine Consistency of Individual Preferences: Theory and Experimental Evidence', *Scandanavian Journal of Economics*, forthcoming.
List, J. and J. Shogren (1999), 'Price Signals and Bidding Behavior in Second-Price Auctions with Repeated Trials', *American Journal of Agricultural Economics* **88**, 942–949.
Loewenstein, G. (1999), 'Experimental Economics from the Vantage-Point of Behavioural Economics', *Economic Journal* **109**(February), F25–F34.
Loewenstein, G. and D. Kahneman (1991) 'Explaining the Endowment Effect', Working paper, Carnegie Mellon.
Loomes, G., S. Chris and S. Robert (2003), 'Do Anomalies Disappear in Repeated Markets', *Economic Journal* **113**, C153–C166.
Myagkov, M. and C. Plott (1997), 'Exchange Economies and Loss Exposure: Experiments Exploring Prospect Theory and Competitive Equilibria in Market Environments', *American Economic Review* **87**(5, December), 801–828.
Plott, C. R. (1996), Rational Individual Behavior in Markets and Social Choice Processes: the Discovered Preference Hypothesis, in J. A. Kenneth, C. Enrico, P. Mark and S. Christian, eds., *The Rational Foundations of Economic Behavior* (pp. 225–250). London: Macmillan.
Plott, C. R. and K. Zeiler (2003) 'The WTP/WTA Gap, the ''Endowment Effect'' and Experimental Procedures for Eliciting Valuations', working paper, California Institute of Technology.
Randall, A. and J. R. Stoll (1980), 'Consumer's Surplus in Commodity Space', *American Economic Review* **71**(3), 449–457.
Rowe, R. D., R. C. d'Arge and D. S. Brookshire (1980), 'An Experiment on the Economic Value of Visibility', *Journal of Environmental Economics and Management* **7**, 1–19.
Schulze, W. D., R. C. d'Arge and D. S. Brookshire (1981), 'Valuing Environmental Commodities: Some Recent Experiments', *Land Economics* **57**(2), 151–172.
Simonson, I. and A. Tversky (1992), 'A Choice in Context-Tradeoff Contrast and Extremeness Aversion', *Journal of Marketing Research* **29**(3), 281–295.
Slovic, P. and S. Lichtenstein (1968), 'Relative Importance of Probabilities and Payoffs in Risk-Taking', *Journal of Experimental Psychology Monographs* **78**, 1–18.
Starmer, C. (1999), 'Cycling with Rules of Thumb: An Experimental Test for a New From of Non-Transitive Behaviour', *Theory and Decision* **46**, 141–158.
Sugden, R. (2005) 'Coping with Preference Anomalies in Cost–Benefit Analysis: a Market-Simulation Approach', *Environmental & Resource Economics* **32**, 129–160 (this issue).
Thaler, R. (1980), 'Toward a Positive Theory of Consumer Choice', *Journal of Economic Behavior and Organization* **1**, 39–60.