Journal of Environmental Economics and Management **44**, 426–447 (2002)
doi:10.1006/jeem.2001.1215

A Review of WTA / WTP Studies

John K. Horowitz and Kenneth E. McConnell

Department of Agricultural and Resource Economics, University of Maryland, College Park, Maryland 20742-5535
E-mail: horowitz@arec.umd.edu

Received February 23, 2000; revised November 1, 2000; published online January 31, 2002

Willingness to accept (WTA) is usually substantially higher than willingness to pay (WTP). These constructs have been studied for roughly 30 years and with a wide variety of goods. This paper reviews those studies. We find that the less the good is like an "ordinary market good," the higher is the ratio. The ratio is highest for non-market goods, next highest for ordinary private goods, and lowest for experiments involving forms of money. A generalization of this pattern holds even when we account for differences in survey design: ordinary goods have lower ratios than non-ordinary ones. We also find that ratios in real experiments are not significantly different from hypothetical experiments and that incentive-compatible elicitation yields higher ratios.

1. INTRODUCTION

The difference between willingness to pay (WTP) and willingness to accept (WTA) has been widely studied through both theory (Hanemann [33, 34], Randall and Stoll [66], Willig [82]) and experiments. In a typical experiment, a subject is given some item, like a coffee mug, and offered money to return it to the experimenter. The dollar amount the subject asks for is his WTA. Another subject is not given a mug and instead is asked to pay for one. The amount he offers is his WTP. Previous authors have shown that WTA is usually substantially larger than WTP, and almost all have remarked that the WTA/WTP ratio is much higher than their economic intuition would predict (e.g., Kahneman *et al*. [45], hereafter KKT.)

The pervasiveness of high WTA/WTP ratios and the wide variety of goods that have been used in the experiments have combined to sustain interest in WTA vs. WTP for roughly 30 years. We analyze those studies here. Although large WTA/WTP ratios are well documented, the findings do not seem to have had much of an effect on either economic models or discussions of policy design, as Knetsch [46] has noted. A wider role for these findings has been hampered, we believe, by two issues.

First, it has seemed possible that the high observed ratios are due to "weak" experimental features such as hypothetical payments, student subjects, or elicitation questions that are not incentive-compatible. According to this argument, more realistic experiments, such as those with real money or incentive-compatible elicitation, will yield lower and more reasonable ratios. Likewise, we might suspect that repeating a given experiment with the same group of participants would cause the ratios to fall to closer-to-expected levels over time, as Coursey *et al*. [20] found in some of their experiments. These possibilities have not been investigated on a

0095-0696/02 $35.00

broad-based scale. A finding that design features were playing a major role would almost surely make the literature's WTA/WTP findings of less concern despite their preponderance.

The second issue that has blocked a wider role for the WTA/WTP findings is an absence of a rich set of behavioral patterns that any behavioral model might be expected to cover or explain. For which goods is a high ratio most likely to be found? A few studies have claimed that the ratio is smaller for goods that have close substitutes (Harless [36], Shogren *et al.* [71], hereafter SSHK), but the evidence about this, or other possible trends, has remained diffuse. In addition to providing a guide for future model-building, these patterns should make it easier to understand how the ratio might manifest itself in real-world economic behavior.

This paper uses a collection of WTA/WTP studies to address these issues. We found 45 studies that reported usable data. The studies draw on a remarkable range of goods: chocolates, pens, mugs, movie tickets, hunting licenses, visibility, nuclear waste repositories, nasty-tasting liquids, pathogen-contaminated sandwiches, and many others. This variety allows unique insight into the WTA/WTP ratio. To our knowledge, no other economic issue has been experimentally studied across such a wide variety of goods.

With regard to experiment design, we find that ratios in real experiments are not significantly different from hypothetical experiments, and that incentive-compatible elicitation yields higher ratios, not lower. In other words, survey techniques that would be expected to yield a "truer" picture of preferences lead either to no change or to higher observed ratios. We also found that students tended to have lower, not higher, ratios than the general public, so moving the experiments out of the classroom seems not to lead to lower ratios. The evidence on the effects of repetition is mixed, but there is not strong evidence that the ratio decreases through iteration. Therefore, high WTA/WTP ratios are not the result of experimental design features that would be considered suspect, even apart from their WTA/WTP results.

With regard to patterns in the observed ratios, we find that, on average, the less the good is like an "ordinary market good," the higher is the ratio. The ratio is highest for public and non-market goods, next highest for ordinary private goods, and lowest for experiments involving forms of money. A generalization of this pattern holds even when we account for differences in survey design: ordinary goods have lower ratios than non-ordinary ones. This pattern is the major result we discovered.

This paper does not take up the issue of whether the WTA/WTP findings provide evidence for or against the neoclassical paradigm, even though that potential has been the theme of much of the literature we reviewed. Rather, our goal is to draw broad-based results from this long and rich experimental track record. We are not concerned with whether the observed ratios are consistent with the standard neoclassical model. It is possible to investigate these results without addressing the neoclassical question.

We end this introduction with several examples of the importance of the WTA/WTP ratio. The ratio comes into play in the context of the assignment of property rights, since the difference between a WTA and WTP experiment is a difference in property rights over the item being valued. Therefore the WTA/WTP ratio measures the consequence of assigning a property right one way or the other.

Consider the case of preserving land from development. In our set of studies, the mean WTA/WTP ratio is approximately 7. This number suggests, roughly speaking, that the amount of land that would be preserved if development rights were held by the general public is 7 *times higher* than the amount that would be preserved if the rights were deeded to the landowner and had to be purchased by the public.[1] This difference would substantially alter the balance of environmental preservation and urban development in the United States.

As Knetsch [46] has pointed out, a large difference between WTA and WTP can have other potent effects on environmental policy. These occur when the appropriate welfare measure is willingness to accept (because, in most instances, environmental quality can only deteriorate) but policy analysts use willingness to pay as the measure of benefits. For marketed goods, predictions about volume of trade or gains from trade may be seriously flawed if they are derived without recognizing the frequently large difference between individuals' willingness to pay and their willingness to accept for goods (Borges and Knetsch [10]; see Franciosi *et al*. [26] for the counter-implication). Fischel [24] reviews the arguments about whether the WTA/WTP ratio justifies compensating landowners by more than their property's market value when property is taken by eminent domain.

There seems to be a further lesson implied by these examples. Because holders of some "right" appear to value the right differently from non-holders, one of the most economically consequential decisions will be the *initial* establishment of the property rights, especially for environmental and other public amenities for which property rights are unclear (Knetsch [46]). Such a prediction runs contrary to economists' understanding of the consequences of assigning property rights. It is with this possibility in mind that we undertake this review.

2. THE STUDIES

A list of the studies is given at the end of the paper. A sample of the data is given in Table I. A table with data from all of the studies is available from the authors. We have not re-analyzed the authors' analyses. For example, if the authors chose to remove outliers, we have accepted their decision and used the means and medians they reported. When multiple trials were conducted, we have used the authors' summary of the results. There are 45 studies that are usable for our analysis and five that are not, for reasons explained below.

Let $i = 1 \dots 45$ index the individual studies we used for the analysis. Let $k = 1 \dots K_i$ index the experiments within a study, where K_i is the number of experiments reported in study i. The pair ik is the unit of observation; that is, a line in the data table (see Table I), which we call an *experiment*. Let N_{ik} be the number of subjects in experiment k of study i. There are 208 observations in our data set.

There are six experiments in five studies for which the authors did not report the number of subjects. These experiments are excluded from parts of the analysis.

Dependent variable. The variable that is the focus of most of the analysis is mean WTA/mean WTP. This is labeled RATIO_{ik}. The highest RATIO_{ik} is 2858,

[1] This case presumes that for developers, willingness to pay for development rights equals willingness to sell.

TABLE I
Sample of Data (First 12 Observations)

Study	Med. WTA/ med. WTP	Mean WTA/ mean WTP	Mean WTP	Private/public good	Elicitation: Hyp./real	Elicitation: Technique	Elicitation: Subjects	*N*
#1. Adamowicz *et al.*	—	1.95	\$4.76	Private (movie ticket)	Hyp.	Open	Students	157
	1.85	1.70	\$28.50	Private (hockey ticket w/subst.)	Hyp.	Closed	Students	150
	1.87	1.91	\$36.60	Private (hockey tkt w/o subst.)	Hyp.	Closed	Students	150
#2. Banford *et al.*	2.60	2.78	\$43.10	Public (ocean pier)	Hyp.	Closed, iterated	Local public	71
	4.40	4.24	\$21.97	Public (home postal deliv.)	Hyp.	Closed, iterated	Local public	71
#3. Bateman *et al.*	3.95	2.81	£0.78	Private (10 chocolates)	Random real	Open; IC	Students	89, 96
	2.00	2.09	8.7 chocs.	Private (money—£2.00)	Random real	Open; IC	Students	96, 89
	1.30	2.53	£0.60	Private (4 cans of Coke)	Random real	Open; IC	Students	89, 96
	2.00	2.00	2.5 cans	Private (money—£0.80)	Random real	Open; IC	Students	96, 89
(Pounds sterling were converted at £1 = \$1.55. To calculate mean WTP for treatments 2 and 4, we used the prices implied by treatments 1 and 3.)								
#4. Benzion *et al.*	—	1.76	\$4.90	Private (\$40 in 0 ⇔ 6 mos.)	Hyp.	Open	Students	204
	—	1.92	\$6.80	Private (\$40 in 0 ⇔ 1 yr.)	Hyp.	Open	Students	204
	—	1.96	\$9.80	Private (\$40 in 0 ⇔ 2 yrs.)	Hyp.	Open	Students	204

found by SSHK (Study #38) in the first round of a study of salmonella contamination. We drop this observation because it is 25 times larger than the next highest ratio, 113, found by Brookshire and Coursey (#8). Therefore, there are 201 $(208 - 6 - 1)$ observations in our main sample.

Aggregation: *individual means vs. group means*. Most studies report only mean WTA/ mean WTP, even when open-ended WTA and WTP values were collected from all individuals. Only two studies reported both the mean of individual WTA/WTP ratios and the ratio of mean WTA/mean WTP (Dubourg *et al.* #13; Eisenberger and Weber, #15). In those two experiments, the result

$$\sum_i \frac{\mathrm{WTA}_i}{\mathrm{WTP}_i} > \frac{\sum_i \mathrm{WTA}_i}{\sum_i \mathrm{WTP}_i} \tag{1}$$

was obtained, where i represents the individual subject. The right-hand side is our dependent variable. Therefore, the literature's high reported ratios are not due to using aggregated measures.

Kachelmeier and Shehata [44] found a correlation between WTA and WTP of 0.35, and Borges and Knetsch [10] found a correlation of 0.24. These relatively low correlations also suggest that little has been lost by the literature's concentrating on the ratio of means.

Median vs. mean. There are 41 experiments that reported ratios of both means and medians. The ratio involving means was greater than the ratio involving medians in close to 80% of the experiments. An unweighted regression of the former on the latter yielded

$$\frac{\text{Mean WTA}}{\text{Mean WTP}} = \underset{(0.72)}{2.67} + \underset{(4.15)}{1.58} \frac{\text{Median WTA}}{\text{Median WTP}} \qquad n = 41. \tag{2}$$

T-ratios are in parentheses. Studies #24 and #42 reported only ratios of medians. We predicted their ratio of means with the use of (2) and used those ratios in our regressions. To test sensitivity to using the predicted ratio, we also estimated our main model without studies 24 and 42.

Mean WTP. This is the denominator of RATIO_{ik}. In the regressions, mean WTP is deflated to 1983 dollars, using the Consumer Price Index.

Private/public good. We label as a public good any good for which a collective decision was being made, even if the good might actually be private (e.g., hunting permits) or have both private (access) and public (quality) dimensions (e.g., home postal delivery). The item types are discussed further in Section 3.

Increments and decrements. Most studies ask for willingness to pay to go from A to B and willingness to accept the change from B to A, which is the comparison that has been the subject of most theoretical analysis. A few studies ask about willingness to pay for the change from A to $A + \Delta$ and willingness to accept the change from A to $A - \Delta$ (Brookshire and Coursey, #8; Jones-Lee *et al.* #49; Thaler, #42; Viscusi, Magat, and Huber, #44); see Hanemann [34] for a discussion of the difference between this comparison and the standard one. We did not distinguish between these in our analysis.

Hypothetical, real, and random real. A study is hypothetical if the valuation question was purely hypothetical, and real if it was carried out for real money and real exchange. A study is "random real" if several valuation questions were asked and (only) one of them was chosen at random and carried out for real money and real exchange. Random real experiments are classified as REAL = 1 in the regressions.

Elicitation technique and incentive compatibility. The main techniques are the following: (1) A simple open-ended question such as "What is the maximum you would be willing to pay to obtain X?" These questions do not provide subjects with incentives to reveal their true maximum WTP, but strategic bias was typically considered unimportant when many of these studies were conducted (see Schulze *et al.* [68]). (2) An incentive-compatible (IC) open-ended question. These use Vickrey auctions, Becker–deGroot–Marschak mechanisms, or something similar. (3) Payment card. A payment card has several values printed on it, and the subject then circles his own WTP or WTA. These may be IC (e.g., combined with a Vickrey auction) or not. (4) Single closed-ended, yes–no question. The individual is asked whether he or she would pay or accept some specified amount, call it c. The amount c varies across the sample, but each individual answers only one valuation question. WTA and WTP means must be estimated from the group's responses. We used the authors' estimates unless otherwise noted. Simple closed-ended questions are incentive compatible. (5) Iterated closed-ended question. Each individual answers several closed-ended questions with different values of c. If c is varied a fixed number of times (usually twice), then WTA and WTP means must still be estimated from group responses. If c is varied until the subject is roughly indifferent between yes and no, then each subject's individual value is eventually observed and no estimation is necessary. Iterated closed-ended questions are not incentive compatible.

Because we are interested in the effects of these survey methods, we used the same categorization for hypothetical as for real questions. Incentive compatibility can be important even for hypothetical experiments if responses have at least some probability of affecting a real-world outcome, as Carson *et al.* [17] point out.

Subjects. Many of the subjects were students. When a local issue was being studied with non-students, subjects were the "local public." When an issue without a clear location-specific connection was being studied with non-students, the subjects were the "public." For empirical purposes, we distinguish only between students and the rest of the subjects.

Number of observations. If only a single number of observations is listed, then each of these N_{ik} subjects answered both a WTP and WTA question. If there were missing values for some subjects (but all subjects were asked both questions), we list the smaller number of responses, if available. If two numbers are listed, separate subject groups answered the WTA and WTP questions. The first number is WTA observations, the second is WTP observations.

Rules for including studies. We included every study we could find, including studies we do not have copies of, but whose ratios were reported in other studies. There are seven such studies in data set. Because these are frequently several years old and unpublished, we decided not to try to gather more information about them.

TABLE IIA
Summary Statistics (Weighted)

Statistic	N	Mean	Standard error
RATIO_{ik}	201	7.17	0.93
RATIO_{ik} (excludes estimated RATIOs)[a]	175	7.18	1.02
RATIO_{ik} (excludes study #4)	169	7.86	1.07
Median WTA/median WTP	66	5.52	1.03
Mean WTP ($1983)	169	$175	22.40
REAL (real = 1, hypothetical = 0)	201	0.22	0.03
IC (incentive-compatible = 1)	201	0.25	0.03
STUDENT (student subjects = 1)	201	0.35	0.03
OPEN-ENDED (open-ended = 1)	201	0.56	0.04

[a] Excludes 25 experiments in study #24 and 1 experiment in #42 in which only ratios of medians were reported.

We exclude from the analysis three studies that used dichotomous choice but had only one offer price (#46, #47, #48). We estimated WTA and WTP from these studies using the Turnbull estimator, but such estimates are imprecise (Haab and McConnell [31]).

A few studies elicited open-ended WTA values but dropped responses that said "I will not accept this trade under any condition." Therefore, their calculated mean WTA is a lower bound on true mean WTA. Of these, we include Dubourg *et al.* (#13) and Viscusi *et al.* (#44) but drop Jones-Lee *et al.* (#49) because its percentage of "never-accept" responses is quite large (81%). We further exclude one study from 1968 in which the mean WTP was negative but the mean WTA was positive (Slovic and Lichtenstein, #50). All analysis is based on studies #1 through #45.

After completing this research, we learned of several studies that were not included. To facilitate future research, we list them here.[2] They are Bell and Leaworthy [4], Brookshire *et al.* [14], Brown and Mathews [16], Franciosi *et al.* [26], Hartman *et al.* [37], Jones-Lee [42], and McClelland and Schulze [57].

Weights. In calculating summary statistics in Table IIA, we weighted observations by $\text{sqrt}(N_{ik})/\text{sqrt}(K_i)$. The numerator gives higher weight to experiments with more subjects, but at a decreasing rate. The denominator treats different

[2] We thank Richard Carson, Praveen Kujal, Carol Mansfield, and an anonymous referee for these citations.

TABLE IIB
Summary Statistics (Unweighted): Quantiles

Statistic	Minimum	10%	25%	50%	75%	90%	Maximum
RATIO_{ik} ($n = 206$)	0.74	1.16	1.66	2.60	6.12	10.52	112.67
Median WTA/median WTP ($n = 67$)	1.00	1.33	1.67	2.33	3.00	7.28	42.94
Mean WTP ($1983) ($n = 173$)	$0.12	$0.34	$0.91	$3.73	$34.51	$280	$5847

TABLE IIIA
Ratio by Type of Good

Good	Mean RATIO	Standard error	Number of experiments
Public or non-market goods	10.41	2.53	46
Health and safety	10.06	2.28	32
Ordinary private goods	2.92	0.30	59
Lotteries	2.10	0.20	25
Timing	1.95	0.17	39
All goods	7.17	0.93	201
Unknown number of subjects	6.71	Not calculated	6

Public or non-market: studies #2, 5, 6, 8, 9, 12, 17, 18, 20, 26, 31, 35, 36, 37, 39, 43, 45.
Health and safety: studies #13, 16, 32, parts of #38, 44.
Ordinary private goods: studies #1, 3, 7, 24, 29, 30, 34, parts of #38, 41.
Lotteries: studies #11, 15, 19, 23, 25, 27, 40.
Timing: studies #4, 21, 28, 42.
Unknown number of subjects: studies #10, 14, 22, last observation of #24, 33.

experiments within the same study as providing neither completely independent information (in which case the denominator would be 1) nor fully duplicated information (in which case the denominator would be K_i). There are six experiments that have no weights because authors did not report the number of subjects.

3. TYPE OF GOOD

The mean of RATIO_{ik} is 7.17 with standard error 0.93 ($n = 201$). The median is 2.60. The data are summarized in Table II. Statistics are weighted by $\text{sqrt}(N_{ik})/\text{sqrt}(K_i)$ in Tables IIA and III.

The studies contain a wide variety of goods. We classify them in five broad categories: health and safety, lotteries, ordinary private goods, the time at which a good will be received or given up, and public or non-market goods not included in any of the other categories. We also subdivide the timing and public/non-market categories. Summary statistics for mean RATIO by good type are given in Tables IIIA, IIIB, and IIIC.

The main result is that the farther a good is from being an ordinary private good, the higher the ratio. The pattern is striking (see Table IIIA). Ratios are highest for health/safety and public/non-market goods, next highest for ordinary private

TABLE IIIB
Ratio by Type of Good: Timing Studies

Type of good	RATIO		
	Mean	S.E.	n
Timing of receipt of private goods (study #28)	2.82	0.39	5
Timing of receipt of money (Studies #4, 21, 42)	1.84	0.18	34

TABLE IIIC

Ratio by Type of Good: Public/Non-Market Goods

Type of good	RATIO Mean	S.E.	n
Misc. public and non-market goods (#2, 6, 8, 18, 31, 39, 43, 45)	27.57	7.50	19
Hunting (#5, 9, 17, 20)	10.47	5.29	8
Visibility (#35, 36)	7.40	2.31	7
Siting (#26, 37)	4.14	1.83	4
Sucrose octa-acetate (#12)	3.99	0.47	8

goods and lotteries, and lowest for surveys that involve the time at which a good is received. The health/safety and public/non-market ratios are almost identical. The latter finding is not surprising since health and safety are themselves public/non-market goods.

The pattern continues: The closer the good comes to being actual money, the smaller the ratio. Lotteries, which were all based on money payments, have lower ratios on average than ordinary private goods. Timing studies that involve money have lower ratios than timing studies that involve goods (Table IIIB). In this regard, it is reassuring (if a little self-evident) to note that no significant difference between WTA and WTP is observed when the good is money, as in experiments using tokens [45, p. 1328].

Timing studies that involve goods behave like ordinary private goods, although the sample size is small. Timing studies that involve money behave like lotteries. This latter finding is particularly striking given the close connection economists recognize between choice over time and choice under uncertainty (Prelec and Loewenstein [63], Quiggin and Horowitz [64]).

In the public and non-market goods category (Table IIIC), the lowest ratio is for tasting of sucrose octa-acetate, a bitter non-hazardous substance that is perhaps more like an ordinary private good than any of the other items in Table IIIC. The elicited ratio is just slightly higher than for the ordinary private goods category. The pattern for other studies in the public and non-market goods category is similar but weaker, although the number of observations in many of the subcategories is low.

There are two anomalies, siting and hunting, that merit discussion because they pertain to issues that have been proposed to be resolved through the issuance of property rights.

Hunting permits are private goods and might therefore be expected to have lower ratios. But if subjects believed that their responses would also be used to make other wildlife policy decisions, then the hunting permit survey is not merely a question about a simple private good, it is implicitly a question about wildlife management, a public good.[3]

For siting studies, the relatively low ratios may be explained by subjects believing that they *did* essentially hold a property right. The fact that often only one suitable

[3] A subject's response may affect not just whether he gets a hunting permit but whether he will have to pay for permits in the future or whether hunting opportunities will be expanded/curtailed or may affect other wildlife programs.

waste site is available in a community and that the siting survey may lead subjects to believe that any other siting-type decisions would also be subject to citizen review may make the siting problem more like a private good, i.e., one that subjects had relatively clear control over; and this yields a lower ratio.[4] The sample size for this conclusion is small, however, and the explanation deserves further scrutiny.

The result in Table IIIA is consistent with Hanemann's finding that the lower the substitution elasticity between a bundle of market goods and the rationed good is, the higher the WTA/WTP ratio will be [33]. If non-market goods have lower substitution elasticities than ordinary goods, then our results are consistent with Hanemann's model.

4. REGRESSION ANALYSIS

This section uses a random effects model to look at the effects of type of good, survey design, mean WTP, and year. Three major survey design features are examined through regressions: hypothetical or real payoffs, elicitation technique, and student or non-student subjects. For type of good, we divided the sample into (a) ordinary goods and (b) all others. A more precise disentangling of good type and survey design is desirable, but it is nearly impossible because finer classifications of good types do not contain the full range of survey design features, and so separate effects cannot be discerned.

Econometric Model

The econometric model we use is

$$\text{RATIO}_{ik} = \beta x_{ik} + u_i + \varepsilon_{ik}, \tag{3}$$

where u_i is a study-specific error and x_{ik} is a vector of experiment and study characteristics. We weight observations by $\text{sqrt}(N_{ik})$ to give experiments with more subjects greater weight.

This model has the structure of an unbalanced random effects model. The advantage of the random effects model is that it allows covariance among the experiments in a given study. In some studies, there is only one experiment, so a fixed effects model cannot be estimated.

The model can be estimated as a maximum likelihood model by fixing $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_\varepsilon^2)$, estimating β conditional on ρ, and then iterating on ρ until the likelihood function is maximized (Nerlove [62, Chap. 19]). The value of ρ will be between 0 and 1, with $\rho = 0$ being the ordinary least-squares value.

In the regressions, all values of ρ are less than or equal to 0.12, with two regressions having $\rho = 0$ (#4 and #6). A low value of ρ indicates that between-experiment random variation is large relative to within-study random variation. This is a desirable finding; it suggests that quantifiable survey characteristics are capturing the greater part of the explainable variation.

The results of the weighted random effects model are given in Table IV. The basic regression is the first column (regression 1). We also ran each of the Table IV regressions, using (i) OLS with the observations weighted by $\text{sqrt}(N_{ik})/\text{sqrt}(K_i)$,

[4] One community's responses might affect decisions made for another community. If subjects were concerned only about their own community, this would be sufficient for our explanation.

TABLE IV
Regression Results: Dependent Variable = RATIO

	#1	#2	#3	#4	#5	#6	#7	#8[a]	#9[b]
INTERCEPT	6.08	4.71	2.90	8.01	2.94	4.90	7.44	5.16	18.59
	(1.12)	(1.01)	(0.59)	(1.37)	(0.77)	(1.05)	(1.04)	(0.79)	(1.96)
ORDINARY	−6.12	−6.32	−5.92	−8.14	−5.02	−8.22	−16.44	−6.73	−7.84
(Ordinary = 1)[c]	(1.78)	(1.77)	(1.70)	(1.98)	(1.62)	(1.14)	(2.86)	(1.81)	(2.13)
REAL	−7.09	1.17	−0.27	−6.10	−6.75	−5.54	−6.83	−7.79	−8.02
(Real = 1)	(1.47)	(0.34)	(0.08)	(1.23)	(1.78)	(1.31)	(1.43)	(1.51)	(1.56)
IC	10.59	—	—	11.14	9.34	11.18	11.95	10.87	6.93
(Inc. Comp. = 1)	(2.45)			(2.55)	(2.56)	(2.90)	(2.81)	(2.39)	(1.44)
STUDENT	−13.24	−11.71	−12.50	−14.49	−10.04	−13.44	−15.71	−13.05	−15.28
(Student = 1)	(3.75)	(3.16)	(3.47)	(3.81)	(3.20)	(3.73)	(4.41)	(3.46)	(3.97)
MEAN WTP	−0.005	−0.005	−0.005	−0.005	—	—	−0.005	−0.005	−0.03
(Deflated to $1983)	(1.72)	(1.80)	(1.95)	(1.71)			(1.81)	(1.63)	(3.83)
OPEN-ENDED	—	−0.54	—	—	—	—	—	—	—
(Open-ended = 1)		(0.21)							
CLOSED-ENDED	—	—	−5.17	—	—	—	—	—	—
(Closed-ended = 1)			(1.73)						
MEAN WTP ×	—	—	—	0.34	—	—	—	—	—
ORDINARY				(0.89)					
REAL ×	—	—	—	—	—	−0.26	—	—	—
ORDINARY, β_5						(0.02)			
IC ×	—	—	—	—	—	−6.99	—	—	—
ORDINARY, β_6						(0.55)			
STUDENT ×	—	—	—	—	—	11.53	14.13	—	—
ORDINARY, β_7						(1.86)	(2.23)		
YEAR	0.07	0.10	0.15	0.05	0.09	0.07	0.07	0.07	0.02
	(1.06)	(1.60)	(2.11)	(0.65)	(1.65)	(1.06)	(0.82)	(0.98)	(0.19)
ρ	0.08	0.11	0.10	0[d]	0.12	0[d]	0.05	0.07	0.07
R^2	0.20	0.16	0.18	0.20	0.14	0.17	0.23	0.20	0.23
N	169	169	169	169	201	201	169	152	137

T-statistics are in parentheses.
[a] Ordinary goods: #1, 3, 7, 24, 29, 30, 34, parts of 38, and 41.
[b] Excludes most of #24 (KKT) and all of 42, which reported only ratio of medians.
[c] Excludes #4 (BRY).
[d] Equations estimated by OLS.

and (ii) random effects with unweighted observations. The regressions gave the same patterns of sign and significance as appear in Table IV. An unweighted OLS regression yielded different conclusions.

Type of Good

Non-ordinary goods have significantly higher ratios—they are typically 6 to 8 points higher than ordinary goods. This effect occurs even when we take survey design features and mean WTP into account. The pattern that we uncovered in Section 3, that the farther a good is from being an ordinary private good the higher the ratio, remains a prominent and robust feature of the observed WTA/WTP behavior.

Only "ordinary private goods" (see Table IIIA) are counted as ordinary goods in these regressions. Lotteries and timing experiments are counted as non-ordinary goods. We adopt this division even though the previous analysis shows lotteries and timing to be like ordinary private goods in many ways. The reason for our classification is that *ex ante*—before any statistical analysis was conducted—we believe that most economists would have proffered that experiments with lotteries or timing were more unusual than experiments with goods like mugs and pens; that they were more like non-market goods. We did not want the observed differences in the ratios to influence how we classified the goods for our analysis of survey design effects.[5]

Survey Design

We next turn our attention to survey design features. The main question that confronts us is whether there is any evidence that high ratios are an experimental artifact. Much has been written elsewhere about survey design for valuing non-market goods, and our compilation of studies provides some insight into those general survey design issues. Our main focus, however, is whether high WTA/WTP ratios are the result of questionable survey design and easily identifiable influences, or instead truly present a broad-based picture of preferences.

Hypothetical vs. real payoffs. Real experiments do not yield ratios that are significantly different from those of hypothetical experiments. In some instances, such as when we account for whether an experiment uses closed-ended elicitation (regression 3), the effect of realness essentially disappears entirely. Realness has its statistically strongest effect only when the mean WTP is not included. This effect loses its significance when the mean WTP is included.

The high values of the WTA/WTP ratio initially led some researchers to claim that hypothetical surveys were unsuitable for eliciting preferences. Our results show that real experiments do not yield significantly lower ratios. Thus, any claim about the suitability of hypothetical surveys must rest on evidence other than the size of the WTA/WTP ratio.

Elicitation technique. Studies that are incentive compatible (IC) have significantly higher ratios. This result is unexpected. If high ratios were the result of "strategizing" by the subjects rather than a feature of true preferences then we

[5] To test the effect of the classification of lottery and timing goods, we also estimated regression 1 with the lottery and timing experiments excluded. The following table shows that the pattern of inferences is not sensitive to their omission. The patterns of sign and size are quite similar to the specification in Table IV.

Intercept	Ordinary	Real	IC	Student	Mean WTP	Year	ρ	R^2	N
18.70	−7.69	−12.83	11.13	−14.89	−0.03	−0.03	0.07	0.22	116
(1.62)	(1.53)	(1.71)	(1.70)	(2.60)	(3.30)	(0.20)			

would expect incentive-compatible experiments to result in lower ratios. They do not.

We also looked at other categorizations of elicitation techniques. Open-ended studies, typically construed as not being incentive compatible, had no statistically significant effect (regression 2). Closed-ended questions, which are considered incentive compatible, do yield lower ratios than non-closed-ended questions (regression 3), although the statistical significance is not strong. Furthermore, the lower ratios of closed-ended questions may be due to functional form and range-of-integration assumptions imposed by the researcher; these assumptions would likely lead to a lower WTA being inferred.

Intuition about the effect of elicitation method is complicated, leading to weak hypotheses about the effect of elicitation on the ratio. Under some approaches, subjects might either overstate WTA or WTP (if they want a good to be provided and feel they will not have to pay full price) or under-report them, if they think they might thereby get the good at a cheaper rate [17]. The overall effect that elicitation will have on the ratio is unknown.

Students vs. non-students. Students exhibit significantly and substantially lower ratios than non-students. This result is unexpected. Its main implication for our research is that high WTA/WTP ratios are exhibited by the general public, not just college undergraduates.

For non-ordinary goods, students exhibit lower ratios than do non-students. For ordinary goods, students exhibit the same ratios as non-students, because in regressions 6 and 7 we cannot reject the hypothesis that $\beta_{\text{STUDENT}} + \beta_{\text{STUDENT} \times \text{ORDINARY}} = 0$ ($F = 0.17$ with a p value of 0.68 for regression 6; and $F = 0.06$ with a p value of 0.79 for regression 7).

Ordinary goods and survey features. We also checked whether these survey design features had the same effect on both ordinary and non-ordinary goods. The hypothesis that coefficients on the three survey features crossed with ORDINARY are jointly zero cannot be rejected in regression 6 ($F = 1.96$). In other words, with respect to realness, elicitation approach, and student subjects, the responses are the same for both ordinary and non-ordinary goods.

If we look at these effects individually, however, students do appear to behave differently for ordinary and non-ordinary goods. For ordinary goods, students have the same ratios as non-students, since the coefficients on STUDENT and STUDENT × ORDINARY sum to approximately zero (as discussed above). For non-ordinary goods, students have lower ratios.

Likewise, the hypothesis that MEAN × ORDINARY is zero cannot be rejected (regression 4). In words, we find that the relationship between mean WTP and WTA/WTP is the same for ordinary and non-ordinary goods.

Other Findings

We next take up other survey design patterns and possible influences on the ratio. These yield less direct evidence about the "reliability" of the ratio but are useful because they help form a broad picture of WTA/WTP behavior.

Mean WTP. This variable requires us to exclude 32 experiments that do not report the mean WTP. We find that the higher the WTP is, the lower the

WTA/WTP ratio is. On average, a \$200 increase in mean WTP causes the ratio to decrease by 1 point. The relationship between RATIO and mean WTP is the same for both ordinary goods and other goods (regression 4).

It is possible that mean WTP is endogenous. We performed a Hausman test for endogeneity of the right-hand-side mean WTP, both in an OLS model ($\rho = 0$) and for the model with the optimal value of ρ , for regressions 1–3 and 7–9. In all cases, the test statistic does not rise to a level that would lead to rejection of the null hypothesis of no endogeneity.

Year. There has been a slight increase in the ratio over the 30 years that it has been studied. This result does not have an obvious behavioral explanation. One possible explanation is that as the existence of a disparity has been established over the years, researchers have tended to study situations where it might arise.

When Benzion *et al*. (#4, hereafter BRY) is excluded, YEAR has a small and statistically insignificant coefficient (regression 9). However, in the analog to regression 5 with BRY excluded (not shown), the YEAR coefficient is larger and statistically significant. In other words, the increase in the ratio over the years is a relatively robust finding that is only mildly sensitive to inclusion of BRY.

Income. Only a few studies have looked at the relationship between WTA/WTP and income. Therefore, we do not include income as an explanatory variable in the regressions. Adamowicz *et al*. [1] found that WTA-WTP is decreasing in income, but the coefficient is not significant. Horowitz [39] found no significant relationship between WTA/WTP and subjects' wealth.

Individual studies. In regressions 8 and 9, we re-ran regression 1 without KKT and Thaler (#24 and #42) and BRY (#4), respectively. Neither exclusion affects the results much. KKT (#24) and Thaler (#42) reported only ratios of medians, but their techniques were similar to the other studies, so it is not surprising that their results are consonant with the overall findings. BRY (#4), however, differed substantially from the other studies. In BRY, each student subject was given a 64-question survey that asked the subject, for example, to "state an amount of money \$$x$ so that he or she would be indifferent between paying \$$yt$ time periods from now or paying \$$x$ immediately" [5, p. 275]. The sheer number of questions (i.e., experiments) makes this study stand out. Also, for some responses it was necessary for us to convert a future WTA or WTP to its present value; we used the discount rate implicit in the question. Dropping this study has almost no effect on the coefficients but does slightly increase the standard errors of some estimates, as might be expected. We also ran the other regressions without BRY (not shown). There was little change. Regressions 1–7 therefore include all of the studies.

Sample selection. The studies in Table I are not a random sample. They must be selected by editors or, for unpublished studies, by the other researchers that cite them and, at a minimum, by the authors themselves, who felt the experiment worth conducting and the results worth writing up. We see three ways in which sample selection might affect our results. First, experiments in which the elicited WTAs were extremely high are less likely to enter our sample. The main consequence is that our observed mean RATIO is below the population mean. We suspect that this will be more of a problem for the types of goods that tend to exhibit high

RATIOS.[6] Therefore, the coefficient on ORDINARY is probably biased upward, and the difference between ordinary and non-ordinary goods may be greater than our regressions indicate.

Studies using non-incentive-compatible techniques also likely have a lower probability of being published. Even in our own analysis, for example, we have in places excluded the BRY study because of the suspicious incentive structure. However, the fact that some elicitation methods are not included in the empirical analysis does not necessarily bias the coefficient on IC as long as the relationship between the ratio and the elicitation method is consistent and the exclusion does not remove all variation in IC.

5. DO PRACTICE AND FAMILIARITY LOWER THE WTA / WTP RATIO?

Several authors have suggested that repeating an experiment for the same subjects might lower the ratio, primarily because WTA would be reduced as subjects realized they would be content to take home a smaller amount of real money than they first thought. KKT refer to the "conclusion reached in some other studies that the WTA-WTP discrepancy is greatly reduced by market experience" [45, p. 1335]. We list the relevant studies in Tables VA and VB.

The evidence is mixed. Brookshire and Coursey (#8) found that the ratio decreased. Their study design is not readily generalizable because they informed their subjects of the compensation fund that was available and then elicited WTA bids until either the total WTA was less than the fund or five trials had been conducted; in their WTP experiments, they told subjects the cost of the item and elicited bids until they covered that cost or reached a maximum of five trials.

Coursey *et al*. (#12, hereafter CHS) also showed a decrease in WTA, but their sample size was small, and, like Brookshire and Coursey (#8), they repeated their experiment with an explicit goal, namely until a consensus was reached, with a

[6] An example is study #49, a health and safety study in which at least 80% of subjects implicitly claimed an infinite WTA. Another example is our own censoring of the largest SSHK observation, also a health and safety study.

TABLE VA

Studies in Which Hypothetical Elicitation(s) Preceded Real Elicitation(s)

Study	Procedure	Sequence of mean WTA/mean WTP
#7. Boyce *et al.*	Ten practice rounds[a]	Not reported
	One binding round	1.66 (no-kill), 2.36 (kill)
#12. CHS	One practice round[a]	3.79
	Iterated practice rounds[a]—first and final bids	5.26, 3.80
	Four practice rounds[b]	3.95, 6.13, 3.90, 3.49
	Maximum of six rounds until no subject objected[b]	1.59
#24. KKT	Three practice rounds[c] (pens)	6.00, 6.00, 5.00
	One binding round	5.00

[a] No information on other bids was announced after a round.
[b] Other subjects' bids were announced after each round.
[c] The market-clearing bid was announced after each round.

TABLE VB
Studies with Repeated, Random-Real Elicitations

Study	Repetitions	Information revealed after each round	Sequence of mean WTA/mean WTP
#24. KKT	4 (mug)	Market-clearing bid announced	1.91, 2.33, 2.33, 2.33
	4 (pen)	Market-clearing bid announced	3.33, 2.33, 3.00, 2.33
	4 (mug)	Market-clearing bid announced	2.71, 2.11, 2.11, 1.89
	4 (binocs.)	Market-clearing bid announced	1.67, 1.67, 2.33, 2.33
	5 (mug)	Market-clearing bid announced	3.8, 2.8, 2.2, 1.8, 1.8
#34. Morrison	5 (chocolates)	No information on others' bids announced	0.99, 1.09, 1.09, 1.13, 1.13
	5 (mug)	No information on others' bids announced	2.01, 2.22, 2.42, 2.29, 2.19
#38. SSHK	5 (candy)	High bidder and reigning price announced	1.28, 1.16, 0.98, 0.93, 0.93
	20 (sandwich) (rounds 1, 7–10, and 17–20)	High bidder and reigning price announced (?)	8.74, 2.11, 2.60 (pathogen 1) 2858, 3.39, 2.20 (pathogen 2) 4.00, 3.22, 3.66 (pathogen 3) 16.09, 6.42, 6.61 (pathogen 4) 34.04, 3.05, 4.65 (pathogen 5)
	10 (mug)	High bidder and reigning price announced (?)	2.76, 1.74, 1.10, 1.05, 1.07, 1.45, 1.29, 1.24, 1.16, 0.74[a] 2.74, 1.98, 1.27, 1.03, 1.21, 1.31, 0.97, 1.19, 1.23, 0.80[a]
#8. Brookshire and Coursey	Repeated until WTP covered cost or WTA did not exceed fund. Max. of 5 trials	Sum of WTP or WTA announced	3.92, 3.63, 2.90, 2.28 (25 tree) 8.08, 8.19, 11.16, 8.28, 7.39 (50 tree)

[a] WTA was elicited in two ways, but only one set of WTPs was elicited. Thus, the denominators are the same in the two sequences. In the first WTA set, an identical mug was for sale just outside the experiment. In the second set, no mug was available. See SSHK [71] for details.

maximum of 10 trials. In a reassessment, Gregory and Furby [30] emphasized the smallness of CHS's samples and claimed that the paper's reported convergence "depends upon inclusion of... suspicious outlying groups" (p. 285). There appears to be no convergence in CHS's initial repeated hypothetical rounds; there was convergence only in the final real rounds.

Knez *et al*. [48] looked at the number of instances in which an individual subject's WTP exceeded his WTA and concluded that this number fell when the experiment was repeated. Under this result, the WTA/WTP ratio would likely have been *rising* with repetition. They did not report any ratios, so actual results are unknown.

The strongest evidence of a falling ratio comes from SSHK, who showed a significant decrease in the ratio between their first and middle rounds, for both contaminated sandwiches and mugs. This was observed in five separate experiments with a total of roughly 60 subjects. In a subsequent experiment, Shogren *et al*. ([72], cited in Shogren and Hayes [70]) showed that the ratio fell in a Vickrey auction but not in a Becker–DeGroot–Marschak auction, and argue that the former is more like a market.

Studies that explicitly claim that WTA/WTP did not fall include KKT (#24) and Morrison (#34).

In summary, the idea that the ratio will fall as subjects become familiar with an experiment may be intuitively compelling, but the evidence is weak. Some experimental techniques appear to aid convergence, such as repeating the experiment to aim for a goal and having an outside market for the good, as KKT did for their last mug experiment. In many cases, even when the ratio does fall, it falls to levels that still seem high.

Last, we should note that even if the ratio were to fall with practice, the implications would be limited because the ratio has its most important economic role in environmental and public policy decisions for which familiarity and practice are likely to be absent.

6. CONCLUDING COMMENTS

Our research investigates a body of empirical work that has appeared to challenge both the neoclassical model of the consumer and a belief in the "neutrality" of property rights. Before this challenge can stand it has been necessary to establish two related points: (i) The high observed WTA/WTP ratios do not appear to be experimental artifacts. Our claim is based on the findings that hypothetical or non-incentive-compatible experiments do not yield statistically significantly higher ratios; that high ratios are exhibited by a broad-based (i.e., non-student) population; and that familiarity with the experiments does not uniformly lead to lower ratios.

(ii) A robust and economically useful response pattern exists. We find that the farther a good is from being an "ordinary private good," the higher the ratio. The pattern prevails even when we account for possible differences in survey designs. The extensive literature on the WTA/WTP ratio provides sturdy evidence for these claims.

We leave for future papers the two major unanswered questions that this research raises. First, and we believe most important: To what extent can a disparity between WTA and WTP, or anticipation of it, be observed in real-world economic choices? Are there "remedies," and, if so, are they desirable? Second, does the WTA/WTP disparity provide sufficiently broad and deep evidence against the neoclassical model? Does that evidence warrant substantially modifying that model, at least in some situations for which economists' expertise might be called upon?

LIST OF STUDIES

See references for full citation.

#1. Adamowicz *et al*. [1].
#2. Banford *et al*. [2].
#3. Bateman *et al*. [3].
#4. Benzion *et al*. [5].

#5. Bishop *et al.* [9]. Further details appear in Bishop and Heberlein [7].
#6. Bowker and MacDonald [11].
#7. Boyce *et al.* [12].
#8. Brookshire and Coursey [13].
#9. Brookshire *et al.* [15].
#10. Casey [18], cited in Eisenberger and Weber [23].
#11. Coombs *et al.* [19].
#12. Coursey *et al.* [20].
#13. DuBourg *et al.* [21].
#14. Eby [22], cited in Meyer [56].
#15. Eisenberger and Weber [23].
#16. Gerking *et al.* [27].
#17. Hammack and Brown [32].
#18. Hanemann *et al.* [35]. Further details appear in Hoehn and Loomis [38] and Mansfield [54].
#19. Harless [36].
#20. Bishop and Heberlein [8]. See also Welsh [81], cited in Mansfield [54].
#21. Horowitz [39].
#22. Hueth *et al.* [40], cited in Bergstrom [6].
#23. Kachelmeier and Shehata [44].
#24. Kahneman *et al.* [45].
#25. Knetsch and Sinden [47].
#26. Kunreuther and Easterling [49].
#27. Lichtenstein and Slovic [50].
#28. Loewenstein [51].
#29. Loewenstein and Adler [52].
#30. Loomis *et al.* [53].
#31. Mantymaa [55].
#32. McDaniels [58].
#33. Meyer [59], cited in Hyman and Stiftel [41]. Further details appear in Meyer [60].
#34. Morrison [61].
#35. Rae *et al.* [65].
#36. Rowe *et al.* [67].
#37. Schulze *et al.* [69], cited in Fisher *et al.* [25].
#38. Shogren *et al.* [71].
#39. Sinclair [73], cited in Banford *et al.* [2] and Gordon and Knetsch [28].
#40. Singh [74].
#41. Smith [76].
#42. Thaler [77].
#43. Van Kooten and Schmitz [78].
#44. Viscusi *et al.* [79].
#45. Welle [80], cited in Mansfield [54].
#46. Gregory [29].
#47. Marshall *et al.* [56].
#48. McDaniels [58].
#49. Jones-Lee *et al.* [43].
#50. Slovic and Lichtenstein [75].

REFERENCES

1. W. L. Adamowicz, V. Bhardwaj, and B. Macnab, Experiments on the difference between willingness to pay and willingness to accept, *Land Econom*. **69**, 416–427 (1993).
2. N. D. Banford, J. Knetsch, and G. Mauser, Feasibility judgments and alternative measures of benefits and costs, *J. Bus. Admin*. **11**(1 & 2), 25–35 (1979/1980).
3. I. Bateman, A. Munro, B. Rhodes, C. Starmer, and R. Sugden, A test of the theory of reference-dependent preferences, *Q. J. Econom*. **CXII**, 479–505 (1997).
4. F. W. Bell and V. R. Leaworthy, "An Economic Analysis of the Importance of Saltwater Beaches in Florida," Department of Economics, Florida State University (1985).
5. U. Benzion, A. Rapoport, and J. Yagil, Discount rates inferred from decisions: An experimental study, *Management Sci*. **35**, 270–284 (1989).
6. J. Bergstrom, Concepts and measures of the economic value of environmental quality: A review, *J. Environ. Management* **31**, 215–228 (1990).
7. R. C. Bishop and T. Heberlein, Measuring values of extramarket goods: Are indirect measures biased? *Amer. J. Agric. Econom*. **61**, 926–930 (1979).
8. R. C. Bishop and T. Heberlein, Does contingent valuation work, *in* "Valuing Environmental Goods" (R. Cummings, D. Brookshire, and W. Schulze, Eds.) Rowman and Allanheld, Totowa, NJ (1986).
9. R. C. Bishop, T. A. Heberlein, and M. J. Kealy, Contingent valuation of environmental assets: Comparisons with a simulated market, *Nat. Resour. J*. **23**, 619–633 (1983).
10. B. Borges and J. Knetsch, Tests of market outcomes with asymmetric valuations of gains and losses: Smaller gains, fewer trades, and less value, *J. Econom. Behavior Organ*. **33**, 185–193 (1998).
11. J. M. Bowker and H. F. MacDonald, An economic analysis of localized pollution: Rendering emissions in a residential setting, *Can. J. Agric. Econom*. **41**, 45–59 (1993).
12. R. R. Boyce, T. C. Brown, G. H. McClelland, G. L. Peterson, and W. D. Schulze, An experimental examination of intrinsic values as a source of the WTA-WTP disparity, *Amer. Econom. Rev*. **82**, 1366–1373 (1992).
13. D. S. Brookshire and D. L. Coursey, Measuring the value of a public good: An empirical comparison of elicitation procedures, *Amer. Econom. Rev*. **77**, 554–566 (1987).
14. D. Brookshire, D. Coursey, and W. Schulze, Experiments in the solicitation of private and public values: an overview, *in* "Advances in Behavioral Economics" (L. Green and J. Kagel, Eds.), Vol. III, Ablex, Norwood, NJ (1990).
15. D. S. Brookshire, A. Randall, and J. Stoll, Valuing increments and decrements in natural resource service flows, *Amer. J. Agric. Econom*. **62**, 478–488 (1980).
16. G. Brown and S. Mathews, Economic valuation of the 1967 sport salmon fishery of Washington, Washington Department of Fisheries Bulletin, Technical Report 2 (1970).
17. R. Carson, T. Groves, and M. Machina, Incentive and informational properties of preference questions, unpublished manuscript, Department of Economics, University of California, San Diego (2000).
18. J. T. Casey, Predicting buyer-seller gaps for risky and riskless options, working paper, State University of New York at Stony Brook (1990).
19. C. H. Coombs, T. G. Bezembinder, and F. M. Goode, Testing expectation theories of decision making without measuring utility or subjective probability, *J. Math. Psych*. **4**, 72–103 (1967).
20. D. L. Coursey, J. L. Hovis, and W. D. Schulze, The disparity between willingness to accept and willingness to pay measures of value, *Q. J. Econom*. **102**, 679–690 (1987).
21. W. R. DuBourg, M. Jones-Lee, and G. Loomes, Imprecise preferences and the WTP-WTA disparity, *J. Risk Uncertainty* **9**, 115–133 (1994).
22. P. A. Eby, "The Value of Outdoor Recreation: A Case Study," Master's thesis, University of British Columbia (1975).
23. R. Eisenberger and M. Weber, WTP and WTA for risky and ambiguous lotteries, *J. Risk Uncertainty* **10**, 223–233 (1995).
24. W. Fischel, The offer/ask disparity and just compensation for takings: A constitutional choice perspective, *Internat. Rev. Law Econom*. **15**, 187–203 (1995).
25. A. Fisher, G. McClelland, and W. Schulze, Measures of WTP versus WTA: Evidence, explanations and potential reconciliation, *in* "Amenity Resource Valuation: Integrating Economics with Other Disciplines" (G. Peterson, B. Driver, and R. Gregory, Eds.), Venture Publishing, State College, PA (1989).

26. R. Franciosi, P. Kujal, R. Michelitsch, V. L. Smith, and G. Deng, Experimental tests of the endowment effect, *J. Econom. Behavior Organ.* **30**, 213–226 (1996).
27. S. Gerking, M. De Haan, and W. D. Schulze, The marginal value of job safety: A contingent value study, *J. Risk Uncertainty* **1**, 185–199 (1988).
28. I. M. Gordon and J. Knetsch, Consumers surplus and the evaluation of resources, *Land Econom.* **55**, 1–10 (1979).
29. R. Gregory, Interpreting measures of economic loss: evidence from contingent valuation and experimental studies, *J. Environ. Econom. Management* **13**, 325–337 (1986).
30. R. Gregory and L. Furby, Auctions, experiments, and contingent valuation, *Public Choice* **55**, 273–289 (1987).
31. T. Haab and K. E. McConnell, Referendum models and negative willingness to pay: Alternative solutions, *J. Environ. Econom. Management* **32**, 251–270 (1997).
32. J. Hammack and G. M. Brown, "Waterfowl and Wetlands: Toward Bioeconomic Analysis," Johns Hopkins University Press for Resources for the Future, Baltimore, MD (1974).
33. M. Hanemann, Willingness to pay and willingness to accept: How much can they differ? *Amer. Econom. Rev.* **81**, 635–647 (1991).
34. M. Hanemann, The economic theory of WTP and WTA, *in* "Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries" (I. Bateman and K. G. Willis, Eds.), Oxford Univ. Press, London (1999).
35. M. Hanemann, J. B. Loomis, and B. Kanninen, Statistical efficiency of double-bounded dichotomous choice contingent valuation, *Amer. J. Agric. Econom.* **73**, 1255–1263 (1991).
36. D. W. Harless, More laboratory evidence on the disparity between willingness to pay and compensation demanded, *J. Econom. Behavior Organ.* **11**, 359–379 (1989).
37. R. S. Hartman, M. J. Doane, and C. K. Woo, Consumer rationality and the status quo, *Q. J. Econom.* **106**, 141–162 (1991).
38. J. P. Hoehn and J. B. Loomis, Substitution effects in the valuation of multiple environmental programs, *J. Environ. Econom. Management* **25**, 56–75 (1993).
39. J. K. Horowitz, Discounting money payoffs: An experimental analysis, *in* "Handbook of Behavioral Economics" (S. Kaish and B. Gilad, Eds.), Vol. 2B, JAI Press, Greenwich, CT (1991).
40. D. Hueth, S. Voorhies, and R. Cosagrande, Estimating benefits from controlling nuisance pests: An application of the interactive bidding technique, Paper presented at the Annual Meeting of the American Agricultural Economics Association, Clemson (1981).
41. E. L. Hyman and B. Stiftel, "Combining Facts and Values in Environmental Impact Assessment," Westview, Boulder, CO (1988).
42. M. W. Jones-Lee, "The Value of Life: An Economic Analysis," Univ. of Chicago Press (1976).
43. M. W. Jones-Lee, M. Hammerton, and P. Philips, The value of safety: results of a national sample survey, *Econom. J.* **95**, 49–72 (1985).
44. S. Kachelmeier and M. Shehata, Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China, *Amer. Econom. Rev.* **82**, 1120–1141 (1992).
45. D. Kahneman, J. L. Knetsch, and R. H. Thaler, Experimental tests of the endowment effect and the Coase theorem, *J. Polit. Econom.* **98**, 1325–1348 (1990).
46. J. Knetsch, Environmental policy implications of disparities between willingness to pay and compensation demanded measures of values, *J. Environ. Econom. Management* **18**, 227–237 (1990).
47. J. L. Knetsch and J. A. Sinden, Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value, *Q. J. Econom.* **99**, 507–521 (1984).
48. P. Knez, V. Smith, and A. Williams, Individual rationality, market rationality, and value estimation, *Amer. Econom. Rev.* **75**, 397–402 (1985).
49. H. Kunreuther and D. Easterling, Gaining acceptance for noxious facilities with economic incentives, *in* "The Social Response to Environmental Risk" (D. Bromley and K. Segerson, Eds.), pp. 151–186, Kluwer Academic, Dordrecht (1992).
50. S. Lichtenstein and P. Slovic, Reversals of preference between bids and choices in gambling decisions, *J. Exp. Psych.* **89**, 46–55 (1971).
51. G. F. Loewenstein, Frames of mind in intertemporal choice, *Management Sci.* **34**, 200–214 (1988).
52. G. Loewenstein and D. Adler, A bias in the prediction of tastes, *Econom. J.* **105**, 929–937 (1995).

53. J. Loomis, G. Peterson, T. Brown, P. Champ, and B. Lucero, Estimating WTA using the method of paired comparison and its relationship to WTP estimated using dichotomous choice CVM, unpublished manuscript (1996).
54. C. Mansfield, "Despairing over Disparities: An Empirical Analysis of the Difference Between Willingness to Pay and Willingness to Accept," Ph.D. dissertation, Department of Economics, University of Maryland (1994).
55. E. Mantymaa, Willingness to pay and willingness to accept: An empirical test of differences with environmental commodities in a CVM field study, unpublished manuscript (1997).
56. J. Marshall, J. Knetsch, and J. Sinden, Agents' evaluations and the disparity in measures of economic loss, *J. Econom. Behavior Organ.* **7**, 115–127 (1986).
57. G. McClelland and W. Schulze, The disparity between willingness to pay and willingness to accept as a framing effect, *in* "Frontiers in Mathematical Psychology" (D. R. Brown and J. E. K. Smith, Eds.), Springer-Verlag, Berlin/New York (1991).
58. T. McDaniels, Reference points, loss aversion, and contingent values for auto safety, *J. Risk Uncertainty* **5**, 187–200 (1992).
59. P. Meyer, "A Comparison of Direct Questioning Methods for Obtaining Dollar Values for Public Recreation and Preservation," Environment Canada, Vancouver, BC (1976).
60. P. Meyer, Publicly vested values for fish and wildlife: Criteria in economic welfare and interface with the law, *Land Econom.* **55**, 223–235 (1979).
61. G. Morrison, Willingness to pay and willingness to accept: Some evidence of an endowment effect, *Appl. Econom.* **29**, 411–417 (1997).
62. M. Nerlove, Likelihood inference in econometrics, unpublished manuscript (1999).
63. D. Prelec and G. Loewenstein, Decision-making over time and under uncertainty: A common approach, *Management Sci.* **37**, 770–786 (1991).
64. J. Quiggin and J. K. Horowitz, Time and risk, *J. Risk Uncertainty* **10**, 37–55 (1995).
65. D. A. Rae, J. Hausman, and J. Wickham, Benefits of visual air quality in Cincinnati: Results of a contingent ranking survey, Charles River Associates, Research Project 1742 for Electric Power Research Institute (1982).
66. A. Randall and J. Stoll, Consumer surplus in commodity space, *Amer. Econom. Rev.* **70**, 449–455 (1980).
67. R. Rowe, R. d'Arge, and D. Brookshire, An experiment on the economic value of visibility, *J. Environ. Econom. Management* **7**, 1–19 (1980).
68. W. Schulze, R. d'Arge, and D. Brookshire, Valuing environmental commodities: Some recent experiments, *Land Econom.* **57**, 151–172 (1981).
69. W. D. Schulze, G. McClelland, B. Hurd, and J. Smith, Estimating benefits for toxic waste management: An application of the property value method, Draft Report, U.S. Environmental Protection Agency, Washington, DC (1985).
70. J. Shogren and D. Hayes, Resolving differences in willingness to pay and willingness to accept: Reply, *Amer. Econom. Rev.* **87**, 241–244 (1997).
71. J. F. Shogren, S. Y. Shin, D. J. Hayes, and J. R. Kliebenstein, Resolving differences in willingness to pay and willingness to accept, *Amer. Econom. Rev.* **84**, 255–270 (1994).
72. J. Shogren, R. Wilhelmi, C. Koo, S. Cho, C. Park, and P. Polo, Auction institutions and the measurement of WTP and WTA, unpublished manuscript (1996).
73. W. F. Sinclair, "The Economic and Social Impact of the Kemano II Hydroelectric Project on British Columbia's Fisheries Resources," Federal Department of the Environment, Fisheries, and Marine Services, Vancouver, BC (1976).
74. H. Singh, The disparity between willingness to pay and compensation demanded: Another look at laboratory evidence, *Econom. Lett.* **35**, 263–266 (1991).
75. P. Slovic and S. Lichtenstein, Relative importance of probabilities and payoffs in risk taking, *J. Exp. Psych. Monogr.* **78**, 1–18 (1968).
76. V. L. Smith, Comments, *in* "Valuing Environmental Goods: An Assessment of the Contingent Valuation Method" (R. G. Cummings, D. S. Brookshire, and W. D. Shulze, Eds.), Rowman & Allanheld, Totowa, NJ (1986).
77. R. Thaler, Some empirical evidence on dynamic inconsistency, *Econom. Lett.* **8**, 201–207 (1981).
78. G. C. Van Kooten and A. Schmitz, Preserving waterfowl habitat on the Canadian prairies: Economic incentives versus moral suasion, *Amer. J. Agric. Econom.* **74**, 79–89 (1992).
79. W. K. Viscusi, W. Magat, and J. Huber, An investigation of the rationality of consumer valuations of multiple health risks, *RAND J. Econom.* **18**, 465–479 (1987).

80. P. Welle, "Potential Economic Impacts of Acid Deposition: A Contingent Valuation Study of Minnesota," Ph.D. dissertation, Department of Agricultural Economics, University of Wisconsin–Madison (1986).
81. M. Welsh, "Exploring the Accuracy of the Contingent Valuation Method: Comparisons with Simulated Markets," Ph.D. dissertation, Department of Agricultural Economics, University of Wisconsin–Madison (1986).
82. R. Willig, Consumer's surplus without apology, *Amer. Econom. Rev.* **66**, 589–597 (1976).