# Preference Uncertainty, Preference Learning, and Paired Comparison Experiments

***David C. Kingsley and Thomas C. Brown***

ABSTRACT. *Results from paired comparison experiments suggest that as respondents progress through a sequence of binary choices they become more consistent, apparently fine-tuning their preferences. Consistency may be indicated by the variance of the estimated valuation distribution measured by the error term in the random utility model. A significant reduction in the variance is shown to be consistent with a model of preference uncertainty allowing for preference learning. Respondents become more adept at discriminating among items as they gain experience considering and comparing them, suggesting that methods allowing for such experience may obtain more well founded values.* (JEL C25, D83)

## I. INTRODUCTION

A fundamental assumption of neoclassical microeconomic theory is that preferences are transitive. This intuitive assumption implies that among a series of binary choices preferences cannot cycle. For example, if a consumer prefers A to B and B to C then it follows that he or she also prefers A to C. Paired comparison experiments involve multiple binary choices between items in the choice set, allowing researchers to test the transitivity axiom.

Past research shows that for all but very small choice sets, respondents' paired choices are rarely fully transitive, but that as respondents progress through a random sequence of paired choices they become more consistent, apparently fine-tuning their preferences (Brown et al. 2008). This fine-tuning implies imprecision of preference, or in other words, preference uncertainty. Preference uncertainty was described by Thurstone (1927) as reflecting an underlying valuation distribution from which an individual randomly draws a value at a given instant. Allowing for preference uncertainty, the respondent becomes a potential source of error within choice models. Respondent error and the existence of preference uncertainty is an increasingly important topic being investigated within choice experiments and valuation studies. Indeed, an emergent theme within nonmarket valuation is to allow respondents to express levels of uncertainty (Alberini, Boyle, and Welsh 2003; Champ et al. 1997; Evans, Flores, and Boyle 2003; Li and Mattsson 1995; Welsh and Poe 1998).

Preference uncertainty implies that the design of the experiment or valuation survey may affect respondent choice. Researchers have examined the effect of experimental design using the error variance of the random utility model as a measure of preference uncertainty. Increasing the complexity of the choice set was found to increase the variance of the error term in a heteroskedastic logit model (DeShazo and Fermo 2002). Deshazo and Fermo hypothesize that the variance of choice models and choice consistency are inversely related. Similarly, it has been shown that the difficulty of the choice, referred to as task demand, has a nonlinear effect on the error term. Both very easy and very difficult choices were more random

*Land Economics* • August 2010 • 86 (3): 530–544
ISSN 0023-7639; E-ISSN 1543-8325


The authors are, respectively, assistant professor, Department of Economics and Management, Westfield State College, Westfield, Massachusetts; and research economist, Rocky Mountain Research Station, Fort Collins, Colorado. This paper benefited from discussions with Patricia Champ, Nick Flores, Jason Shogren, and Donald Waldman, as well as conference and seminar participants at the AERE sessions at the ASSA 2006 annual meetings, University of Colorado at Boulder, the Environmental Protection Agency, and Stephen F. Austin State University. All errors remain our own.

(Swait and Adamowicz 1996). These papers suggest preference uncertainty but do not address preference learning.

This step was taken in two studies that looked at the effect that repeated choices had on the mean and variance of elicited preferences (Holmes and Boyle 2005; Savage and Waldman 2008). A reduction in the error variance through the choice experiment implies preference learning, while an increase implies respondent fatigue or boredom. Results from Savage and Waldman (2008) were mixed; in their web sample fatigue was supported, but in their mail sample the error was constant. Holmes and Boyle (2005) found that error variance did decline over a sequence of choices, implying that respondents were better able to discriminate between choices made later in the experiment.

In this paper we show that the increasing choice consistency observed by Brown et al. (2008) is accompanied by a significant reduction in the error variance of a random utility model fit to the paired comparison data. We interpret this finding as preference learning. This result implies that the data become less noisy over choice occasions and indicates that respondents are better able to discriminate between items in later choices. Further, we find, as expected, that greater utility difference between items significantly reduces the probability of an inconsistent choice, and that inconsistent choices are likely to be switched when retested at the end of the experiment.

Taken together, these findings suggest that even hypothetical market experience provided through simple paired comparisons may affect respondents' choices and that nonmarket valuation techniques that rely on only one or a few responses may not be obtaining well-founded values. This finding is in line with the recent report by Bateman et al. (2008) that respondents to a dichotomous-choice contingent valuation survey require repetition and experience with the choice task in order to express preferences consistent with economic theory. As described in more detail in the Discussion section (Section V), our finding is also not inconsistent with the discovered preference hypothesis (Plott 1996), which maintains that stable underlying preferences are uncovered through experience with a choice task.

## II. PREFERENCE UNCERTAINTY AND LEARNING

Random utility models provide a general framework within which researchers investigate individual choice behavior (McFadden 2001). Consistent with economic theory, these models assume that individuals always choose the alternative yielding the highest level of utility (Marschak 1959). Utility is described as a random variable in order to reflect the researcher's observational deficiencies, not individuals' uncertainty about their own preferences (Ben-Akiva and Lerman 1985).

The model that Marschak (1959) proposed was an interpretation of what was probably the first stochastic model of choice, introduced by L. L. Thurstone in 1927 under the name of "the law of comparative judgment." Unlike the modern random utility model, in Thurstone's model, utility is represented by a distribution about a fixed point of central tendency (Thurstone 1927). This representation of utility has important implications concerning the source of error in choice models and represents the fundamental difference between these models (Brown and Peterson 2009; Flores 2003). Thurstone's model is now referred to as a constant utility model (Ben-Akiva and Lerman 1985). The constant utility model allows individuals to sample their utility from a distribution; choices are made based on the realization of utility on a particular choice occasion. This uncertainty may cause observed preferences to appear inconsistent (i.e., violate transitivity).

The law of comparative judgment was developed to explain common results from psychometric choice experiments involving binary choices (Bock and Jones 1968; Brown and Peterson 2003; Torgerson 1958). For Thurstone, a choice between two alternatives involved draws from two underlying preference or judgment distributions (McFadden 2001). Subjects might, for example, be presented with numerous pairs of objects and asked, for each pair, to say which object is heavier. The main finding, which dates back at least to Fechner's work (1860), was, not surprisingly, that the closer the items were in weight the more common incorrect selections became.

Allowing for researcher error is common practice in economic models. Although allowing for the existence of uncertain preferences and sources of error beyond the researcher is less common, it has not been ignored. For example, Bockstael and Strand (1987) examined the effect the source of error has on the estimation of economic values in a framework they called "random preferences." More recent research suggests that each respondent has an implicit valuation distribution (Wang 1997). For Wang, respondents answer dichotomous choice questions as if their values reflect distributions rather than fixed points. Similarly, Li and Mattsson (1995) assume that respondents have incomplete knowledge of their preferences and thus can give the wrong answer to a dichotomous choice question. They find that respondents are a significant source of error and so exacerbate the standard deviation of the estimated valuation distribution.

This paper assumes that both sources of error, researcher and respondent, are present in individual choice. The term "preference uncertainty" reflects respondent error, which translates to draws from an underlying valuation distribution unknown to both the respondent and the researcher. These random draws may contribute to inconsistency and increase the noise measured in the data. If respondent uncertainty could be reduced, perhaps through market experience or experimental design, choice consistency would increase and the data would become less noisy. This process will be referred to as preference learning and will be evident through a reduction in the standard deviation of the estimated valuation distribution measured by the error variance in the random utility model.

*Dichotomous Choice Contingent Valuation*

In a standard dichotomous choice contingent valuation study, respondents are asked to respond "yes" or "no" to a question such as, "Would you be willing to pay $t_i$ dollars to obtain environmental improvement $k$?" The individual's valuation function is defined as follows:

$$u_{ik} = \alpha_k + \varepsilon_{ik}, \qquad [1]$$

where $u_{ik}$ is individual $i$'s unobserved utility of item $k$, the deterministic component of value is represented by $\alpha_k$, and $\varepsilon_{ik}$ represents the stochastic component. Note that we assume a homogeneous set of individuals with respect to $\alpha_k$. It is common to express $\alpha_k$ as linear in parameters, $x_i'\beta$, where $x_i$ is a set of variables describing the characteristics of either the individual or the item. The respondent is assumed to choose "yes" whenever $u_{ik} \geq t_i$. Therefore,

$$P(yes) = P(u_{ik} \geq t_i) = P(\varepsilon_{ik} \geq t_i - \alpha_k), \qquad [2]$$

where $P$ indicates probability. Allowing the stochastic error term, $\varepsilon_{ik}$, to be normally distributed with mean zero and constant variance, $\sigma_\varepsilon^2$, we have the following expressions:

$$P(yes) = 1 - \Phi\left(\frac{t_i - \alpha_k}{\sigma_\varepsilon}\right) \qquad [3]$$

and

$$P(no) = 1 - P(yes), \qquad [4]$$

where $\Phi$ is the standard normal cumulative distribution. Then $\sigma_\varepsilon$ represents the standard deviation of the estimated valuation distribution, which has mean $\alpha_k$.

It is worth noting that within dichotomous choice contingent valuation settings, the assumption of a symmetric valuation distribution means that the scale of the model has little consequence, such that preference uncertainty leads to no bias in the estimated mean or median. The importance of recognizing preference uncertainty and preference learning becomes evident within choice experiments where respondents make several choices between items. Common examples of such choice experiments include attribute-based methods and paired comparison experiments.

*Paired Comparison Experiments*

Consider the choice between two items, labeled $r$ and $c$. The utilities of the items to individual $i$ are distributed as follows:

$$u_{ir} = \alpha_r + \varepsilon_{ir} \tag{5}$$

and

$$u_{ic} = \alpha_c + \varepsilon_{ic}. \tag{6}$$

Under the assumption that $\varepsilon_{ik}$ is a mean zero random variable distributed i.i.d. normal, the choice between items $r$ and $c$ can be written probabilistically, where $P_{rc}$ is the probability that item $r$ is chosen over item $c$:

$$\begin{aligned} P_{rc} &= P(u_{ir} > u_{ic}) = P(\alpha_r + \varepsilon_{ir} > \alpha_c + \varepsilon_{ic}) \\ &= P(\varepsilon_{ic} - \varepsilon_{ir} < \alpha_r - \alpha_c) \end{aligned} \tag{7}$$

or

$$P_{rc} = \Phi\left(\frac{\alpha_r - \alpha_c}{\sqrt{2}\sigma_\varepsilon}\right) \tag{8}$$

and

$$P_{cr} = 1 - P_{rc}, \tag{9}$$

where $\sqrt{2}\sigma_\varepsilon$ is the standard deviation of $\varepsilon_{ic} - \varepsilon_{ir}$.

Consider the density functions of items $r$ and $c$ depicted in Figures 1 and 2, which represent the underlying valuation distribution assumed to exist for each individual. In expectation item $r$ is preferred to item $c$, since $\alpha_r > \alpha_c$. But for a given choice, individuals act on their instantaneous values, not their expected values, which are unknown. Figure 1 shows the instantaneous value of item $r$, $u_{ir}$, above (to the right of) the instantaneous value of item $c$, $u_{ic}$. Thus, Figure 1 depicts a consistent choice because the choice based on these instantaneous values is consistent with the individual's underlying preferences represented by the expected values. As depicted in Figure 2, these two density functions allow for an inconsistent choice, wherein $u_{ic} > u_{ir}$ despite $r$ being preferred in expectation.

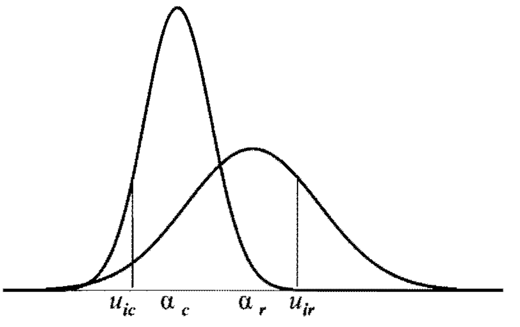


FIGURE 1
A CONSISTENT CHOICE

The expression for $P_{rc}$ provides two intuitive results. First, for a given standard deviation, $\sigma_\varepsilon$, the greater the utility difference, $\alpha_{rc} = \alpha_r - \alpha_c$, the more likely the choice will be consistent (item $r$ being chosen over item $c$), and conversely, the less likely an inconsistent choice becomes

$$\frac{d\Phi\left(\frac{\alpha_{rc}}{\sqrt{2}\sigma_\varepsilon}\right)}{d\alpha_{rc}} = \phi\left(\frac{\alpha_{rc}}{\sqrt{2}\sigma_\varepsilon}\right)\frac{1}{\sqrt{2}\sigma_\varepsilon} > 0. \tag{10}$$

Second, for a given utility difference, $\alpha_{rc}$, the narrower the distribution (the smaller is

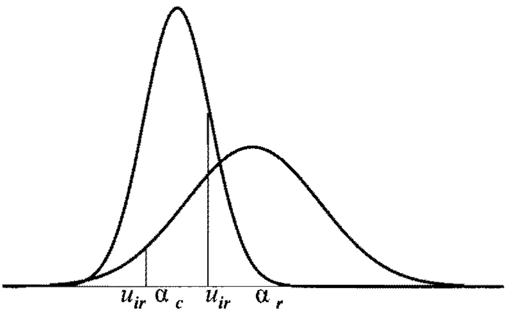


FIGURE 2
AN INCONSISTENT CHOICE

TABLE 1

ITEMS INCLUDED BY PETERSON AND BROWN (1998)

| | |
|---|---|
| 1 | A meal at a restaurant of the respondent's choice not to exceed $15 (Meal) |
| 2 | A nontransferable $200 gift certificate to a clothing store of the respondent's choice (Clothes) |
| 3 | Two tickets and transportation to a cultural or sporting event in Denver estimated at $75 (Tickets and transportation) |
| 4 | A nontransferable $500 certificate good for travel on any airline (Airline tickets) |
| 5 | A 2,000-acre wildlife refuge in the mountains west of Fort Collins, Colorado, purchased by the university (Wildlife refuge) |
| 6 | An agreement among Colorado State University, local business, and government to improve the water and air quality in Fort Collins (Clean air arrangement) |
| 7 | An annual no-cost on-campus weekend music festival open to all students (Spring festival) |
| 8 | A no-fee service providing video tapes of all class lectures in the university library (Video service) |
| 9 | An expansion to the parking garage system on campus so that parking is always easy to find and convenient (Parking capacity) |
| 10 | An expansion of the eating area in the student center (Eating area) |

$\sigma_\varepsilon$) the more likely a consistent choice becomes, and the wider the distribution the more likely an inconsistent choice becomes:

$$\frac{d\Phi\left(\frac{\alpha_{rc}}{\sqrt{2}\sigma_\varepsilon}\right)}{d\sigma_\varepsilon} = -\phi\left(\frac{\alpha_{rc}}{\sqrt{2}\sigma_\varepsilon}\right)\frac{\alpha_{rc}}{\sqrt{2}\sigma_\varepsilon^2} < 0. \qquad [11]$$

In psychometric experiments, inconsistent choices are easily identified because the expected value of each item is objective (e.g., the weight of an object). However, in economic valuation studies the expected value must be estimated, and inconsistency, particularly within an individual, is not easily identified. Peterson and Brown (1998) developed a simple technique (discussed in the next section) used with paired comparison experiments that identifies a respondent's likely set of inconsistent choices.

## III. PAIRED COMPARISON METHODOLOGY

The paired comparison method has successfully been used in nonmarket valuation studies (Champ and Loomis 1998; Kingsley 2006; Loomis et al. 1998; Peterson and Brown 1998).[1] In this paper we reanalyze paired comparison data collected by Peterson and Brown (1998). In the Peterson and Brown experiment all items were economic gains. Respondents were instructed to choose the item in each pair they would prefer if they could have either at no cost. The paired choices were drawn from a set of four private goods and six locally relevant public goods (Table 1) along with 11 monetary amounts.[2] Items were not paired with themselves, and dollar amounts were not compared (it was assumed that larger dollar amounts were preferred). Each respondent made 155 choices, 45 between items and 110 between an item and a dollar amount. For presentation, pairs were randomized across respondent and choice occasion. The pairs were presented on a personal computer, and the time respondents took to enter each choice was recorded. Three hundred and thirty students from Colorado State University participated in the study. Four were dropped because of missing data, leaving a total of 326 respondents, providing 50,530 individual observations. In addition, the experiment retested 10 consistent and all inconsistent choices within an individual after the initial choices were made. The respondents had not been informed that some choices would be repeated, and there was no break in the presentation of pairs to

[1] This paper is primarily concerned with what the measure of the error term of the model of individual choice reveals about the decision making process, and not with the estimated values of the items assessed. All data related to the estimated means are available upon request.

[2] Dollar amounts were 1, 25, 50, 75, 100, 200, 300, 400, 500, 600, and 700.

indicate that a new portion of the experiment had begun.

Given a set of $t$ items, the paired comparison method presents them independently in pairs as $(t/2)(t-1)$ discrete binary choices. These choices yield a preference score for each item, which is the number of times the respondent prefers that item to other items in the set. A respondent's vector of preference scores describes the individual's preference order among the items in the choice set, with larger integers indicating more-preferred items. In the case of a 21-item choice set, an individual preference score vector with no circular triads contains all 21 integers from 0 through 20. Circular triads (i.e., choices that imply A > B > C > A) cause some integers to appear more than once in the preference score vector, while others disappear.

For a given respondent, a pair's preference score difference (PSD) is simply the absolute value of the difference between the preference scores of the two items of the pair. This integer, which can range from 0 to 20 for a 21-item choice set, indicates on an ordinal scale the difference in value assigned to the two items.

The number of circular triads in each individual's set of binary choices can be calculated directly from the preference scores. 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 the maximum.[3] The coefficient varies from one, indicating no circular triads in a person's choices, to zero, indicating the maximum possible number of circular triads.

When a circular triad occurs, it is not unambiguous which choice is the cause of the circularity. This is easily seen by considering a choice set of three items whose three paired comparisons produce the following circular triad: A > B > C > A. Reversing any one of the three binary choices removes the circularity of preference; selection of one to label as "inconsistent" is arbitrary. However, with more items in the choice set, selection of inconsistent choices, though still imperfect, can be quite accurate. For each respondent, we selected as inconsistent any choice that was contrary to the order of the items in the respondent's preference score vector, with the condition that the order of items with identical preference scores was necessarily arbitrary. Simulations show that the accuracy of this procedure in correctly identifying inconsistent choices increases rapidly as the PSD increases. In simulations with a set of 21 items and assuming normal dispersion distributions, the accuracy of the procedure rises quickly from 50% at a PSD of 0 to nearly 100% at a PSD of 5.[4]

## IV. RESULTS AND ANALYSIS

In this section, we first report on the likelihood of an inconsistent choice and the likelihood of a preference reversal. This analysis provides support for the notion of preference learning. We then take a closer look at preference uncertainty and preference learning, fitting a heteroskedastic probit model to the paired comparison data.

### *Probability of an Inconsistent Choice*

The proportion of choices identified as inconsistent, using the decision rule described above, falls quickly and then levels off as respondents progress through the experiment (Figure 3). Nearly 14% of respondents' first choices are identified as inconsistent, but by the twentieth choice the

[3] The maximum possible number of circular triads, $m$, is $(t/24)(t^2-1)$ when $t$ is an odd number, and $(t/24)(t^2-4)$ when $t$ is even, where $t$ is the number of items in the set. Letting $a_i$ equal the preference score of the $ith$ item and $b$ equal the average preference score, in other words, $(t-1)/2$, the number of circular triads is (David 1988) $c=(t/24)\ (t^2-1)-{}^1\!/\!_2\sum(a_i-b)^2$. The coefficient of consistency (Kendall and Smith 1940) is then defined as $1-c/m$.

[4] A thorough explanation of the procedure for specifying inconsistent choices is found in Chapter 4 of Brown and Peterson's book (2009), available at www.treesearch.fs.fed.us/pubs/31863.

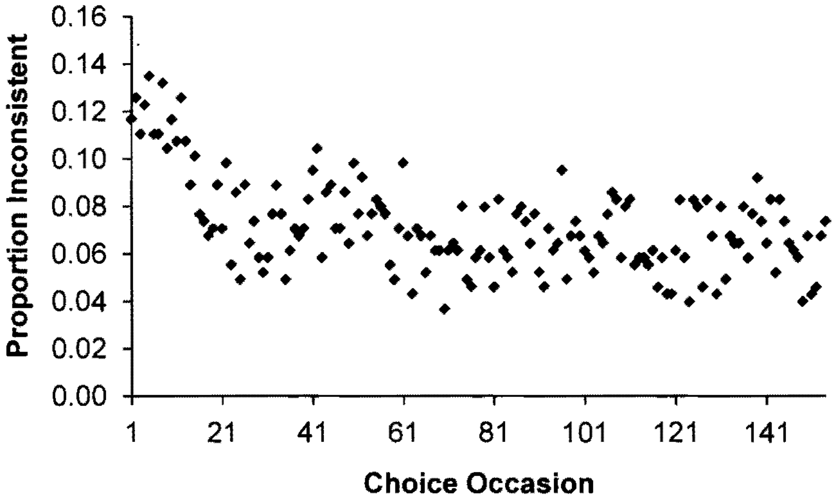


FIGURE 3
PROPORTION OF CHOICES IDENTIFIED AS INCONSISTENT

proportion has dropped to around 5% to 6%. Because the order in which the pairs were presented was random for each respondent, this result is unrelated to presentation order and thus seems driven by respondents' experience comparing the items.

A probit model is estimated to predict the probability of a choice being identified as inconsistent. The dependent variable, $y_{ij}$, equals 1 if the choice is identified as inconsistent and 0 otherwise, where $i$ denotes the individual and $j$ denotes the choice occasion. Exogenous variables include Choice occasion, ln(Choice occasion), PSD, a public good dummy, and the amount of time in seconds required to make the choice. Demographic variables included are Gender, Age, and Education.

The variable Choice occasion represents the order in which a given respondent encountered the pairs of items and made a choice; given Figure 3, the probability of an inconsistent choice is expected to decrease over choice occasion. ln(Choice occasion) captures the decreasing and leveling off relationship between choice inconsistency and choice occasion depicted in Figure 3. PSD is used as an approximate measure of the utility difference, the expectation being that respondents are less likely to commit an inconsistent choice the greater is this difference.[5] The public good dummy represents any choice involving a public good (the public good may be paired with another public good, a private good, or a dollar amount). Respondents are assumed to face greater uncertainty when the choice involves a public good, as opposed to choices involving only private goods or a private good and a monetary amount, because they lack experience making public good choices. Greater uncertainty is expected to lead to greater choice inconsistency. Similarly, the amount of time required to make the choice reflects the difficulty of the choice and is expected to be positively related to choice inconsistency. Therefore,

$$P(y_{ij}=1)=\Phi(x'_{ij}\beta) \qquad [12]$$

and

[5] PSD should clearly be related to the probability of an inconsistent choice. It is included in the model to verify this expectation. Its inclusion raises the question of endogeneity. With this in mind, we ran several other probit models that systematically excluded individual independent variables in order to observe whether results were sensitive to inclusion of PSD or any of the other independent variables. Results remained consistent and the coefficient on PSD varied very little.

TABLE 2

PROBABILITY OF AN INCONSISTENT CHOICE

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Choice occasion | −0.000077 (0.0000113)** | | −0.000077 (0.0000178)** | |
| Ln(Choice occasion) | | −0.00498 (0.000518)** | | −0.00498 (0.000796)** |
| PSD | −0.0111 (0.000265)** | −0.0111 (0.000266)** | −0.0111 (0.000575)** | −0.0111 (0.000579)** |
| Public | 0.00687 (0.000979)** | 0.00677 (0.000969)** | 0.00687 (0.00122)** | 0.00677 (0.0012)** |
| Time | 0.000505 (0.0000787)** | 0.000316 (0.0000815)** | 0.000505 (0.000131)** | 0.000316 (0.000118)** |
| Age | −0.000588 (0.000202)** | −0.000579 (0.000199)** | −0.000588 (0.000253)** | −0.000579 (0.000249)** |
| Gender | −0.000781 (0.000994) | −0.00075 (0.000983) | −0.000781 (0.00192) | −0.00075 (0.00189) |
| School | 0.000422 (0.000380) | 0.000423 (0.000376) | 0.000422 (0.0006) | 0.000423 (0.000593) |
| Log likelihood | −10,159 | −10,134 | −10,159 | −10,134 |
| AIC | 20,332 | 20,282 | 20,332 | 20,282 |
| $N$ | 50,530 | 50,530 | 50,530 | 50,530 |

*Note:* Dependent variable, $y_{ij}$, equals 1 if the choice was identified as inconsistent. Marginal effects reported. Standard errors in parentheses. Columns (1) and (2) contain models that treat each observation as independent. Columns (3) and (4) adjust the standard errors for clustering by individual. AIC, Akaike information criterion; PSD, preference score difference.
** Coefficient is significant at the $p < 0.05$ level.

$$P(y_{ij}=0)=1-\Phi(x'_{ij}\beta). \quad [13]$$

All 50,530 choices are pooled over individuals $i$ and choice occasions $j$, so the likelihood function becomes

$$L(y_{ij};\beta)=\prod_{i}^{n}\prod_{j}^{J}\Phi(x'_{ij}\beta)^{y_{ij}}[1-\Phi(x'_{ij}\beta)]^{1-y_{ij}} \quad [14]$$

Results, shown in Table 2, support the intuition of the model. Note that only marginal effects are reported and that Columns (1) and (2) present models that treat each observation as independent, while Columns (3) and (4) adjust the standard errors to account for the repeated choices made within individual. As expected, controlling for repeated responses per individual raises the standard errors; however, the marginal effects remain significant. A negative and significant PSD indicates, as predicted in the conceptual model, that inconsistency is less likely the greater the utility difference between the items. Additionally, choices including a public good are more likely to be inconsistent, perhaps suggesting that there is greater uncertainty and thus a wider valuation distribution for these goods. Choices that require more time to make are also more likely to be inconsistent, and older students tend to be more consistent.

Importantly, Choice occasion (Columns 1 and 3) is negative and significant, suggesting that an inconsistent choice becomes less likely as respondents progress through the experiment. When Choice occasion is modeled in log form (Columns 2 and 4), Choice occasion is also significant, implying that the probability of an inconsistent choice decreases quickly and levels off. The Akaike information criterion (AIC) statistic reported in Table 2 is a measure of fit; as is shown, the log form provides a substantially lower AIC and therefore provides a better fit to the data. Figure 4[6] illustrates the two functional

[6] Figure 4 was created using the estimated coefficients from Table 2 Columns (3) and (4). All variables except dummy variables are set to their averages: PSD = 6.93, Age = 19.8, School = 13.2, Time = 3.2, Gender = 0 (Female), and Public = 0.

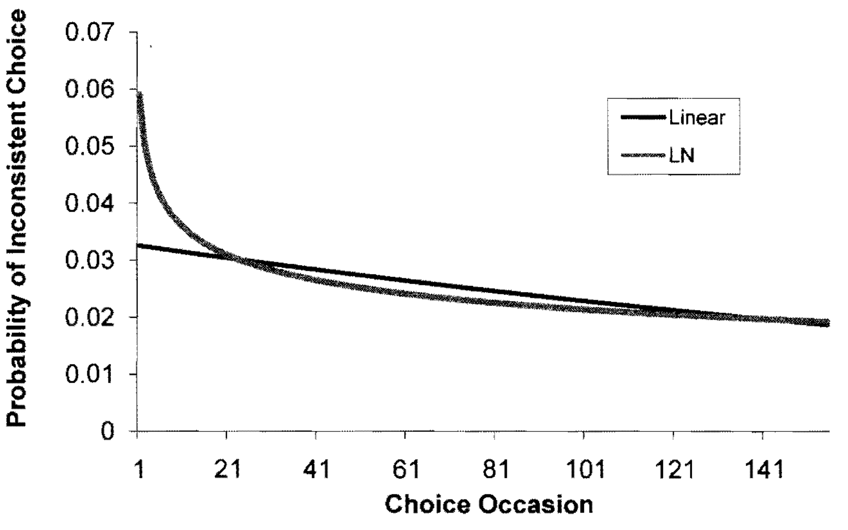


FIGURE 4
SPECIFICATION EFFECTS ON THE PROBABILITY OF AN INCONSISTENT CHOICE

forms based on the data in Columns (3) and (4).

*Probability of a Preference Reversal*

The experiment retested 10 randomly selected consistent choices and all inconsistent choices made by an individual after the initial 155 choices were made. The data show substantial differences in the reversal rates of originally consistent as opposed to originally inconsistent choices. Of the 3,260 consistent choices retested, 289, or 8.9%, were reversed, whereas 2,250 of 3,679 (61.2%) inconsistent choices were reversed (Table 3). This indicates that respondents were not only striving for consistency with their previous choices but also expressing their preferences.

The data are further broken out in Table 3 into the types of choices made. A choice may be between two public goods, two private goods, a public and private good, a public good and a dollar amount, or a private good and a dollar amount. As is shown, there are only slight differences among the types of choices in the rate of preference reversal, indicating that type of choice did not play a large role in preference reversals.

A probit model was used to test for the significance of factors hypothesized to affect the probability of a preference reversal. As noted above, two distinct sets of data were retested, the set of originally inconsistent choices and a random selection of 10 originally consistent choices. For both sets of choices the dependent variable, $y_{ij}$, equals 1 if the choice was reversed and 0 otherwise. Six exogenous variables are included: PSD,

TABLE 3
CHOICE SWITCHING

| Type of Choice | Number Retested | Number Switched | Proportion Switched |
|---|---|---|---|
| | Identified Inconsistent Choices | | |
| Public vs. public | 366 | 214 | 0.58 |
| Public vs. money | 1,493 | 970 | 0.65 |
| Private vs. private | 153 | 95 | 0.62 |
| Private vs. money | 921 | 518 | 0.56 |
| Public vs. private | 746 | 453 | 0.61 |
| Total | 3,679 | 2,250 | 0.61 |
| | Identified Consistent Choices | | |
| Public vs. public | 257 | 26 | 0.10 |
| Public vs. money | 1,307 | 121 | 0.09 |
| Private vs. private | 157 | 18 | 0.11 |
| Private vs. money | 1,056 | 82 | 0.08 |
| Public vs. private | 483 | 42 | 0.09 |
| Total | 3,260 | 289 | 0.09 |

TABLE 4
PROBABILITY OF A PREFERENCE REVERSAL

| | Inconsistent | | Consistent | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Choice occasion | −0.00194 (0.000179)** | −0.00194 (0.000202)** | −0.000078 (0.0000728) | −0.000078 (0.0000738) |
| PSD | 0.0858 (0.00473)** | 0.0858 (0.00521)** | −0.0162 (0.000961)** | −0.0162 (0.00116)** |
| Public | 0.0638 (0.0185)** | 0.0638 (0.0177)** | 0.0141 (0.00644)** | 0.0141 (0.00753)** |
| Age | 0.00207 (0.00357) | 0.00207 (0.00352) | −0.00181 (0.00151) | −0.00181 (0.00148) |
| Sex | 0.0211 (0.0169) | 0.0211 (0.0174) | −0.000396 (0.00662) | −0.000396 (0.0082) |
| School | −0.0000226 (0.00639) | −0.0000226 (0.00614) | 0.00539 (00.00272)** | 0.00539 (0.00275)** |
| Log likelihood | −2,178 | −2,178 | −806 | −806 |
| AIC | 4,368 | 4,368 | 1,624 | 1,624 |
| $N$ | 3,679 | 3,679 | 3,260 | 3,260 |

*Note:* Dependent variable, $y_{ij}$, equals 1 if the original choice was reversed. Marginal effects reported. Standard errors in parentheses. Columns (1) and (3) contain models that treat each observation as independent. Columns (2) and (4) adjust the standard errors for clustering by individual. AIC, Akaike information criterion; PSD, preference score difference.
** Coefficient is significant at the $p < 0.05$ level.

the public good dummy, the choice occasion when the original choice was made, and the three demographic variables.

The results are presented in Table 4. Columns (1) and (2) investigate the probability of reversing a choice that was originally identified as inconsistent, whereas Columns (3) and (4) investigate the probability of reversing a choice that was originally identified as consistent. Columns (2) and (4) adjust the error terms to account for repeated observations within individual. As is shown, accounting for repeated observations has little impact on the marginal effects. Results suggest that the greater the PSD between the items the more (less) likely an originally inconsistent (consistent) choice is reversed. Choices involving a public good are more likely to be reversed in both the inconsistent and consistent subsets. Choice occasion has a significant and negative effect on the probability of reversing an originally inconsistent choice, showing that more recent inconsistent choices are less likely to be reversed. However, Choice occasion has no significant effect on reversing an originally consistent choice.

*Preference Learning*

This paper defines preference learning as a significant reduction in the error variance of the random utility model, that is, a significant narrowing of the estimated valuation distribution.[7] The valuation function for this analysis is as follows:

$$u_{ijk} = \alpha_k + \varepsilon_{ijk}. \qquad [15]$$

Again the index $i$ denotes the individual and $j$ denotes the choice occasion. Note that the data are set up in rows and columns; as such, the item index will be $k = r, c$ for row or column. The row contains the 10 items and the column contains the 10 items along with the 11 monetary amounts. As before, the model assumes that all respondents are identical (that they have the same valuation on each $\alpha_k$). The error term, $\varepsilon_{ijk}$, includes an alternative specific error constant and is normally distributed $N(0,\sigma^2_{\varepsilon_{ijk}})$. The probability contribution to the likelihood function is denoted $P_{rc}$ ($P_{cr}$) for the probability that the row (column) item is chosen over the column (row) item.

First consider the choice between an item and a dollar amount. Each bid amount appears only in the column and is denoted by $t_{ijc}$. The probability that the item is

[7] As noted previously, the primary focus of this paper is the error process over choice occasion and the ability of respondents to discriminate between choices as they progress through the paired comparison experiment. Therefore we limit our presentation here to the error process.

preferred to the dollar amount is

$$P_{rc} = P(u_{ijr} > t_{ijc}) = P(\alpha_r + \varepsilon_{ijr} > t_{ijc}), \quad [16]$$

and thus

$$P_{rc} = 1 - \Phi[(t_{ijc} - \alpha_r)/\sigma_r]. \quad [17]$$

Next, consider the choice between two items

$$P_{rc} = P(u_{ijr} > u_{ijc}) = P(\alpha_r + \varepsilon_{ijr} > \alpha_c + \varepsilon_{ijc}), \quad [18]$$

and thus,

$$P_{rc} = \Phi\left[\frac{\alpha_r - \alpha_c}{(\sigma^2_{\varepsilon_r} + \sigma^2_{\varepsilon_c})^{1/2}}\right], \quad [19]$$

where $(\sigma^2_{\varepsilon_r} + \sigma^2_{\varepsilon_c})^{1/2}$ is the standard deviation of $\varepsilon_c - \varepsilon_r$. The dependent variable, $y_{ijk}$, equals 1 if the column item is chosen and 0 otherwise. The likelihood function is written as follows:

$$L(y_{ijk}; \alpha_k, \sigma_{\varepsilon_k}) = \prod_i^n \prod_j^J P_{rc}^{1-y_{ijk}} P_{rc}^{y_{ijk}}. \quad [20]$$

In order to test for preference learning, a heteroskedastic probit model is introduced that allows the standard deviation of the error term to adjust over Choice occasion as follows:

$$\varepsilon_{ijk} \sim N\left(0, \sigma^2_{\varepsilon_{kj}}\right). \quad [21]$$

Two functional forms are considered: linear $\sigma_{\varepsilon_{kj}} = \lambda_k + \beta(j)$ and nonlinear $\sigma_{\varepsilon_{kj}} = \lambda_k + \beta(1/j)$. Furthermore, this reduced form specification sets the standard deviation to be a function of the three available demographic variables. It has been shown that one can identify the error variance of choice models by exploiting the variation in monetary thresholds (Cameron 1988; Cameron and James 1987). In particular, using the variation across dollar amounts in this paired comparison experiment, we are able to identify the standard deviation within the model. Furthermore, using choice occasion, we are able to identify changes to this error structure over the course of the experiment.

TABLE 5
HETEROSKEDASTIC PROBIT MODEL

| | (1) | (2) |
|---|---|---|
| $\lambda$ – Video service | 356 (27.9)** | 325 (27.8)** |
| $\lambda$ – Parking capacity | 334 (28.0)** | 299 (27.5)** |
| $\lambda$ – Meal | 264 (27.8)** | 233 (27.6)** |
| $\lambda$ – Clothes | 165 (24.8)** | 133 (24.6)** |
| $\lambda$ – Wildlife refuge | 464 (31.8)** | 432 (31.6)** |
| $\lambda$ – Spring festival | 335 (27.8)** | 302 (27.6)** |
| $\lambda$ – Tickets and transportation | 199 (25.6)** | 168 (25.4)** |
| $\lambda$ – Eating area | 362 (29.6)** | 325 (29.4)** |
| $\lambda$ – Airline tickets | 231 (25.5)** | 199 (25.4)** |
| $\lambda$ – Clean arrangement | 427 (29.9)** | 397 (29.8)** |
| Gender | 13.83 (4.84)** | 13.3 (4.79)** |
| Age | 4.95 (1.38)** | 4.91 (1.36)** |
| Education | −5.91 (2.39)** | −5.88 (2.36)** |
| $\beta$ | −0.26 (0.05)** | 348.99 (60.66)** |
| Log likelihood | −26,029 | −26,002 |
| AIC | 52,106 | 52,052 |
| $N$ | 50,530 | 50,530 |

*Note:* Standard errors in parentheses. The model for Column (1) is $\sigma_{\varepsilon_{kj}} = \lambda_k + \beta(j)$. The model for Column (2) is $\sigma_{\varepsilon_{kj}} = \lambda_k + \beta(1/j)$. AIC, Akaike information criterion.
** Coefficient is significant at the $p < 0.05$ level.

This specification, using the standard deviation as a function of choice occasion, made it impossible to cluster the error terms by individuals, as the full model was unidentified. However, previous models show little impact of clustering and we do not believe this to be a serious problem.

Interpretation of these parameters is as follows. Researcher error as well as any error generated by the respondent unrelated to choice occasion will be picked up by $\lambda_k$. This term can be interpreted as an alternative specific error constant that is a measure of the error introduced into the model by each item. In both forms a significant $\beta$ represents a significant change in the scale over choice occasion (i.e., a change in respondent error). A reduction in the scale is interpreted as preference learning, and an increase is interpreted as fatigue or boredom. The hypothesis to be tested is $H_0 : \beta = 0$, implying that choice occasion has no effect on the scale of the model.

Results for the parameterized error structure are shown in Table 5. The important result is the significance of $\beta$ in both the linear (Column 1) and nonlinear (Column 2) forms. Both models suggest that the scale of the choice model decreases as respondents progress through a sequence of randomly ordered choices. This implies that the data become less noisy as the respondent continues through the experiment. In order to believe that this reduction stems from researcher error, it would need to be the case that some unobservable characteristics of the choices became less significant to the respondent as the experiment progressed.

As shown in Figure 3, the proportion of choices identified as inconsistent drops quickly and levels off, suggesting that the nonlinear form is the more appropriate representation of the preference learning described in these data. This is also supported by the lower AIC statistic reported in Table 5 for the nonlinear model. This evidence supports the inverse relationship between choice consistency and the scale of random utility choice models (DeShazo and Fermo 2002; Savage and Waldman 2008).

The above analysis implies that the error change parameter, $\beta$, is equivalent across items. In order to relax this assumption we ran a model allowing each item to have a $\beta$. Table 6 presents the $\beta$ coefficients; to simplify the presentation the 10 $\lambda$ rows are not shown. The $\beta$ terms for 6 of the 10 items remain significant; for the 4 others the $\beta$ is found to be insignificant. It is not entirely surprising that across a variety of items some may benefit from learning while others may not. Why some items are susceptible to learning while others are not is unknown and may be experiment specific. However, it is important to note that the results are not driven to be a single item or even a subset of items (i.e., only the public goods or only the private goods) and that the learning indicated in Table 5 is supported across a variety of items.

TABLE 6
HETEROSKEDASTIC PROBIT MODEL INDIVIDUAL BETA TERMS

| | (1) |
|---|---|
| $\beta$ – Video service | 3064 (653.5)** |
| $\beta$ – Parking capacity | −18.4 (97.60) |
| $\beta$ – Meal | 128.6 (130.50) |
| $\beta$ – Clothes | 285.1 (106.4)** |
| $\beta$ – Wildlife refuge | 320.5 (243.30) |
| $\beta$ – Spring festival | 793.9 (354.1)** |
| $\beta$ – Tickets and transportation | 730.2 (250.9)** |
| $\beta$ – Eating area | 296.2 (147.8)** |
| $\beta$ – Airline tickets | 342.9 (146.9)** |
| $\beta$ – Clean arrangement | 237.2 (231.70) |
| Gender | 12.59 (4.78)** |
| Age | 4.89 (1.36)** |
| Education | −5.75 (2.36)** |
| Log likelihood | −25,983 |
| AIC | 52,031 |
| $N$ | 50,530 |

*Note:* Based on the model $\sigma_{\varepsilon_{kj}} = \lambda_k + \beta_k(1/j)$. Standard errors in parentheses. AIC, Akaike information criterion.
** Coefficient is significant at the $p < 0.05$ level.

## V. CONCLUSION AND DISCUSSION

This paper combines common notions of preference uncertainty and preference learning found in the literature. Preference uncertainty implies that an individual's choice on a particular occasion represents a draw from an underlying valuation distribution. Therefore, the respondent is a potential source of error within choice models. The existence of preference uncertainty allows for the possibility that market experience or experimental design may affect respondent choice. A change in preference uncertainty, depending on the direction of the change, can be interpreted as preference learning or respondent fatigue.

Our investigation of changes in the error process over choice occasion points to some process of preference learning. It was found that the probability of an inconsistent choice decreases with choice occasion and utility difference between items of a pair. It was also found that the greater the utility difference between the items the more (less) likely an originally inconsistent (consistent) choice is reversed upon retrial. Both of these results are implied by the conceptual model presented.

Further, using a heteroskedastic probit model, it is shown that as respondents progress through the experiment, the scale of the random utility model decreases significantly. This result implies that the data become less noisy over choice occasion and indicates that the respondents are better able to discriminate between items in later choices. This result supports the ability of individuals to learn their preferences as they progress through a choice experiment in a way consistent with the discovered preference hypothesis (Plott 1996). This hypothesis proposes that stable underlying preferences are "discovered," or "learned," through experience with a choice task that provides relevant feedback and incentives (see also Braga and Starmer 2005). Respondents in our paired comparison sessions can be said to be learning their underlying preferences in the sense that they refined or fine-tuned them. Further, we observed no indication that the underlying preferences were not stable, since, as reported by Brown et al. (2008), there was no change in mean preference scores over choice occasion.[8] It is notable that this refinement occurred even though the paired comparison task provides no feedback or incentives that would encourage respondents to focus more carefully or otherwise alter their behavior. This suggests that mere experience in choosing between pairs of a set of items is sufficient for preference learning to occur.[9]

The result that respondents are able to learn their preferences as they progress through a choice experiment raises important questions for the economic valuation literature. Notably, one must consider the efficiency, if not the appropriateness, of single-shot surveys that assume well-founded preferences exist a priori. Furthermore, the question of whether elicited preferences may be altered through this process is of central concern to the literature and is deserving of further investigation.

[8] An alternative to the discovered preference hypothesis is the constructed preference hypothesis (Gregory, Lichtenstein, and Slovic 1993), which posits that preferences are quite labile and context dependent and that they are constructed during and possibly influenced by the choice task at hand. We presented the paired choices in only one context and, thus, did not test the extent to which respondents' choices were subject to contextual cues. We do note, however, that the paired comparison task is relatively free of contextual cues, in that the choices are binary and randomly ordered and the respondents do not interact. Further, in our implementation of the task, payment—and all the concerns that arise with payment—were not an issue. The method of paired comparisons thus offers a credible way to assess underlying preferences without being overly concerned about the influence of unwanted contextual cues.

[9] This finding is also not inconsistent with that of List (2003, 2004), although the kind of field market experience he examined is quite different from the experience gained during a paired comparison exercise.

## References


Alberini, Anna, Kevin J. Boyle, and Michael P. Welsh. 2003. "Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty." *Journal of Environmental Economics and Management* 45 (1): 40–62.

Bateman, Ian J., Diane Burgess, W. George Hutchinson, and David I. Matthews. 2008. "Learning Design Contingent Valuation: NOAA Guidelines, Preference Learning and Coherent Arbitrariness." *Journal of Environmental Economics and Management* 55 (2): 127–41.

Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: MIT Press.

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

Bockstael, Nancy E., and Iver E. Strand, Jr. 1987. "The Effect of Common Sources of Regression Error on Benefit Estimates." *Land Economics* 63 (1): 11–20.

Braga, Jacinto, and Chris Starmer. 2005. "Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis." *Environmental and Resource Economics* 32 (1): 55–89.

Brown, Thomas C., David Kingsley, George L. Peterson, Nick Flores, Andrea Clarke, and Andrej Birjulin. 2008. "Reliability of Individual Valuations of Public and Private Goods: Response Time, Preference Learning, and Choice Consistency." *Journal of Public Economics* 92 (7): 1595–1606.

Brown, Thomas C., and George L. Peterson. 2003. "Multiple Good Valuation." In *A Primer on Nonmarket Valuation*, ed. Patricia A. Champ, Kevin Boyle, and Thomas C. Brown. Norwell, MA: Kluwer Academic Publishers.

———. 2009. *An Enquiry into the Method of Paired Comparison: Reliability, Scaling, and Thur-*

*stone's Law of Comparative Judgment*. General Technical Report RMRS-GTR-216WWW. Fort Collins, CO: Rocky Mountain Research Station.

Cameron, Trudy Ann. 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *Journal of Environmental Economics and Management* 15 (3): 355–79.

Cameron, Trudy Ann, and Michelle D. James. 1987. "Efficient Estimation Methods for "Closed-Ended" Contingent Valuation Surveys." *Review of Economics and Statistics* 69 (2): 269–76.

Champ, Patricia A., Richard C. Bishop, Thomas C. Brown, and Daniel W. McCollum. 1997. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management* 33:151–62.

Champ, Patricia A., and John B. Loomis. 1998. "WTA Estimates Using the Method of Paired Comparison: Tests of Robustness." *Environmental and Resource Economics* 12 (3): 375–86.

David, Herbert A. 1988. *The Method of Paired Comparisons,* 2nd ed. New York: Oxford University Press.

DeShazo, J. R., and German Fermo. 2002. "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency." *Journal of Environmental Economics and Management* 44 (1): 123–43.

Evans, Mary F., Nicholas E. Flores, and Kevin J. Boyle. 2003. "Multiple-Bounded Uncertainty Choice Data as Probabilistic Intentions." *Land Economics* 79 (4): 549–60.

Fechner, G. T. 1860. *Elemente Der Psychophysik*. Leipzig: Breitkopf and Hartel.

Flores, Nicholas E. 2003. "A Complimentary Approach to the Law of Comparative Judgment and Random Utility Models." Working Paper. Boulder, CO: University of Colorado, Department of Economics.

Gregory, Robin, Sarah Lichtenstein, and Paul Slovic. 1993. "Valuing Environmental Resources: A Constructive Approach." *Journal of Risk and Uncertainty* 7 (2): 177–97.

Holmes, Thomas P., and Kevin J. Boyle. 2005. "Dynamic Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions." *Land Economics* 81 (1): 114–26.

Kendall, M. G., and B. Babington Smith. 1940. "On the Method of Paired Comparisons." *Biometrika* 31 (3–4): 324–45.

Kingsley, David A. 2006. "Multiple Good Valuation Using Paired Comparison Choice Experiments." Working Paper. Boulder, CO: University of Colorado, Department of Economics.

Li, Chuan-Zhong, and Leif Mattsson. 1995. "Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation." *Journal of Environmental Economics and Management* 28 (2): 256–69.

List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118 (1): 41–71.

———. 2004. "Substitutability, Experience, and the Value Disparity: Evidence from the Marketplace." *Journal of Environmental Economics and Management* 47 (3): 486–509.

Loomis, John, George Peterson, Patricia Champ, Thomas Brown, and Beatrice Lucero. 1998. "Paired Comparison Estimates of Willingness to Accept versus Contingent Valuation Estimates of Willingness to Pay." *Journal of Economic Behavior and Organization* 35 (4): 501–15.

Marschak, Jocob. 1959. "Binary-Choice Constraints and Random Utility Indicators." In *Mathematical Methods in the Social Sciences*, ed. Kenneth J. Arrow, Samuel Karlin, and Patrick Suppes. Palo Alto, CA: Stanford University Press.

McFadden, Daniel. 2001. "Economic Choices." *American Economic Review* 91 (3): 351–78.

Peterson, George L., and Thomas C. Brown. 1998. "Economic Valuation by the Method of Paired Comparison, with Emphasis on Evaluation of the Transitivity Axiom." *Land Economics* 74 (2): 240–61.

Plott, Charles R. 1996. "Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis." In *Rational Foundations of Economic Behavior*, ed. Kenneth J. Arrow, Enrico Colombatto, Mark Perleman, and Christian Schmidt. London: Macmillan and St. Martin's.

Savage, Scott J., and Donald M. Waldman. 2008. "Learning and Fatigue During Choice Experiments: A Comparison of Online and Mail Survey Modes." *Journal of Applied Econometrics* 23 (3): 351–71.

Swait, Joffre, and Wiktor Adamowicz. 1996. "The Effect of Choice Complexity on Random Utility Models: An Application to Combined Stated and Revealed Preference Models. Gainesville: Department of Marketing, University of Florida.

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

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

Wang, H. 1997. "Treatment of 'Don't-Know' Responses in Contingent Valuation Surveys: A Random Valuation Model." *Journal of Environmental Economics and Management* 32 (2): 219–32.

Welsh, Michael P., and Gregory L. Poe. 1998. "Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach." *Journal of Environmental Economics and Management* 36 (2): 170–85.