This file was created by scanning the printed publication.
Errors identified by the software have been corrected;
however, some errors may remain.

Economic Valuation by the Method of Paired Comparison, with Emphasis on Evaluation of the Transitivity Axiom

George L. Peterson and Thomas C. Brown

ABSTRACT. *The paired comparison (PC) method is used to investigate reliability, transitivity, and decision time for binary choices among goods and sums of money. The PC method reveals inconsistent choices and yields individual preference order over the set of items being compared. The data reported support the transitivity assumption and demonstrate high reliability for individual preference order. The paper also discusses using the PC method to obtain conservative median and mean estimates of willingness to accept compensation for public and private goods. The PC method may prove useful for valuing public goods, and warrants further study.* (JEL Q26)

I. INTRODUCTION

Important public interest decisions often require information about the economic values people place on public goods. The contingent valuation (CV) method is apparently the only existing approach for estimating the economic value of many such goods but the state of the art is imperfect (Mitchell and Carson 1989). Most economists concur that available CV methods do not yield believable estimates of willingness to accept compensation for losses, and the use of CV to estimate willingness to pay (WTP) for gains in public goods or to avoid losses thereof remains controversial.[1] Thus, there is a need to explore new approaches.

In this paper we investigate the reliability and transitivity of economic preferences observed by the psychometric method of paired comparison (Bock and Jones 1968; David 1988; Fechner 1860; Guilford 1954; Kendall and Gibbons 1990; Kendall and Smith 1940; Thurstone 1927; Torgerson 1958). The paired comparison (PC) method yields an individual respondent's preference order among elements of a choice set by presenting the elements in pairs and asking the respondent to choose the preferred element in each pair. Because the method obtains repeated measures for each element, responses should be more reliable than the single-point estimates generally obtained by CV. PC also allows a test of the transitivity assumption.

When applied to a set of goods only, the method yields the preference order among the goods. When applied to a set containing both goods and sums of money, the method elicits multiple binary choices that enable monetary valuation by stochastic discrete choice analysis, by psychometric scaling, or by simple arithmetic methods.

The method can be used to observe choices among losses or among gains. When choosing among losses, the monetary values indicate WTP, and when choosing among gains, the monetary values indicate willingness to accept compensation (WTA) from the chooser reference point. Other things being equal, the chooser reference point yields a conservative measure of WTA that includes the income effect but avoids behavior like

The authors are, respectively, project leader and economist, Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture, Fort Collins, CO. We thank Paul Bell, Andrej A. Birjulin, Andrea Clarke, and Dan McCollum for help in experimental design and data collection, Randy Rosenberger for help in data analysis, and the following individuals for helpful comments on earlier drafts of this paper: Richard Bishop, Patricia Champ, Rudy King, Jack Knetsch, Beatrice Lucero, and Robert Mendelsohn. We, of course, absolve them from responsibility for the final product.

[1] The controversy over the ability of CV to measure the value of public goods was summarized in recent sets of articles in two journals, the *Journal of Economic Perspectives* and the *Natural Resources Journal.* See Portney (1994) and Cummings and Harrison (1994) for the introductory papers in each set. A commonly used approach to the WTA problem is to measure willingness to pay (WTP) to avoid loss. Loehman, Park, and Boldt (1994) report that WTP to avoid losses in visibility and health exceeds the bid for obtaining gains, particularly for large changes, and that the observed difference is not a logical contradiction. We still do not know, however, if WTP to avoid loss constitutes an acceptable measure of WTA.

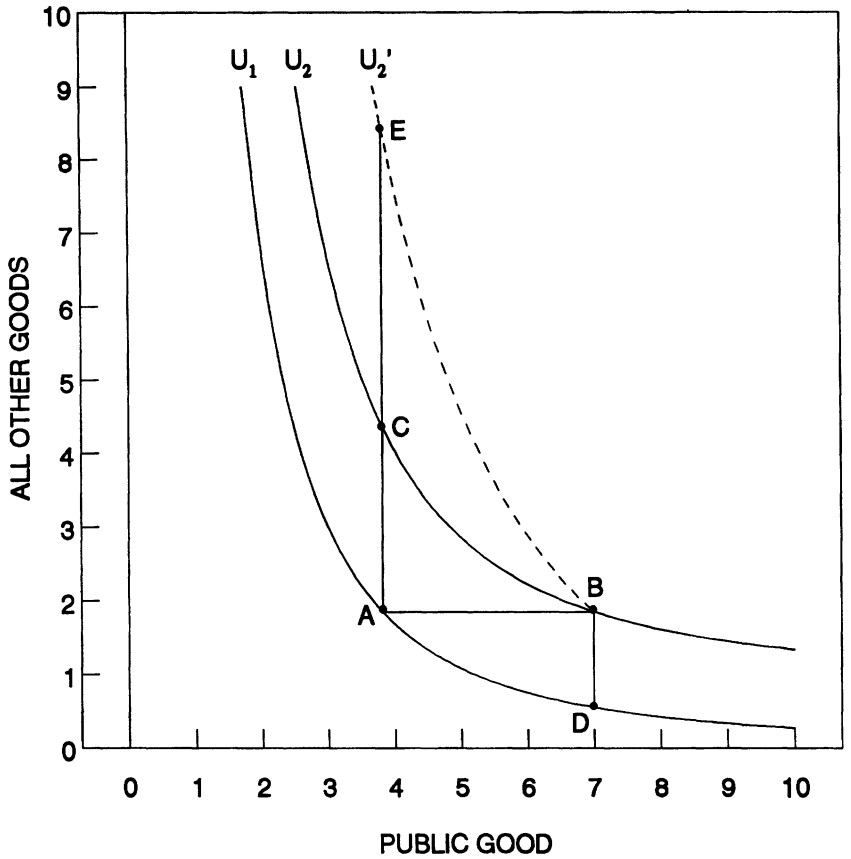


FIGURE 1
WELFARE EFFECT OF A ZERO-PRICE QUANTITY CHANGE IN A PUBLIC GOOD

loss aversion (Kahneman and Tversky 1979; Kahneman, Knetsch, and Thaler 1990; Kahneman, Knetsch, and Thaler 1991; Tversky and Kahneman 1991). Used in this way, the PC method provides an alternative approach to CV.[2]

II. THEORETICAL BACKGROUND

Neoclassical consumer theory argues that an ordinal utility function can describe a person's preferences.[3] Necessary assumptions include the transitivity axiom, that is, a consumer who prefers A to B and B to C, will prefer A to C. If we assume that monetary valuation of public goods does not violate the axioms of utility theory, we can invoke the standard consumer theory (Deaton and Muelbauer 1957; Freeman 1993; Just, Hueth, and Schmitz 1982), as in Figure 1 ignoring the dotted line. The vertical axis in Figure 1 mea-

[2] In related applications of PC to estimate economic values, Magat et al. (1988) obtained *ratings* of preference between pairs of versions of a consumer item that varied in both price and risk of personal injury from its use, and Viscusi, Magat, and Huber (1991) obtained choices between pairs of alternative living places that differed in both cost of living and health risk. Responses allowed estimates of willingness to pay to avoid the risks. Hey and Orme (1994) used a computerized PC experiment to investigate generalizations of expected utility theory. Loehman and Hu De (1982) studied the relationship between changes in a public good and changes in income by means of a payment card, which they interpreted as an application of the PC method.

[3] Heiner (1985) (see also Friedman 1985; Smith 1985) offers an interesting critique of classical preference theory and the generalizability of experimental results.

sures the monetary value of the consumer's endowment of private goods, including money, and the horizontal axis measures the public good endowment. U_1 and U_2 are indifference curves describing the trade-off between public and private endowments.

Assume a consumer at point B on U_2 faces a loss in the endowment of public goods to A on U_1. According to neoclassical consumer theory, AC measures WTA to accept the loss AB. Similarly, if a consumer at point A on U_1 faces a choice between a zero-price increase in the endowment of public goods to B on U_2 and a zero-price increase in other goods (and/or money) to C on U_2, AC measures WTA to forego the gain AB.

Discovery of indifference curves is a difficult task. Assume, however, that we can define a set of incentive compatible choices that includes the public good(s) in question, several sums of money, and perhaps some familiar private goods as well. Inclusion of several well chosen goods in addition to the target good(s) should increase the likelihood that the respondent will think more carefully about the characteristics and relative worth of the target good(s) in the course of addressing the required comparisons. It also helps avoid the human tendency to exaggerate the importance of an object or event to which attention has been drawn. It is then only necessary to order the consumer's preferences among the elements of this set in order to bound either WTP or WTA for the target good(s), depending on the choice question asked. If we select the other goods and sums of money judiciously, we can capture the target good(s) within narrow boundaries, thereby obtaining a reasonably precise estimate of WTP or WTA.

Given a set of t objects, the PC method presents them independently in pairs as $(t/2)(t-1)$ discrete binary choices.[4] The pairs should be randomly ordered for each respondent to control for order effects. The respondent simply chooses the preferred item in each pair.[5] Even if indifferent between the two items in a pair, the respondent must still make a choice.[6] As with dichotomous choice CV, the decision task is simpler than with open-ended CV because the respondent need only recognize preference, not state a monetary amount. Multiple responses by each individual allow a test of reliability.

The method also allows investigation of the transitivity axiom, because pairwise comparison of elements of a choice set reveals inconsistent choices as circular triads, that is, choices that imply A > B > C > A. A circular triad may have one of several causes, including systematic intransitive preference, random choice in cases too close to call, incompetence of the respondent, or simple mistakes. We therefore differentiate in this paper between (1) intransitivity, defined as systematic and repeatable decision behavior that violates the transitivity axiom, and (2) inconsistency, as indicated by non-repeatable circular triads. To test the transitivity axiom, one must design an experiment that allows the causes of observed circular triads to be identified.

The PC method can be applied to either choices between losses or to choices between gains. When asked to choose between losses, the respondent is at point B in Figure 1,

[4] When the number of objects is large, various methods can be used to reduce the number of choices. See, for example, Green and Srinivasan (1978) on conjoint analysis.

[5] If $U(Q)$ represents the utility of the respondent's current endowment Q, the paired comparison choice is essentially $U(Q, x_i) > < U(Q, x_j)$, where x_i and x_j represent two items from the choice set $x_1, \ldots, x_t$.

[6] Whether to allow the respondent an indifference option or to require a choice is debatable. One argument is that indifference is a valid behavioral response and ought to be measured. Allowing the indifference option might therefore increase generality and give added insight into switching behavior and the occurrence of circular triads. We argue, however, that an indifference option allows the respondent to be lazy with close calls, that is, to choose "indifference" in cases where discernment of preference is possible, thus reducing the amount of information obtained. Forcing a choice in all cases maximizes discernment of difference while revealing indifference stochastically. Across respondents, or across repetitions of the choice for the same respondent, the expected effect of indifference is an equal number of selections of each item in the pair. The requirement of a choice is similar to dichotomous choice contingent valuation where the respondent is allowed only two options, "yes" or "no." Regarding application of the "law of comparative judgment" in the design of PC experiments, Torgerson (1958) states, "No equality judgments are allowed. This is consistent with the derivation of the law, wherein the probability of a zero discriminal difference is vanishingly small."

where BD and AB represent losses to which the respondent is indifferent. To estimate WTP to avoid the loss AB in the public good endowment, PC would offer choices between the loss AB and several monetary losses. These choices reveal the breakpoint at which the respondent switches from choosing the loss AB to choosing a monetary loss, thus yielding an estimate of WTP to avoid the loss. Similarly, when asked to choose between alternative gains, the respondent is at point A in Figure 1, where AC and AB represent gains to which the respondent is indifferent. To estimate WTA to forego the gain AB in the public good endowment, PC would offer choices between the gain AB and several monetary gains.[7] These choices reveal the breakpoint at which the respondent switches from choosing the gain AB to choosing the monetary gain. When respondents are asked to choose between alternative gains, WTA is said to be estimated from the chooser reference point.

WTA > WTP if, as depicted in Figure 1, there is an income effect of the change in public good endowment. Experimental estimates of WTA from the seller reference point generally exceed WTP, however, by an amount greater than any reasonable income effect (Fisher, McClelland, and Schulze 1988; Gordon and Knetsch 1979; Kahneman and Tversky 1984; Kahneman, Knetsch, and Thaler 1991; Knetsch 1984; Tversky and Kahneman 1981). One plausible explanation is that the pain of loss is greater than the pleasure of gain, other things being equal. If so, for a consumer at point B, reduction of the public good endowment by an amount equal to AB can only be compensated by an increase in all other goods greater than AC, such as AE in Figure 1. Under this theory, the correct measure of minimum WTA is AE, not AC, and indifference curve U_2 has a discontinuity at B.[8] Kahneman, Knetsch, and Thaler (1990) found that WTA for sellers (those at point B in Figure 1) was considerably greater than what we define as WTA for choosers (those at reference point A). We hypothesize that loss aversion does not occur in decisions made from the chooser reference point, and that an estimate of WTA from the chooser reference point should therefore be more conservative than an estimate from the seller reference point.[9]

The PC method brings to the economic valuation task more than a hundred years of psychometric research on measurement of preferences for subjective stimuli. An extensive literature provides a theoretical foundation, experimental design guidelines for achieving validity, and rigorous methods for testing hypotheses and estimating parameters (Bock and Jones 1968; David 1988;

[7] Where respondents are asked to consider their preferences for two alternatives that differ both in monetary cost and in other attributes (as in Magat, Viscusi, and Huber (1988) and Viscusi, Magat, and Huber (1991), respondents are essentially asked to choose between two *net* gains (the utility of each alternative net of its respective cost). Because—depending on the cost magnitudes used—a net gain may be negative (i.e., a net loss), the respondent may be asked to choose between a net gain and a net loss, or even between two net losses. Such choice data can be used to estimate the monetary value of the attributes, but the monetary value may reflect a mixture of WTA (from the chooser reference point) and WTP. The mixing of WTA and WTP should not matter if income effects are nil.

[8] It is "correct" in the sense that AE measures the true monetary value of the de facto felt loss under the loss aversion theory. Whether the consumer loses a de jure property right when the public good endowment changes from B to A is an issue beyond the scope of this paper.

[9] We do not argue that a chooser reference point is a more appropriate stance than the seller reference point from which to measure WTA for purposes of public policy. Some authors argue that WTA is, by definition, a behavioral phenomenon, not a theoretical construct. If loss aversion is real, for example, it is by argument of consumer sovereignty a legitimate part of WTA (Kahneman, Knetsch, and Thaler 1990). Economic theory, they say, is based on a premise of global rationality that is not descriptive of the bounded rationality of human choice behavior (Simon 1985). This question remains controversial among many economists, however, who argue that we must define WTA by economic theory, that observed differences between WTA and WTP beyond theoretical limits are artifacts of poor measurement, not valid assigned values, or that there is no legal property right to the added value. We make no attempt here to resolve these controversies, and only assert that valuation from the chooser reference point may yield a more conservative estimate of WTA. That is, the estimate from the chooser reference point should be more conservative than an estimate from the seller or loser reference point. Without further experimentation, however, we do not know whether the chooser reference point produces an estimate of WTA that is conservative relative to the actual difference in income effect between WTA and WTP.

Edwards 1957; Fechner 1860; Guilford 1954; Kendall and Gibbons 1990; Thurstone 1927; Torgerson 1958). However, validity and reliability in any application, including application to economic valuation, depend largely on the specific scenario of a given experiment. Content, construct, and criterion validity of PC responses to hypothetical economic choices remain untested. This paper does not directly test the validity of PC; rather, it takes a small step by exploring reliability and transitivity.[10] We have not found previously published applications of the method to economic choices except as we have referenced herein.

Various methods are available for summarizing the preference information contained in the PC data. The simplest approach is to accept the responses on face value and use the number of times each item in the choice set is preferred over other items as an ordinal measure of preference. If monetary amounts are among the items and if they were judiciously chosen, values for the goods can be bracketed by monetary amounts in the preference order. Then linear interpolation within the brackets by mean or median preference scores can be used to value the items.

More rigorous methods of analysis are available, however. Given a random error process, such as a random discriminal dispersion that describes intrapersonal responses or random interpersonal variation in response, various stochastic methods can be used to estimate preference order and monetary bids. Under certain assumptions explained in the psychometric literature, it is possible to test specific stochastic hypotheses and derive an interval scale of preference. Inclusion of monetary magnitudes in the choice set anchors the scale and gives it ratio properties.[11] The responses obtained by PC also lend themselves to stochastic investigation by economic methods of discrete choice analysis, such as binary logit and double-bounded logit analysis, as well as by psychometric scaling techniques and simple heuristic arithmetic calculation.[12] Discrete choice methods can estimate bid functions for trade-off relationships among goods in the choice set, as well as for trade-off relationships between goods and dollars, depending on the objectives of the study and the composition of the choice set.[13] When the experimental design includes multiple sums of money, each respondent makes multiple binary choices between the good(s) and sums of money. Multiple random samples of one discrete choice from each respondent provide a set of conventional discrete choice experiments that can be used to estimate dispersions and confidence intervals.

Comparison of the various approaches to estimating monetary values from PC data is a substantial undertaking and beyond the scope of this paper. The purpose here is to

[10] Loehman and Hu De (1982) argue that PC as utilized in marketing and psychological studies "is considered to be a reliable method for obtaining responses since the choices required are relatively simple." They also cite Stouffer et al. (1950) and Shaw and Wright (1967) as evidence that responses to questionnaires of this type are "reasonably valid." Peterson, Bishop, and Michaels (1973) provide evidence of criterion validity of aggregate PC results obtained in a hypothetical context. They compared the probability of children's first choice of playground equipment predicted from PC responses to photographs with the first choice probability measured by unobtrusive time lapse photography observation of actual behavior in the real playground from which the photographs were taken.

[11] Kendall and Gibbons (1990) and David (1988) describe the applicable probability theory and statistical tests. Edwards (1957), Bock and Jones (1968), Guilford (1954), and Torgerson (1958) explain the analytical methods and underlying assumptions for interval scale estimation. Maxwell (1974) provides a simplifying analytical procedure based on the logistic transformation.

[12] The psychometric theory of PC has an economic counterpart in utility-maximizing discrete choice theory. Luce (1959, 1977) formalized Arrow's (1951) "independence of irrelevant alternatives" (IIA) assumption into a choice axiom. This model has been shown to be essentially equivalent to Thurstone's (1927) "law of comparative judgment" (Case V), if Thurstone's assumption of independent, normally distributed random variables is replaced by one of double-exponential, random disturbances (McFadden 1973; Yellott 1977). The difference distribution of two independent double-exponential random variables is the logistic distribution, which is the basis for the logit model. And Thurstone's assumption of normally distributed random variables leads to the probit model.

[13] Given a sufficient number of items in the set, it is also possible to apply discrete choice within each individual as well as across a pooled set of individuals.

investigate the reliability and transitivity of PC responses. Work in progress is exploring the application of stochastic discrete choice methods, comparison of psychometric and econometric estimation methods, and comparison of PC and CV estimates, among other things.[14] In this paper, we illustrate mean and median value estimation by two simple arithmetic methods: (1) from empirical observation of WTA as the number of respondents who reject each dollar offer for each good, and (2) from the monetary bracketing of goods in the individual respondent preference orders.[15]

As with any method, PC requires effective specification of the goods for which we require choices, or in the words of Arrow et al. (1993), "Adequate information must be provided to respondents about the environmental program that is offered. It must be defined in a way that is relevant. . . ." As stated by Lipman (1991), "given an agent's perception of his world, we must assume that he chooses what he perceives to be best for him." Herein lies a formidable challenge for any method, that is, to create through the information scenario the perception to which we desire a response and to attribute the response obtained to a correct description of the agent's perception, that is, to achieve content validity. PC incurs an additional challenge in this regard when more than one good must typically be described to respondents.[16]

As with CV, researchers who apply PC to valuation of public goods must beware of strategic behavior, hypothetical bias, context effects, and other plausible disturbances that might bias the results (Mitchell and Carson 1989). To be valid, the PC presentation must be incentive compatible. For example, we cannot be certain that free-riding behavior does not affect PC choices that involve public goods. Our evaluation of transitivity and reliability of PC response is a small step toward better understanding of the method and of economic choices in general, but it does not answer these broader questions. We believe, however, that the promising results reported in this paper encourage further research.

III. EXPERIMENTAL DESIGN

The choice set in this experiment consists of six locally relevant public goods, four private goods, and eleven sums of money. Each respondent made 155 choices between public goods, between private goods, between public and private goods, between public goods and sums of money, and between private goods and sums of money. The respondents did not choose between sums of money. Three hundred thirty students at Colorado State University participated in the study for a total of 51,150 binary choices.

Table 1 lists the goods and gives brief descriptions. The four private goods are familiar market goods with suggested retail prices. They were included to encourage respon-

[14] In addition to direct estimation of monetary values and the transitivity and reliability questions addressed in this paper, PC might be used to test or control other specific aspects of the value response, such as the embedding effect. For example, the items could be selected so as to include various levels of a good, such as the levels of environmental services at issue in Kahneman and Knetsch's (1992) embedding study. Having respondents choose among different levels that range from the most general to the most specific relevant level of a good might control the embedding effect. Such questions offer opportunities for further research beyond the scope of this paper.

[15] Although such methods simply accept the sample data on face value in a positivistic sense, they do allow statistical tests of hypotheses and calculation of confidence intervals, and they avoid the extensive assumptions hidden behind more sophisticated methods. We do not intend to suggest that such simple approaches are preferable or even acceptable. Our intent is simply to illustrate that preference order as measured by PC implies monetary value when the experiment includes sums of money. Further research is needed to compare and evaluate alternative ways to estimate monetary value from PC data.

[16] We recognize that the descriptions used for the goods in this study do not meet the rigorous requirements specified by Arrow et al. (1993). With the exception of the "Wildlife Refuge" and "Clean Arrangement," however, the respondents already are quite familiar with the goods used. Operational use of PC, like any application of CV, should include more thorough descriptions. However, we believe that use of more effective descriptions in this study would have reduced the variance among respondents with similar preferences, produced even fewer circular triads, and achieved greater reliability, thereby strengthening the already strong results.

TABLE 1
GOODS INCLUDED IN THE PAIRED COMPARISON EXPERIMENT

PRIVATE GOODS

1. A meal at a Fort Collins restaurant of your choice, not to exceed $15. (**Meal**)
2. Two tickets and transportation to *one* of the following:
   A) A Colorado ski area of your choice.
   B) A concert of your choice in Denver (Contemporary or Classical).
   C) A Broncos, Rockies, or Nuggets game.
   D) A cultural event of your choice at the Denver Center for the Performing Arts.
   Estimated value: $75 (**Tickets**)
3. A nontransferable $200 certificate for clothing at a Fort Collins store of your choice. (**Clothes**)
4. A nontransferable certificate for you to make $500 worth of flights on an airline of your choice. (**Air Travel**)

PUBLIC GOODS

1. A no-fee library service that provides videotapes of all course lectures so that students can watch tapes of lectures for classes they are not able to attend. (**Videotape Service**)
2. Parking garages to increase parking capacity on campus such that students are able to find a parking place at any time, without waiting, within a five-minute walk of any building at no increase in the existing parking permit fee. (**Parking Capacity**)
3. Purchase by CSU of 2,000 acres of land in the mountains west of Fort Collins as a wildlife refuge for animals native to Colorado. (**Wildlife Refuge**)
4. A CSU-sponsored, on-campus springtime weekend festival with a variety of live music and student participation events with no admission fee. (**Spring Festival**)
5. Expansion of the eating area in the Lory Student Center to ensure that any student can find a seat at any time. (**Eating Area**)
6. A cooperative arrangement between CSU, local business groups, and the citizens of the community that would ensure the air and water of Fort Collins would be at least as clean as the cleanest 1% of the communities in the U.S. (**Clean Arrangement**)

dents to consider a wide range of goods and trade-offs, to avoid inducing value by focusing too much attention on any one good, and to examine WTA for familiar private goods with suggested prices. The six "public goods" are of mixed type. "Wildlife Preserve" and "Clean Arrangement" are pure public environmental goods, that is, environmental goods that are nonrival and nonexcludable in consumption. The remaining four public goods—"Spring Festival," "Video Service," "Parking Capacity," and "Eating Area"—are excludable by nature but stated as nonexcludable by policy. They are also nonrival until demand exceeds capacity. Wildlife Preserve and Clean Arrangement benefit all people in the broader community, whereas the other goods benefit only the students. Respondents had Table 1 in front of them during the experiment and were free to refer to it at any time.

The eleven sums of money were $1, $25, $50, $75, and $100 through $700 in intervals of one hundred dollars. The public and private goods used in the experiment were derived from pilot studies in order to have good variation and distribution across the dollar magnitudes. Respondents were asked to choose one item or the other under the assumption that either would be provided at no cost to the respondent.

The experiment was administered by means of a computer code that presented the items on the monitor in random order for each respondent. The goods had short names which appeared side by side on the monitor, with their position (right versus left) also randomized. The respondent entered a choice by pressing the right or left arrow key and could correct mistakes by pressing "backspace." At the end of the 155 paired comparisons, the computer code repeated in random order those pairs for which the individual's binary choice was not consistent with the dominant preference order, as defined by the preference scores, that is, the number of times the

individual preferred each good over all other goods and sums of money (see the definitions of "preference score" and "preference order" in Section IV). The computer also randomly selected ten consistent pairs. The individual pairs in these two sets of repeated choices were randomly intermixed, and there was nothing to indicate to the respondent when the 155 original pairs ended and the repeats began.

The computer program recorded (1) the respondent's choice for each pair in an ordered matrix, (2) the time in seconds required for each choice, (3) the sequence number of each choice, (4) the pairs that were involved in circular triads, and (5) preference switches, if any, for the two types of repeated choices.

IV. DATA SUMMARIZATION METHODS

The Preference Score and Preference Profile

The paired comparison method yields a preference score for each item, which is the number of times the respondent prefers that item over other items in the set. A respondent's vector of preference scores, called the preference profile, describes the individual's preference order among the items in the choice set, with larger integers indicating more preferred items. In this case of a 21-item choice set, an individual preference profile with no circular triads contains all 21 integers from 0 through 20. Circular triads cause some integers to appear more than once in the preference profile while others disappear.

Value Contrast

In some of the analyses that follow, within-pair value contrast, that is, the difference in value assigned by an individual to the two items in a given paired comparison, is an important variable. It is convenient for certain analyses to use the integer difference between preference scores as an index of within-pair value contrast. This index will be an integer that ranges in magnitude from 0 to 20, because it is the difference of two integers from the same range. Although useful, such an index is approximate, because the preference scores are ordinal numbers as far as strength of preference is concerned; however, they are cardinal with respect to the number of times each item was preferred over other items.

Monetary Estimation

If appropriate monetary amounts are among the items in the choice set, the preference profiles allow estimation of dollar values for each of the goods for individual respondents by the simple arithmetic approach of bracketing by sums of money based on preference score and linear interpolation within the brackets.[17] The PC method also allows estimation of aggregate sample values by aggregation of individual dollar estimates as bracketed in the individual preference profiles, or by observing the proportion of respondents rejecting each dollar offer for each good. These proportions are sample estimates of WTA probabilities. In this paper we use two of these simple methods, aggregation of bracketed dollar values for individual respondents and observed bid rejection proportions, to calculate sample mean and median WTA values. Evaluation of such simple methods in comparison with more sophisticated approaches, such as psychometric scaling and discrete choice analysis, is the topic of another paper.

Coefficient of Consistency

The number of circular triads in each individual's responses can be calculated directly from the preference profile. The number of items in the set determines the maximum possible number of circular triads. The individual respondent's coefficient of consistency is calculated by subtracting the observed number of circular triads from the maximum number possible and dividing by

[17] The individual preference profile is the product of 155 choices and is highly reliable. Numerous circular triads in an individual's responses may prevent estimation of dollar values for the individual respondent, however.

the maximum.[18] The coefficient varies from zero to one. A coefficient of one means there are no circular triads in a person's choices. The coefficient approaches zero as the number of circular triads approaches the maximum possible number of circular triads. The maximum possible number is $(t/24)(t^2 - 1)$ when t is an odd number or $(t/24)(t^2 - 4)$ when t is even, where t is the number of items in the set (Kendall and Smith 1940, David 1988).

V. RESULTS

The data obtained by PC offer numerous avenues of exploration. This paper investigates the time respondents took to make their choices, evaluates the reliability of the responses, investigates the transitivity axiom and related decision behavior, and demonstrates estimation of monetary values for the goods by two simple methods.

Decision Time

Average total time to complete the 155 paired comparisons was about 10 minutes, not including the time required to become familiar with the goods and the instructions. Mean decision times for the first five choices were 29, 10.9, 8.8, 7.4, and 6.8 seconds, respectively. Mean decision time continued to decline gradually until it stabilized at about 2.5 seconds halfway through the experiment (Figure 2). A correlation of −0.97 between the log of mean decision time and the log of decision sequence over the 155 sequence numbers reveals a strong tendency for decision time to decline at a decreasing rate as the experiment progresses.[19]

The short average decision time raises a question about how carefully the respondents considered each choice. The short and declining decision time suggests several hypotheses, including: (1) respondents progressively construct or discover their values, (2) they quickly develop simplifying rules, or (3) they grow weary and careless as the experiment progresses. The care taken is partially indicated by the reliability and transitivity of their choices, which we discuss in following subsections. Here we present evidence based solely on decision time as it varies with the nature of the choice.

If carelessness were the cause of the short decision times, we would expect the likelihood of inconsistency to increase as the experiment progresses, and we would expect decision time to be less for inconsistent choices than for consistent choices. In fact, the opposite is true. Average sequence is 77.54 (out of 154) for consistent choices and 70.04 for inconsistent choices. Average decision time is 3.03 seconds for consistent choices and 4.71 seconds for inconsistent choices. Analysis of variance shows these differences to be significant at the 0.000 level in both cases. These numbers demonstrate that inconsistent choices tend to occur earlier in the experiment and require more decision time than consistent choices.

Except for the aforementioned relation of decision time to sequence, the primary relation involving decision time appears to be between decision time and within-pair difference in preference scores. Decision time declines at a slightly decreasing rate with increasing difference in preference scores (Figure 3). The correlation of log of mean decision time with log of mean difference is

---

[18] David (1988) gives the number of circular triads as:

$$c = \frac{t}{24}(t^2 - 1) - \frac{1}{2}\Sigma(a_i - a\bullet)^2$$

where $a\bullet = \Sigma \frac{a_i}{t} = \frac{1}{2}(t - 1)$,

t = the number of elements in the choice set, and
a_i = the number of elements in the choice set dominated by the ith element.

The maximum possible number of circular triads is:

$$k' = \left(\frac{t}{24}\right)(t^2 - 1),\ t \text{ odd},\ k' = \left(\frac{t}{24}\right)(t^2 - 4),\ t \text{ even}.$$

Kendall and Smith (1940) define the coefficient of consistency as:

$$Z = 1 - \frac{24c}{t(t^2 - 1)},\ t \text{ odd};\ Z = 1 - \frac{24c}{t(t^2 - 4)},\ t \text{ even}.$$

Note that $Z = (K' - c)/K'$.

[19] The disaggregate correlation over the 51,150 cases is −0.35.

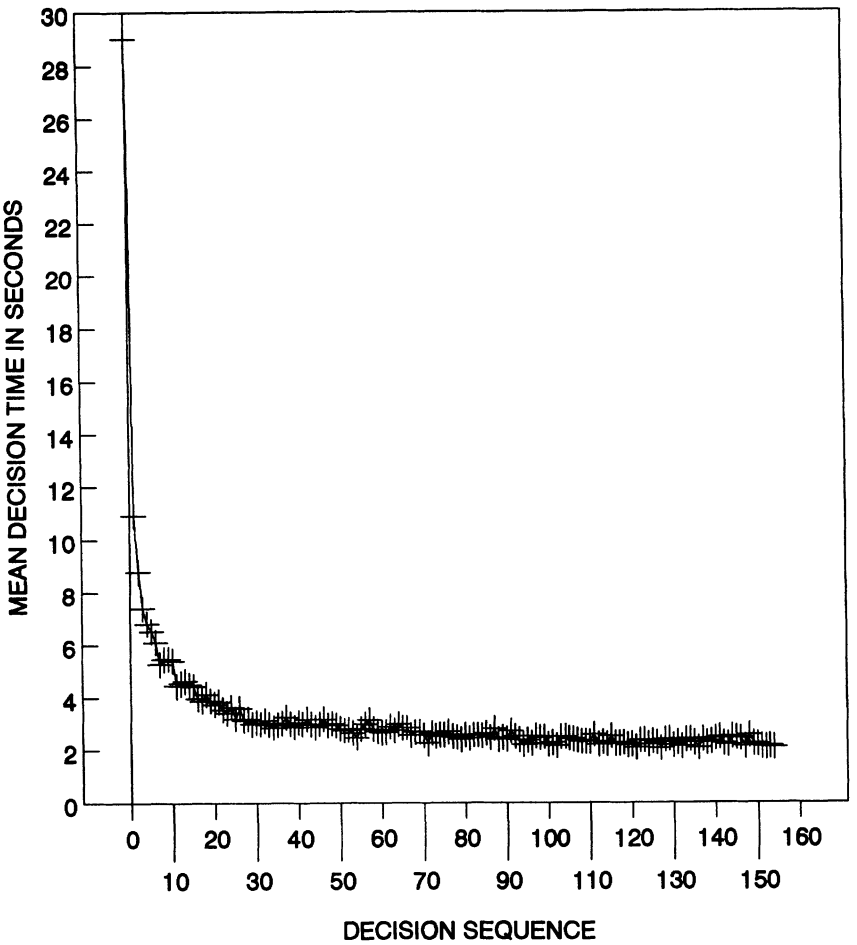


FIGURE 2
MEAN DECISION TIME VERSUS DECISION SEQUENCE

−0.967.[20] Decision time declines from about 5 seconds for choices with the least within-pair difference to about 2 seconds for the choices with the greatest within-pair difference, perhaps reflecting the degree of difficulty in making the choice.

Decision time was also found to vary with type of choice. Mean decision time was 3.80 seconds when a public good was compared with another public good, 3.55 seconds for a public–private good comparison, 3.39 seconds for a private–private good comparison, 2.96 seconds when a public good was compared with a monetary amount, and 2.92 seconds for a private good–money comparison. Independent t-tests comparing these five means indicated that the only ones not significantly different from the others at the 0.05 level are 3.55 versus 3.39 and 2.96 versus 2.92. The means suggest that comparisons involving monetary amounts are easiest and comparisons between public goods are the most difficult, with comparisons between goods involving private goods—which have market prices—falling between these two

[20] This correlation is significant at the 0.001 level. The corresponding mean correlation between log of individual decision time and log of individual difference is −0.231, and is also significant at 0.001.

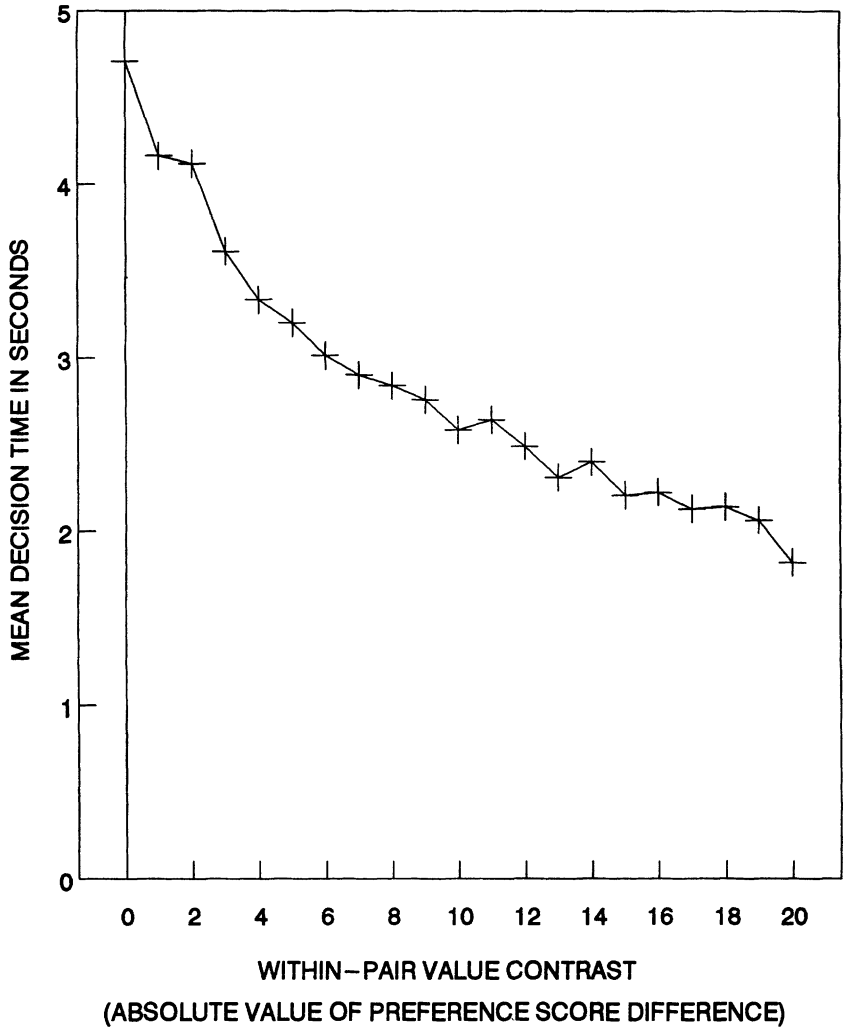


FIGURE 3
MEAN DECISION TIME VERSUS WITHIN-PAIR VALUE CONTRAST

extremes. This finding is consistent with the possibility that respondents convert each good to a monetary value before making a choice, with the conversion being easier for private than public goods—but of course, this is merely conjecture. Finally, we note that a two-way analysis of variance of decision time by within-pair difference in preference score and type of comparison indicates that both preference-score difference and type of comparison significantly vary with decision time ($p < 0.000$), and that the two-way interaction is not significant ($p = 0.524$).

The economics literature discusses the effect of decision cost on choices, and has sometimes considered decision time to be an indicator of decision cost. For example, Wilcox (1993) experimentally measured decision cost by measuring decision time and concluded that "decision time is a potentially rich explanatory and dependent variable, and so should not be an omitted one" (p. 1416). Our data do not necessarily indi-

cate that decision time was considered by respondents as a cost, as we have no indication that the quality of the choices deteriorated with time to make them. A more plausible conclusion, based on our findings, is that decision time simply increased with difficulty of the decision, with difficulty being a function of within-pair closeness in preference space of the items being compared, lack of a monetary referent for some of the goods, and lack of respondent familiarity with exchanges between money and public goods.

Reliability

We estimate the expected reliability of binary choices by examining choice switching behavior. Recall that respondents repeated inconsistent choices and a random sample of ten consistent choices after completion of the initial set of paired comparisons. Of the total set of 51,150 initial choices, 3,688 were inconsistent and 47,462 were consistent with the respondent's dominant rank order. The respondents switched 290 of 3,270 consistent choices and 2,256 of the 3,688 inconsistent choices on retrial, yielding respective reliability estimates of 0.91 and 0.39. The estimated overall expected reliability of a single choice selected at random is 0.87, assuming the same proportion of consistent and inconsistent choices.

Interpersonal correlations over the ten goods provide a lower bound on the expected reliability of the individual preference profiles. If the profiles of different respondents differ only by an independently and identically distributed random error process, the correlations among individual preference profiles constitute a domain sample basis for estimating reliability (Nunnally 1967). The interpersonal correlation is a lower bound, however, because both interpersonal and intrapersonal variance reduce the observed correlation. The maximum, mean, median, and mode of the observed distribution of 54,285 correlations among the 330 respondent preference profiles are 0.98, 0.66, 0.70, and 0.80, respectively. Removal of interpersonal variance would yield much higher reliabilities. To obtain an estimate of the upper bound on preference profile reliability, we had one respondent repeat the experiment twice. Another repeated it five times over an eight-month period. The resulting two estimates of reliability are 0.94 and 0.98, respectively.[21]

Transitivity

The data obtained in this experiment provide several ways to investigate transitivity: (1) consistency of individual responses, (2) the effect of value contrast on consistency, (3) the relationship between consistency and choice switching behavior, and (4) the relationship between choice switching and value contrast.

Consistency of individual responses. The coefficient of consistency as defined above describes the individual respondent's internal consistency as the percent of possible circular triads that are non-circular. Inconsistency does not necessarily imply systematic and repeatable intransitivity, however, because as previously noted, simple mistakes and random response due to indifference or incompetence will also produce circularity. A choice set of very similar elements can produce a low coefficient of consistency by chance alone. The coefficient of consistency is therefore a function of the degree of similarity of items in the choice set, incompetence of the respondent, the tendency for the respondent to make careless mistakes, and systematic intransitivity of response, as noted by David (1988, 3–4):

> a circular triad denotes an inconsistency on the part of the judge, and its simplest explanation is that the judge is at least partially guessing when declaring preferences. The judge may be guessing because of incompetence or because the objects are in fact very similar. . . . But guessing is not the only explanation, for there may be no valid ordering of the three objects even when they differ markedly. Their merit may depend on more than one characteristic, and it is then somewhat artificial to attempt an ordering on a linear scale. Under these circumstances the judge must mentally construct some function of the relevant characteristics and use this as a basis for comparison.

[21] These estimates give an upper bound because they remove interpersonal variance but may not be independent.

It is not surprising that in complicated preference studies the function is vague and may change from one paired comparison to the next, especially when different pairs of objects may cause the judge to focus attention on different features of the objects. This last point helps to account for situations where a particular circular triad occurs frequently in repetitions of the experiment. However, circularity can occur even with a well-defined preference criterion based on two or more underlying dimensions (Tversky 1969).

Figure 4 compares the distribution of the coefficient of consistency for 150 random trials of the experiment—where, in each of the 150 trials, each of 155 choices was random, that is, 23,250 random choices—with the distribution for the 330 respondents. If the 330 observed responses are random, they will have been sampled from the same population as the 150 random experiments and will have the same expected number of circular triads as the random cases. However, if the 330 responses contain nonrandom information, they will have a lower expected number of circular triads and will have come from a different population.

The two samples are obviously from different populations, because the respondents tend to be highly consistent and nonrandom. Twenty people (6 percent) in the lower tail of the real respondent distribution produced 25 percent of all circular triads, which suggests that they may not have understood the task or were different in other ways. The remainder (310) had a coefficient of consistency of at least 82 percent, and 50 percent (165) had a coefficient of consistency of at least 94 percent. Pooling all cases gives an overall coefficient of consistency of 92 percent.

The effect of value contrast on consistency. Figure 5 shows the relationship between the within-pair value contrast and the likelihood that a given choice will not agree with the individual's dominant preference order. The likelihood is computed as the ratio of the number of inconsistent choices to the total number of choices at the given level of value contrast. The total number of choices at each level of within-pair value contrast is listed across the top of Figure 5. For example, 1,797 pairs had zero value contrast, meaning that the respondents were apparently indifferent. Of these, 866 pairs (48 percent) were not consistent with the individuals' dominant preference order.

Under low value contrast we would expect the individual to find the choice more difficult. Under high value contrast we would expect easier choices and greater certainty. Figure 5 shows that the likelihood of inconsistent choice falls off rapidly as within-pair value contrast increases. At a value contrast of 10 or more, fewer than 1 percent of the choices were inconsistent. Low value contrast thus appears to be the primary cause of inconsistency.

The relationship between consistency and choice switching behavior. Under pure indifference or uncertainty, the expected probability that respondents will switch preference (i.e., reverse the choice when it is presented again) is 0.5. Under pure certainty without mistake (i.e., perfect repeatability), the expected switch probability is 0. If the respondent made a mistake with a choice for which preference is certain and perfectly repeatable (i.e., mistakenly hit the wrong key), the expected switch probability approaches 1.0.[22] It is possible, however, that some apparently consistent choices are cases of indifference or uncertainty that are consistent only by chance. Under the hypothesis that indifference, uncertainty, or mistake cause inconsistent choice and certainty moderated by chance causes consistent choice, we would expect reversal of 50 percent or more of inconsistent choices and 0 percent, or close to 0 percent, of consistent choices.

Figure 6 shows the distribution over respondents of the proportion of originally consistent and inconsistent choices switched on retrial. For example, each respondent repeated 10 randomly selected consistent choices. If a given respondent reversed three of these choices on retrial, the proportion switched is 0.3. The number of inconsistent choices repeated was not a fixed number, but varied over respondents. If a person made 18 inconsistent choices and reversed 12, the proportion switched is 0.67.

[22] The probability will be close to but less than one because of the possibility of repetition of the mistake. The joint probability of making the same mistake twice is small, however.

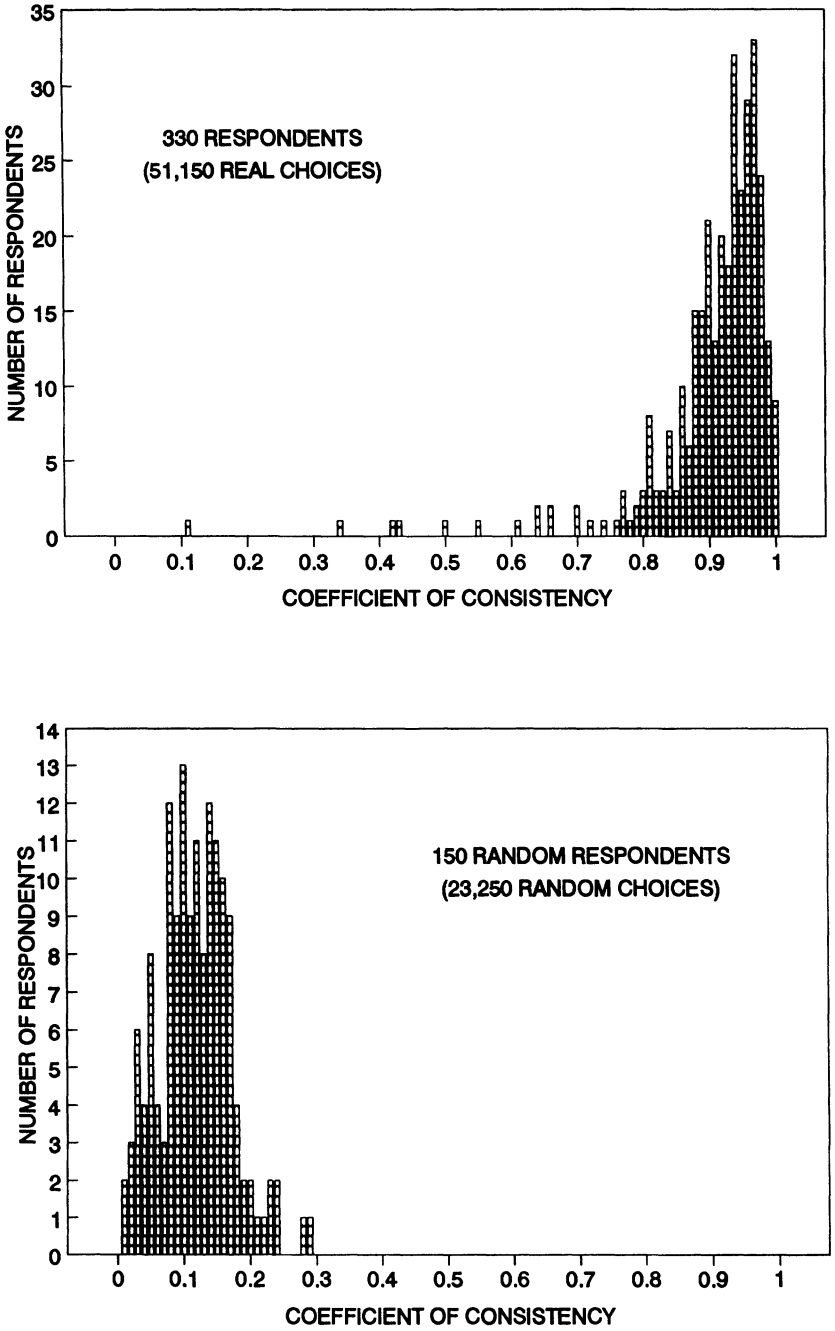


FIGURE 4
COMPARISON OF RESPONDENT CONSISTENCY WITH CONSISTENCY OF RANDOM CHOICES

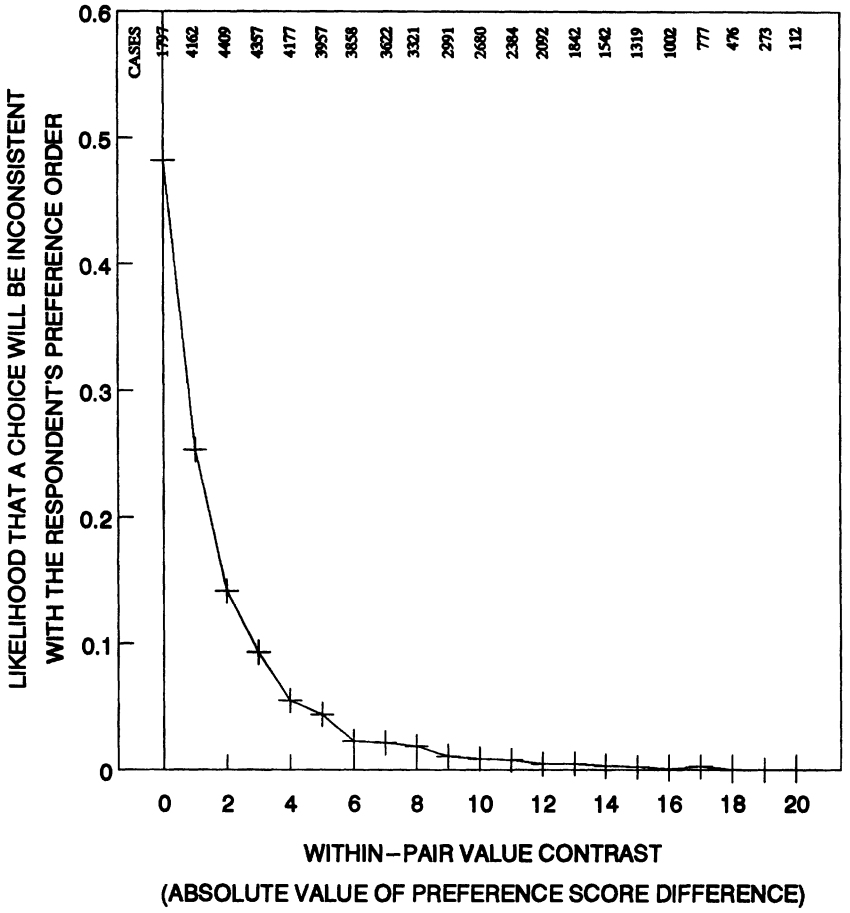
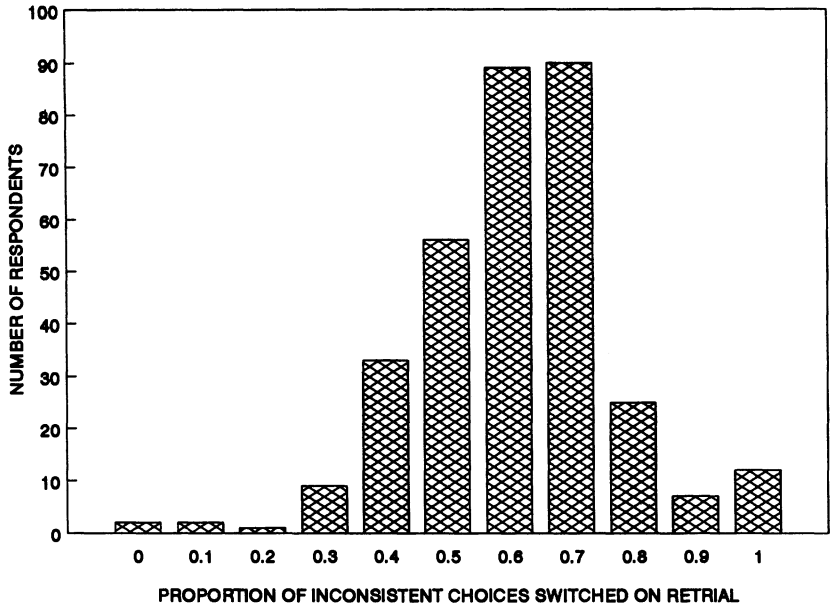


FIGURE 5
INCONSISTENT CHOICE AS A FUNCTION OF WITHIN-PAIR VALUE CONTRAST

For 3,688 originally inconsistent choices, the overall proportion switched is 0.612. For 3,270 originally consistent choices, the overall proportion switched is 0.089. The probability that these two distributions could have been sampled from the same population by chance alone is virtually zero. These findings further support the hypothesis that indifference and mistakes cause inconsistency, whereas certainty (i.e., repeatable preference) strongly influences consistency. That is, the likelihood that a consistent choice will be switched on retrial is reasonably close to zero, whereas the likelihood that an inconsistent choice will be switched is greater than 0.5.

Effect of value contrast on choice switching. Figure 7 shows how the proportion of responses switched changes with value contrast for both originally consistent and originally inconsistent choices. The switch probability declines sharply toward zero with increasing value contrast for consistent choices but increases sharply toward one for inconsistent pairs. At each level of value contrast except zero the probability of switching is significantly different between consistent and inconsistent choices.[23] At zero contrast the probability is close to 0.5 for both catego-

[23] The one-tailed significance of the difference is <0.0000 for all levels of original value contrast from 1 to 15, based on a normal approximation to the binomial test for value contrasts from 1 to 8 and the Fisher exact probability test for value contrasts from 9 to 15. At value contrasts of 16 and 17, the originally inconsistent cases number only 1 and 2, respectively, making the test unreliable.



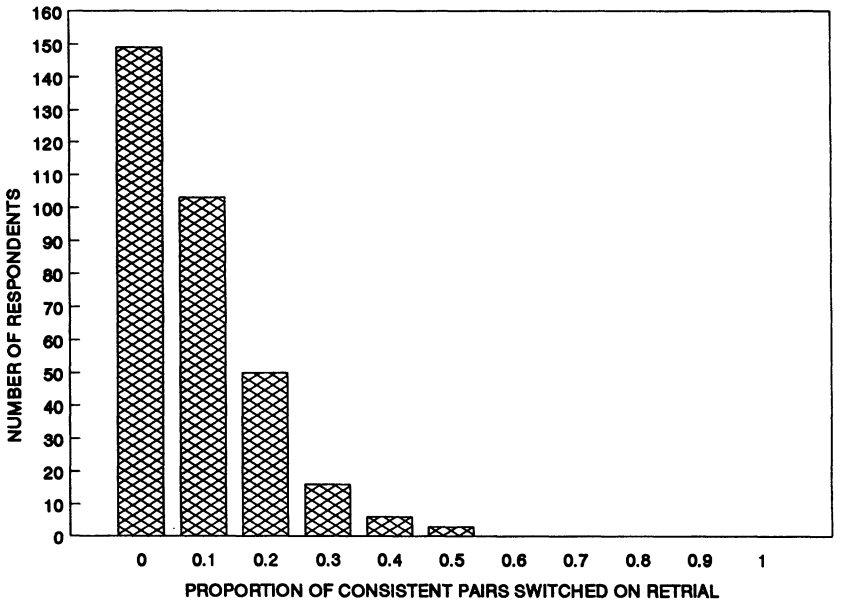


FIGURE 6

DISTRIBUTION OVER RESPONDENTS OF THE PROPORTION OF REPEATED CHOICES SWITCHED ON RETRIAL

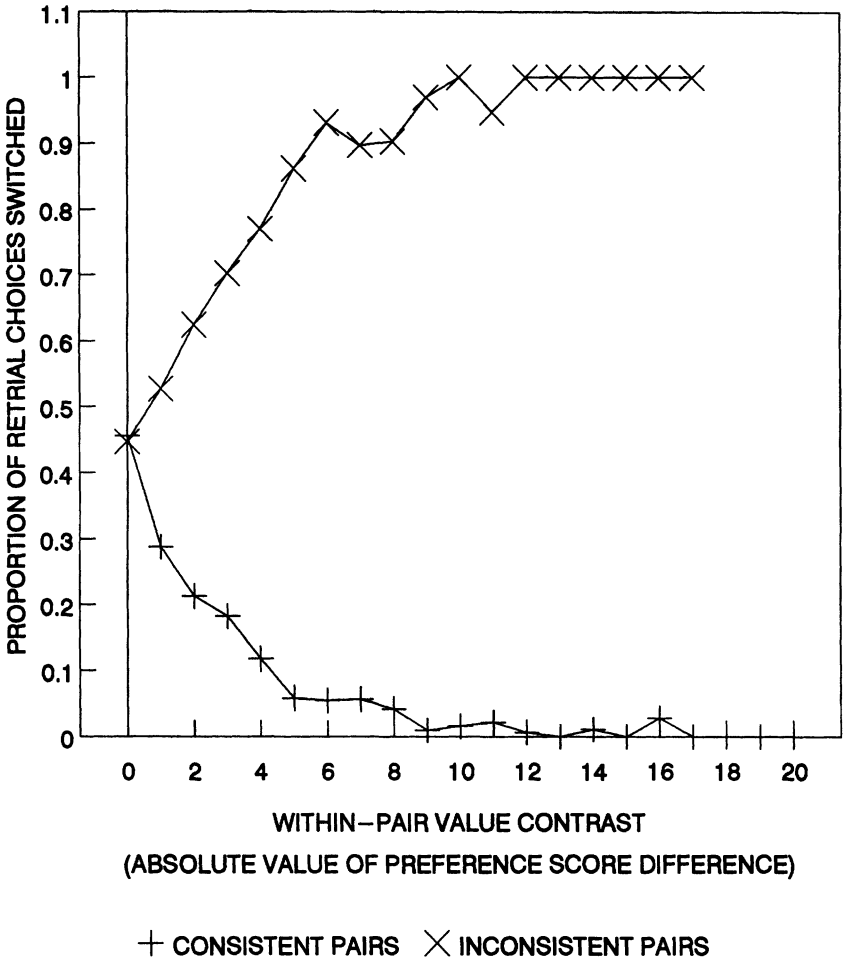


FIGURE 7
PROPORTION OF CHOICES SWITCHED ON RETRIAL AS A FUNCTION OF WITHIN-PAIR VALUE CONTRAST

ries. These results imply that respondents tend to correct mistakes when repeating inconsistent pairs and apply consistent decision criteria when repeating consistent pairs, with a strong tendency for low value contrast to introduce random choice.

Preference Ordering and Monetary Estimates

As previously stated, there are several ways to use PC data to estimate monetary values. David (1988) derives combinatorial and nonparametric methods from rigorous probability theory to test specific hypotheses about PC results. Torgerson (1958), Bock and Jones (1968), Luce (1959), Bradley (1976, 1984), van Putten (1982), and Winsberg and Ramsey (1981) derive and demonstrate various models and psychometric methods for testing hypotheses and estimating parameters. Maximum likelihood methods derived from utility-maximizing discrete choice theory also can be applied to PC data and are well documented in the econometric literature.[24] The fact that each respondent makes multiple discrete choices between

[24] For example, see Loehman and Hu De (1982) for an application of stochastic discrete choice analysis to data obtained by a method similar to PC.

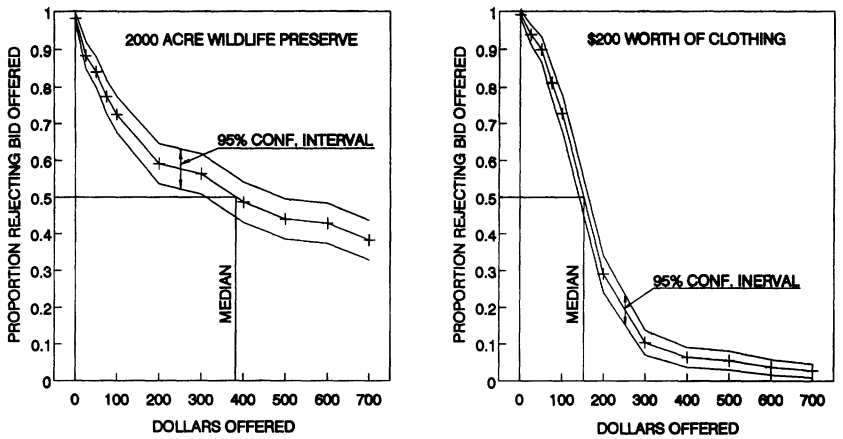


FIGURE 8
EXAMPLES OF THE EMPIRICAL RELATIONSHIP BETWEEN DOLLARS OFFERED AND WTA

goods and several sums of money also allows dispersion and confidence intervals to be observed by random sampling of conventional single-response discrete choice experiments from PC data. The need to apply, compare, and evaluate these various approaches with economic applications of PC opens a new frontier of research opportunity but is beyond the scope of this paper, which focuses on reliability and transitivity of response. In this paper, we illustrate monetary estimation by calculation of aggregate means and medians (1) by bracketing the goods between dollar amounts in the individual preference profiles, and (2) from the observed proportion of respondents rejecting each dollar amount for each good. Figure 8 shows proportions from the latter method for one public good (a 2,000-acre wildlife preserve) and one private good ($200 worth of clothing). These proportions are sample analogues of the probability functions that could be estimated by discrete choice logit analysis.

Table 2 displays an individual preference profile and monetary estimates for one typical respondent. Table 3 shows mean and median estimates for the ten goods as calculated by the two methods described above. The medians in column (1) of Table 3 are obtained by linear interpolation of the proportions illustrated by the two examples in Figure 8. The medians in column (2) come directly from the distribution of individual dollar values estimated from individual preference profiles. The means in column (3) were calculated by averaging individual dollar estimates. The means in column (4) came

TABLE 2
PREFERENCE SCORES AND MONETARY ESTIMATES FOR A TYPICAL RESPONDENT

Item[a]	Preference Score	Estimated $ Value
$700	20	700
$600	18	600
$500 air travel	17	550
Clean air arrangement	17	550
$500	16	500
Video tape service	14	450
Wildlife refuge	14	450
$400	13	400
$200 worth of clothing	12	350
$300	12	300
$200	10	200
Parking capacity	9	150
$100	8	100
$75	7	75
$50	6	50
$75 entertainment tickets	5	38
$15 meal	5	38
$25	4	25
Eating area capacity	3	17
Spring festival	1	9
$1	0	1

[a] See Table 1 for descriptions of the goods.

TABLE 3
ILLUSTRATIVE MEANS AND MEDIANS ESTIMATED BY TWO SIMPLE METHODS

	Medians		Ratio	Means		Ratio
	(1)	(2)	(2)/(1)	(3)	(4)	(4)/(3)
$500 Air travel	$422	$394	0.934	$380	$388	1.021
Wildlife preserve	$382	$372	0.974	$388	$394	1.015
Clean air arrangement	$197	$228	1.157	$323	$328	1.015
$200 Clothing	$152	$148	0.974	$182	$187	1.027
Entertainment tickets	$81	$90	1.111	$151	$154	1.020
Video tape service	$79	$83	1.051	$194	$198	1.021
Parking capacity	$74	$74	1.000	$167	$171	1.024
Spring festival	$62	$67	1.081	$150	$154	1.027
Eating area capacity	$29	$35	1.207	$92	$95	1.033
$15 Meal	$16	$21	1.313	$46	$48	1.043

(1) Interpolated from observed proportions.
(2) From individual preference profiles.
(3) Numerical integration of observed proportions.
(4) From individual preference profiles.

from numerical integration of the observed proportions. The two methods show very similar results.

Means and medians differ substantially in some cases because the distributions are highly skewed. Whether to use mean or median values in economic analysis becomes an important question when they differ significantly. In general, the median is the value that is acceptable to at least 50 percent of the sample and therefore identifies the range of values that a majority will accept. The mean is an estimate of the expected value of response and, when multiplied by the total number of respondents, measures the aggregate value in question. The reader must not generalize the values reported in this paper, however, because they are merely illustrative and do not necessarily represent any population beyond the sample observed. To generalize such values beyond the sample requires rigorous sample design and more rigorous examination of the estimates.

A further caution is suggested by the substantial disagreement among respondents with respect to preference order. Such disagreement implies aggregation error that may or may not be of concern in economic analysis. Regression on personal characteristics may reduce aggregation error. Unlike single point value CV estimates, PC preference profiles contain preference-order information that can be used to classify individuals into value response types, further reducing aggregation error. Such analyses, however, are beyond the scope of this paper.

Taking the data in Table 3 at face value, the two environmental public goods that provide benefit to the general population—"Wildlife Refuge" and "Clean Arrangement"—rank higher than the public goods that benefit only the student population. They also rank higher than all the private goods except "Air Travel."

Inclusion of private goods in the choice set offers potential information about validity of response and WTA when prices are well known or stated ex ante, as in Table 1. "Air Travel" and "Clothes," for example, have suggested values of $500 and $200, respectively. Their medians in Table 3 are approximately $400 and $150, respectively. Their mean values are also lower than the suggested retail prices, although not by as much. The other two private goods, "Tickets" and "Meal" have stated values of $75 and $15, while their medians are slightly higher and their means are much higher. These numbers are only illustrative of the methods used. If they are WTA values as we hypothesize, however, they produce means and medians for the higher-priced goods and medians for the lower-priced goods that are not several times larger than the suggested prices.

VI. CLOSURE

The information content of the binary choices observed in this experiment is extremely high. Specific choices, preference scores, and preference profiles are all very reliable. The observed degree of inconsistency among the binary choices is trivial compared to other errors and biases that might influence the results of any hypothetical contingent market experiment. Responses in general are nonrandom and highly consistent. Systematic intransitivity apparently is not a significant cause of inconsistency among the binary choices. The primary causes of inconsistency appear to be indifference and simple mistakes, as indicated by the following three findings. First, the likelihood of inconsistency declines rapidly with increasing value contrast between the two items in each choice. Second, inconsistent choices are much more likely to be switched on retrial than consistent choices. As hypothesized, the probability that a consistent choice will be switched is close to zero, whereas for inconsistent choices, the probability is greater than 0.5. Third, the proportion of consistent choices switched on retrial declines rapidly to zero with increasing value contrast, whereas the proportion climbs rapidly to one for inconsistent choices. We conclude that the psychometric method of paired comparison is a useful way to explore individual economic decision behavior and shows promise for application to contingent market estimation of the monetary value of nonmarket goods and services, including conservative estimation of WTA.

The richness of information obtained in this experiment and the ease of application encourage further research on the use of paired comparison for economic valuation of nonpriced goods and services. This research could proceed on three fronts. First, improvements in or extensions of what we did are warranted, including (1) looking beyond student populations, (2) testing more rigorous information scenarios for the goods in question, (3) evaluating alternative psychometric and econometric methods for estimating values from PC data, (4) exploring application to a wider variety of goods and services, (5) evaluating the independence of PC choices—for example, whether responses change when the full set of goods is not, as in the current application, listed on the same page, and (6) investigating applications of PC to the WTP context.

The second front is to explore content, construct, and criterion validity of the method, including (1) comparing PC results with more conventional contingent valuation methods and (2) testing PC against real monetary transactions. The third front involves application of the technique, which could occur at a minimum to (1) investigate the embedding (scope) effect and (2) explore application of PC to situations where a preference ordering among goods—gains or losses—is useful in and of itself. For example, as Rutherford et al. (1998) describe, recent national legislation covering oil spills calls for damage schedules to be used to assess damages of smaller (type A) spills. A schedule of relative (interval scale) damages might be developed using public judgments obtained in response to PC questions. Our results suggest that the relative placement of alternative losses is likely to be highly reliable. The attachment of monetary amounts to each item in the scale could then be performed separately, based on the best judgment of the court.

References

Arrow, K. J. 1951. *Social Choice and Individual Values.* New York: Wiley.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58 (10):4602–14.

Bock, R. D., and L. V. Jones. 1968. *The Measurement and Prediction of Judgment and Choice.* San Francisco: Holden-Day, Inc.

Bradley, R. A. 1976. "Science, Statistics, and Paired Comparisons." *Biometrics* 32:213–32.

———. 1984. "Paired Comparisons: Some Basic Procedures and Examples." *Handbook of Statistics* 4, ed. P. R. Krishnaiah and P. K. Sen. Amsterdam, Netherlands: North-Holland.

Cummings, R. G., and G. W. Harrison. 1994. "Was the Ohio Court Well Informed in Its Assessment of the Accuracy of the Contingent

Valuation Method?" *Natural Resources Journal* 34:1–36.

David, H. A. 1988. *The Method of Paired Comparisons.* London: Charles Griffin & Company.

Deaton, A., and J. Muellbauer. 1957. *Economics and Consumer Behavior.* New York: Cambridge University Press.

Edwards, A. L. 1957. *Techniques of Attitude Scale Construction.* New York: Appleton-Century-Crofts.

Fechner, G. T. 1860. *Elemente der Psychophysik.* Leipzig: Breitkopf and Härtel.

Fisher, A., G. H. McClelland, and W. D. Schulze. 1988. "Measures of Willingness to Pay versus Willingness to Accept: Evidence, Explanations, and Potential Reconciliation." In *Amenity Resource Valuation: Integrating Economics with Other Disciplines,* eds. G. L. Peterson, B. L. Driver, and R. Gregory. State College, PA: Venture.

Freeman, A. M., III. 1993. "The Measurement of Environmental and Resource Values: Theory and Methods." Washington, DC: Resources for the Future.

Friedman, D. 1985. "Experimental Economics: Comment." *American Economic Review* 75: 264.

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

Green, P. E., and V. Srinivasan. 1978. "Conjoint Analysis in Consumer Research: Issues and Outlook." *Journal of Consumer Research* 5: 103–23.

Guilford, J. P. 1954. *Psychometric Methods.* New York: McGraw-Hill.

Hey, J. D., and C. Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica* 62: 1291–1326.

Heiner, R. A. 1985. "Experimental Economics: Comment." *American Economic Review* 75: 260–63.

Just, R. E., D. L. Hueth, and A. Schmitz. 1982. *Applied Welfare Economics and Public Policy.* Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., and J. L. Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22 (1):57–70.

Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47:263–291.

———. 1984. "Choices, Values, and Frames." *American Psychology* 39:341–50.

Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98:1325–48.

———. 1991. "The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives* 5:193–206.

Kendall, M., and Jean D. Gibbons. 1990. *Rank Correlation Methods. Fifth Edition.* Suffock, England: Edward Arnold, Hodder & Stoughton.

Kendall, M. G., and B. B. Smith. 1940. "On the Method of Paired Comparisons." *Biometrika* 31:324–45.

Knetsch, J. L. 1984. "Legal Rules and the Basis for Evaluating Economic Losses." *International Review of Law and Economics* 4:5–13.

Lipman, B. L. 1991. "How to Decide How to . . . : Modeling Limited Rationality." *Econometrica* 59:1105–25.

Loehman, E. T., and Vo Hu De. 1982. "Application of Stochastic Choice Modelling to Policy Analysis of Public Goods: A Case Study of Air Quality Improvements." *Review of Economics and Statistics* 64:474–80.

Loehman, E. T., S. Park, and D. Boldt. 1994. "Willingness to Pay for Gains and Losses in Visibility and Health." *Land Economics* 70 (Nov.):478–98.

Luce, R. D. 1959. *Individual Choice Behavior.* New York: Wiley.

———. 1977. "The Choice Axiom after Twenty Years." *Journal of Mathematical Psychology* 15:215–33.

Magat, W. A., W. K. Viscusi, and J. Huber. 1988. "Paired Comparison and Contingent Valuation Approaches to Morbidity Risk Valuation." *Journal of Environmental Economics and Management* 15:395–411.

Maxwell, A. E. 1974. "The Logistic Transformation in the Analysis of Paired-Comparison Data." *British Journal of Mathematical and Statistical Psychology* 27:62–71.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics,* ed. P. Zarembka. New York: Academic Press.

Mitchell, R. C., and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Washington, DC: Resources for the Future.

Nunnally, J. C. 1967. *Psychometric Theory.* New York: McGraw-Hill.

Peterson, G. L., R. L. Bishop, and R. M. Michaels. 1973. "Children's Choice of Playground Equipment: Developing Methodology for Integrating Preferences into Environmental

Engineering." *Journal of Applied Psychology* 58:233–38.

Portney, P. R. 1994. "The Contingent Valuation Debate: Why Economists Should Care?" *Journal of Economic Perspectives* 8:3–17.

Rutherford, M., Knetsch, J. L., and T. C. Brown. 1998. "Assessing Environmental Losses: Judgments of Importance and Damage Schedules." *Harvard Environmental Law Review* 22:51–101.

Shaw, M. E., and J. M. Wright. 1967. *Scales for Measurement of Attitudes.* New York: McGraw-Hill.

Simon, H. A. 1985. "Human Nature in Politics: The Dialogue of Psychology with Political Science." *American Political Science Review* 79:293–304.

Smith, V. 1985. "Experimental Economics: Reply." *The American Economic Review* 75: 265–72.

Stouffer, S. A., et al. 1950. *Measurement and Prediction.* New Jersey: Princeton University Press.

Thurstone, L. L. 1927. "A Law of Comparative Judgment." *Psychology Review* 34:273–86.

Torgerson, W. S. 1958. *Theory and Methods of Scaling.* New York: John Wiley & Sons.

Tversky, A. 1969. "Intransitivity of Preferences." *Psychology Review* 76:31–38.

Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211:452–58.

———. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *The Quarterly Journal of Economics* 106:1039–61.

van Putten, W. L. J. 1982. "Maximum Likelihood Estimation for Luce's Choice Model." *Journal of Mathematical Psychology* 25:163–74.

Viscusi, W. K., W. A. Magat, and J. Huber. 1991. "Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Tradeoffs for Chronic Bronchitis." *Journal of Environmental Economics and Management* 21 (1):32–51.

Wilcox, N. T. 1993. "Lottery Choice: Incentives, Complexity and Decision Time." *The Economic Journal* 103:1397–1417.

Winsberg, S., and J. O. Ramsay. 1981. "Analysis of Pairwise Preference Data Using Integrated B-splines." *Psychometrika* 46:171–86.

Yellott, J. I. 1977. "The Relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgement, and the Double Exponential Distribution." *Journal of Mathematical Psychology* 15:109–44.