# DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 06-06


# Preference Uncertainty, Preference Refinement and Paired Comparison Choice Experiments


David C. Kingsley
University of Colorado, Boulder



October 2006

Center for Economic Analysis
Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

# Preference Uncertainty, Preference Refinement and Paired Comparison Choice Experiments

David C. Kingsley*
University of Colorado, Boulder

October 22, 2006

**Abstract**

Paired comparison choice experiments allow researchers to measure within individual choice consistency and to retest originally inconsistent choices. Results from paired comparison choice experiments suggest that as respondents progress through a random sequence of paired choices they become more consistent, apparently *fine-tuning* their preferences (Brown et al., 2006). This paper investigates the implications of these results within a model of preference uncertainty allowing for preference refinement. Preference uncertainty implies that choices on each choice occasion are driven by realizations from an underlying valuation distribution (Thurstone, 1927; Li & Mattsson, 1995). Preference refinement is defined as a significant reduction in the standard deviation, referred to as the scale, of the valuation distribution. Results support the asserted inverse relationship between choice consistency and the scale of the model (Deshazo & Fermo, 2002). Additionally, the probability of observing an inconsistent choice decreases the greater the utility difference between the involved pair and the narrower the scale becomes. Finally, preference reversals are investigated using originally inconsistent choices and a subset of consistent choices which were retested within the experiment.

*Department of Economics, University of Colorado, 256 UCB, Boulder, CO 80309. david.kingsley@colorado.edu
This paper benefitted from discussions with Tom Brown, Patty Champ, Nick Flores, Jason Shogren and Donald Waldman as well as conference participants at the AERE sessions at the ASSA 2006 annual meetings.

# 1 Introduction

A fundamental assumption of neoclassical microeconomic theory is that preferences exhibit transitivity. This intuitive assumption implies that between pairs of items preferences cannot cycle. For example, if a consumer prefers A to B and B to C then it follows that they also prefer A to C. Paired comparison choice experiments are unique because they involve simple binary choices between pairs of items, allowing researchers to test the transitivity axiom. Choices violating this axiom, referred to as inconsistent, can be quickly identified and retested within individual.

Paired comparison research suggests that as respondents progress through a random sequence of paired choices they become more consistent, apparently *fine-tuning* their preferences (Brown et al., 2006). Brown et al. (2006) also find that the proportion inconsistent for choices involving public goods is higher and remains higher throughout the experiment than the same proportion for choices involving private goods. Furthermore, respondents are more likely to reverse an originally inconsistent choice than an originally consistent choice when retested. Peterson and Brown (1998) conclude that although original choices violate the transitivity axiom when given the opportunity the majority of these violations are reversed (Peterson & Brown, 1998). This paper investigates the implications of these results within a model of preference uncertainty allowing for preference refinement.

Random utility models provide a general framework within which researchers investigate individual choice behavior (McFadden, 2001). First developed by Marschak, random utility models assume that the individual always chooses the alternative yielding the highes level of utility (Marschak, 1960). Despite the assumption of well-behaved preferences, utility is described as a random variable in order to reflect the researcher's observational deficiencies (Ben-Akiva & Lerman, 1985). The model that Marschak proposed was an interpretation of an important paper by L.L. Thurstone (1927). Thurstone's Law of Comparative Judgement marks the first known model of choice. Within Thurstone's model the utility of the alternative would be assumed fixed; however, individuals would imperfectly sample this utility from an underlying distribution (Thurstone, 1927). Therefore, in this model utility would be described as a random variable to reflect preference uncertainty on the part of the respondent. A similar model has been proposed by Li and Mattsson (1995). Thurstone's model would now be referred to as a constant utility model (Ben-Akiva & Lerman, 1985). Although statistically equivalent the random utility and constant utility models are fundamentally different. The important distinction is the source of error within these models. The constant utility, opposed to the random 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 irrational (i.e. violate transitivity).

Preference uncertainty implies that an individual's choice on a particular choice occasion represents a realization from some underlying valuation distribution (Thurstone, 1927; Li & Mattsson, 1995) or perceived utility (McFadden, 2001). Therefore, 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. Almost 20 years ago Bockstael and Strand considered the effect respondent error, stemming from the inherent randomness of preferences, may have when estimating economic benefits (Bockstael & Strand, 1987). Furthermore, researchers have begun to allow respondents to express levels of uncertainty in choice experiments (Li & Mattsson, 1995; Champ, Bishop, Brown, & McCollum, 1997; Evans, Flores, & Boyle, 2003; Welsh & Poe, 1998).

The existence of preference uncertainty generates many important questions. One being whether or not the level of uncertainty is affected by the experimental design or more generally the decision environment? Researchers commonly use the scale, defined as the standard deviation, of the estimated valuation distribution as a measure of preference uncertainty. Increases in scale represent respondent fatigue, confusion or boredom (Deshazo & Fermo, 2002; Swait & Adamowicz, 1996) while decreases in scale are associated with preference learning or refinement[1] (Savage & Waldman, 2004). This paper explicitly defines preference refinement as a significant reduction in the scale of the valuation function around a stable mean. That is, a reduction in preference uncertainty. The mean of the valuation distribution represents the true underlying preferences, only the variance of the distribution is assumed dynamic[2]. True underlying preferences are assumed to be rational and thus to exhibit transitivity. The recognition that the experimental environment has an impact on respondent error is an important development in the literature.

The literature asserts an inverse relationship between individual choice consistency and the scale of a random utility model (Deshazo & Fermo, 2002). This paper supports this assertion by fitting a heteroscedastic probit model to the Peterson and Brown (1998)[3] paired comparison data. As respondents gain experience expressing their preferences they become more consistent and the scale is shown to decrease significantly. The significant narrowing of the scale is interpreted as preference refinement. Furthermore, the model presented suggests and the data support two intuitive factors determining the probability of observing an inconsistent choice. Specifically, the probability of observing an inconsistent choice decreases the greater the utility difference between the involved pair and as the scale of the model narrows.

[1]The term preference refinement is used to differentiate this process from preference learning discussed in related literature (Crocker & Shogren, 1991; Shogren, List, & Hayes, 2000). Crocker and Shogren (1991) develop a model of preference learning in which new or novel items will be valued more highly so that respondents can learn how the item fits into their existing preference set. The respondents are willing to pay a premium to experience new products because they receive the consumption value as well as information about their preferences.

[2]This concept, that respondents need to uncover their stable underlying preferences, has been previously expressed in the discovered preference hypothesis (Plott, 1996).

[3]I gratefully acknowledge and thank Tom Brown for providing me with this data

The paired comparison choice experiments investigated here provide further tests of preference refinement. All inconsistent and a subset of consistent choices were retested within the experiment. If experience expressing their preferences help respondents to refine their preferences than it is hypothesized that a heteroscedastic probit model fit to the retested data will reveal neither preference refinement nor respondent fatigue. Data support this hypothesis. Additionally, the probability of observing a preference reversal is investigated. Greater utility difference between the pair is expected to reduce the probability of reversing a consistent choice and increase the probability of reversing an inconsistent choice. Data support this hypothesis in the consistent subset but does not in the inconsistent subset.

## 2 Preference Uncertainty

In 1927 Thurstone developed the Law of Comparative Judgement in order to explain common results from psychometric choice experiments. Experiments of this type generally involve confronting an individual with a series of stimuli. For example, asking respondents to judge between a pair of items which is heavier. Thurstone's main finding was, not surprisingly, that the closer the items were in weight the more likely the occurrence of a judgement error. The attribute of interest, in this case weight, was measured along a *psychological continuum*. Thurstone claims that there is a decision process by which the choice between the two alternatives is made. No particular source of this decision process is defined, it is simply referred to as a *discriminal process*. That is, the process by which respondents discriminate between the alternatives on the given criteria (in this case weight). The discriminal process is essentially a draw or a sampling from an underlying distribution. For Thurstone the true weight of the stimuli would be represented by $V_i$, but this value is perceived or realized with a normal error $V_i + \epsilon_i$ (McFadden, 2001). The standard deviation of the discriminal process $\sigma$ is referred to as the discriminal dispersion.

Marschak (1960) was the first to interpret the psychosocial continuum, or the attribute of interest as utility and developed the first random utility model (Marschak, 1960). When perceived stimuli, $V_i + \epsilon_i$, is interpreted as utility this model is readily interpreted as an economic model of choice (McFadden, 2001). In both models the stimuli (weight or utility) is treated as a random variable. However, in the transition from a psychological model of choice to a economic model of choice a crucial but often overlooked assumption was made. This assumption concerns the source of the error in the respective models.

In the random utility model respondents are assumed to always choose the item which provides the highest level of satisfaction or utility. The analyst however does not observe all important characteristics and thus models the process as a random variable. Therefore, the researchers observation deficiency is the source of error (Manski, 1973). To the contrary, the Law of Comparative Judgement is a constant utility model (Ben-Akiva & Lerman, 1985) which allows the utility of an item to be fixed, however the utility would be perceived or sampled from the distribution. Thus the respondent is the source of error. Therefore, in a constant utility model judgement errors or inconsistent choices are expected. In fact, between a pair of items A and B the same respondent may choose A on the first choice occasion and on the next choose B without violating this choice model.

This paper will allow both sources of error to exist. The term preference uncertainty reflects respondent error. Utility for each item is considered fixed. The expected value of an individual's valuation distribution represents true underlying preferences. These true preferences are assumed to be well-behaved and thus rational. That is, if respondents can be made certain, perhaps through market experience, they would act according to classical economic theory. It is thus the random sampling or realizations of utility which seem to violate our predictions. Specifically, in this paper underlying preference are assumed transitive while realizations or perceived preferences may exhibit intransitivity. Respondent error creates judgement error which leads to what appear to be inconsistent choices.

Allowing for the existence of uncertain preferences and sources of error beyond researcher error has been considered at least since Bockstael and Strand (1987). Bockstael and Strand look at the effect the source of error has on the estimation of economic values in a framework they called *Random Preferences* (Bockstael & Strand, 1987). More recently Wang (1997) suggested, in order to explain why respondents choose the *Don't Know* option in dichotomous choice contingent valuation studies, that each respondent has an implicit valuation distribution. Respondents answer DCCV questions as if their values follow a distribution (Wang, 1997).

This paper also allows preferences to follow a distribution, however, opposed to Wang (1997) valuations or choices on each choice occasion are realizations from this distribution. As such this paper follows the work of Li and Mattsson closely (Li & Mattsson, 1995). Li and Mattsson assume that respondents have incomplete knowledge of their preferences and thus can give the wrong answer to a dichotomous choice contingent valuation question. First, the implications of respondent uncertainty in a dichotomous choice contingent valuation is discussed, following closely the work of Li and Mattsson (1995).

## 2.1 Dichotomous Choice Contingent Valuation

In a standard DCCV study respondents are asked to vote *yes/no* to a referendum question such as: Would you be willing to pay $t_i$ dollars to obtain some environmental improvement or resource $k$? The individuals valuation function will be defined as a follows.

$$u_{ik} = \alpha_k + \epsilon_{ik} \tag{1}$$

Where $u_{ik}$ is an individuals unobserved utility of item $k$, the deterministic component of value is represented by $\alpha_k$ and $\epsilon_{ik}$ represent the stochastic component. 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 $k$. The respondent will vote *yes* whenever $u_{ik} \geq t_i$. Therefore,

$$Pr(yes) = Pr(u_{ik} \geq t_i) = Pr(\epsilon_{ik} \geq t_i - \alpha_k) \tag{2}$$

and

$$Pr(no) = 1 - Pr(yes) \tag{3}$$

Let the stochastic error term, $\epsilon_{ik}$, be normally distributed with mean zero and constant variance $\sigma_\epsilon^2$. Then $\sigma_\epsilon$ represents the standard deviation, referred to as the scale, of the estimated valuation distribution which has mean $\alpha_k$.

Li and Mattsson introduce preference uncertainty by allowing individual valuations to be realizations from an underlying valuation distribution so that

$$\tilde{u}_{ik} = u_{ik} + \nu_{ik} \tag{4}$$

combining terms we have the unobserved utility function $\tilde{u}_{ik} = \alpha_k + \epsilon_{ik} + \nu_{ik}$. Therefore, respondents reply *yes* whenever $\tilde{u}_{ik} \geq t_i$.

$$Pr(yes) = Pr(\tilde{u}_{ik} \geq t_i) = Pr(e_{ik} \geq t_i - \alpha_k) \tag{5}$$

and

$$Pr(no) = 1 - Pr(yes) \tag{6}$$

Where $e_{ik}$ is a composite error term such that $e_{ik} = \epsilon_{ik} + \nu_{ik}$. Here $\epsilon_{ik}$ represents the stochastic component associated with the researcher arising from omitted variables. On the other hand, $\nu_{ik}$ represents the stochastic component associated with the respondent arising from preference uncertainty. As discussed in the following section the standard deviation of $\nu_{ik}$, $\sigma_\nu$, measures preference uncertainty reductions in which represent preference refinement.

In the presence of preference uncertainty, $\sigma_\nu \neq 0$, the estimated valuation distribution is flattened compared to a model with preference certainty $\sigma_\nu = 0$. This can be seen by comparing the estimated distribution with preference uncertainty $N(\alpha_k, \sigma_e^2)$ and without $N(\alpha_k, \sigma_\epsilon^2)$ where $\sigma_e = (\sigma_\epsilon^2 + \sigma_\nu^2)^{1/2} > \sigma_\epsilon$[4] whenever $\sigma_\nu \neq 0$. Therefore, in the absence of preference uncertainty this model collapses to standard model.

It is worth noting that within DCCV settings assuming a symmetric valuation distribution the scale of the model has little consequence. The presence of preference uncertainty leads to no bias in the estimated mean or median. The model will however tend to overestimate the standard deviation of the valuation distribution. The importance of recognizing preference uncertainty and preference refinement becomes evident within choice experiments where respondents make choices between items. Common examples of such choice experiments include attribute based methods (ABMs) and paired comparison experiments.

## 2.2 Choices Experiments

The Law of Comparative Judgement was designed to compare a series of choices between pairs of items (Torgerson, 1958; Bock & Jones, 1968; David, 1969). Not surprisingly, research found that the closer the two items were in weight the more likely an inconsistent choice was reported[5]. Importantly all error in these experiments stem from the respondent as there is no source of researcher error. This paper will maintain the notation established in section 2.1 recognizing that for Thurstone the error term consisted of only $\nu_{ik}$.

Here the choice will be between item $r$ and item $c$. They are distributed as follows:

$$\tilde{u}_{ir} = \alpha_r + e_{ir} \tag{7}$$

[4]This assumes independence between researcher and respondent error.

[5]An inconsistent choice implies that the heavier item was not chosen and for Thurstone represented a judgment error.

and

$$\tilde{u}_{ic} = \alpha_c + e_{ic} \tag{8}$$

Under the assumption that $e_{ik}$ is a mean zero random variable distributed i.i.d. normal, the choice between item $r$ and $c$ can be written probabilistically.

$$P_{rc} = P(\tilde{u}_{ir} > \tilde{u}_{ic}) = P(\alpha_r + e_{ir} > \alpha_c + e_{ic}) = Pr(e_{ic} - e_{ir} < \alpha_r - \alpha_c) \tag{9}$$

or

$$P_{rc} = \Phi\left(\frac{(\alpha_r - \alpha_c)}{\sqrt{2}\sigma_e}\right) \tag{10}$$

and

$$P_{cr} = 1 - P_{rc} = 1 - \Phi\left(\frac{(\alpha_r - \alpha_c)}{\sqrt{2}\sigma_e}\right) \tag{11}$$

Where $\Phi$ is the standard normal cumulative distribution and $\sqrt{2}\sigma_e$ is the standard deviation of $e_{ic} - e_{ir}$.

Consider the density functions of items $r$ and $c$ depicted in Figure 1. In expectation item $r$ is preferred to item $c$ as $E(u_{ir}) > E(u_{ic})$ or equivalently $\alpha_r > \alpha_c$. If this respondent were certain, or made certain, of his valuation he would always chose item $r$. The top panel represents a consistent choice, the realization of item $r$, $u_{ir}$, is greater (to the right of) $u_{ic}$ the draw on item $c$, therefore item $r$ is chosen over item $c$. However, the bottom panel of Figure 1 depicts an inconsistent choice. The realization of item $c$, $u_{ic}$, is greater than $u_{ir}$ so on this choice occasion the individual would choose item $c$.

In psychometric experiments inconsistent choices are easily identified because the expected value of each item is objective. However, in economic valuation studies the expected value must be estimated and inconsistency particularly within individual is not easily identified. Peterson and Brown (1998) develop a method (discussed in Section 4) which identifies a likely set of inconsistent choices within individual.

The expression for $P_{rc}$ provides two intuitive results. This expression measures the probability of a consistent choice, item $r$ being chosen over

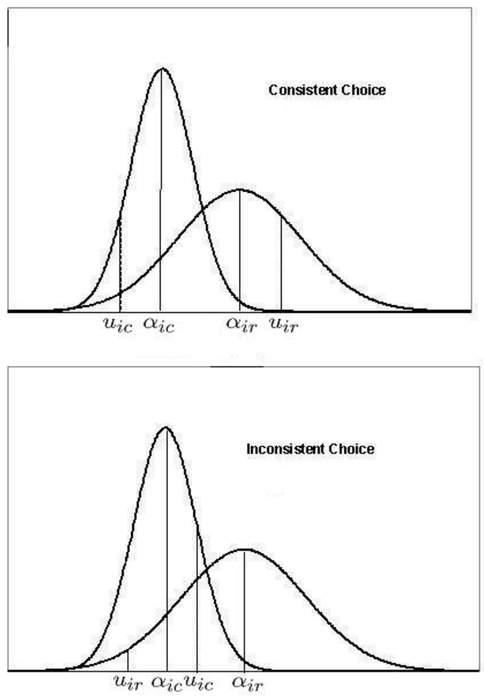


Figure 1: Consistent and Inconsistent Choices

item $c$. First for a given standard deviation, $\sigma_e$, the further apart are the means, $\alpha_{rc} = \alpha_r - \alpha_c$, the greater the utility difference and the more likely the choice is to be consistent (equation 12). Conversely the less likely an inconsistent choice becomes.

$$\frac{d\Phi\left(\frac{\alpha_r - \alpha_c}{\sqrt{2}\sigma_e}\right)}{d\alpha_{rc}} = \phi\left(\frac{\alpha_{rc}}{\sqrt{2}\sigma_e}\right)\frac{1}{\sqrt{2}\sigma_e} > 0 \tag{12}$$

Second, for a given utility difference, $\alpha_{rc}$, the wider the distribution, $\sigma_e$, the less likely a consistent choice (equation 13) becomes. On the other hand, the more narrow the distribution the more likely a consistent choice.

$$\frac{d\Phi\left(\frac{\alpha_{rc}}{\sqrt{2}\sigma_e}\right)}{d\sigma_e} = -\phi\left(\frac{\alpha_{rc}}{\sqrt{2}\sigma_e}\right)\frac{\alpha_{rc}}{\sqrt{2}\sigma_e^2} < 0 \tag{13}$$

As the scale, $\sigma_e$, approaches zero the choice becomes deterministic and increasingly consistent. In the ongoing example item $r$ becomes increasingly likely to be chosen. While as the scale approaches $\infty$ the choice becomes increasingly random. Therefore, decreases in $\sigma_e$ imply preference refinement as respondents become more able to discriminate between the alternatives while increases represent respondent fatigue or confusion (Holmes and Boyle 2006). This is the intuition behind preference refinement to which the paper now turns.

# 3 Preference Refinement

Allowing for preference uncertainty and respondent error one must ask whether the level of uncertainty can be affected by the experimental design. That is, does the task itself (the choice experiment) affect the level of respondent uncertainty. The level of uncertainty, as noted by Thurstone and the above analysis, can be measured by the standard deviation of the distribution.

The relationship between choice consistency and the scale of a random utility model has been investigated since researchers have considered scale as a sign of preference uncertainty (Deshazo & Fermo, 2002; Swait & Adamowicz, 1996). Deshazo and Fermo (2002) consider the impact choice set complexity has on the variance of error term in a heteroscedastic logit model. The variance of the random error term serves as a proxy for choice consistency. The more random are the choices across respondents the less consistent are the choices, thus increasing the variance of the model. In contrast to economists predictions they find that increasing the complexity of their choice set increases the variance of the error term. Deshazo and Fermo are able to affect the complexity by changing the number of alternatives as well as the number and correlation of the attributes in a conjoint framework. Swait and Adamowicz (1996) consider the difficulty of the choice, referred to as task demand, and find a non-linear affect on the error term. Both very easy and very difficult choices are more random. These papers suggest preference uncertainty but do not address preference refinement.

This step was taken in a study that looked at the effect that repeated choices had on the error term in a conjoint framework (Savage & Waldman, 2004). Savage and Waldman consider the standard deviation of the error term to be a measure of the randomness of the unexplained component of utility within a random utility model. They suggest that a reduction in the variance through the choice experiment implies learning (referred to here as refinement) while an increase implies respondent fatigue or boredom. In their web sample fatigue was supported as the error was significantly increased while in the mail sample the error was constant.

In the paired comparison choice experiments discussed in this paper respondents gain experience expressing their preferences among a variety of dissimilar goods. Preference refinements may be a result of these choices and is expected to decrease respondent error and thus $\sigma_\nu$.

# 4 Paired Comparison Choice Experiments

Peterson and Brown developed the method of paired comparisons for economic valuation and used it to investigate the transitivity axiom (Peterson & Brown, 1998). The method has successfully been used to estimate an equivalent surplus for an economic gain (Champ & Loomis, 1998; Loomis, Peterson, Champ, Brown, & Lucero, 1998; Kingsley, 2006a). In this 1998 paired comparison experiment all items were briefly described prior to the beginning of the experiment. Each respondent then sat at a computer and made hypothetical choices between pairs of items. Pairs were randomized

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

Table 1: Items included in Peterson and Brown 1998

across respondent and choice occasion. All respondents see all possible pairs. In a choice set of $t$ items each respondent makes $\frac{t(t-1)}{2}$ choices.

In the Peterson and Brown 1998 experiment all items were economic gains (see Table 1). Respondents were instructed to choose the item they would prefer if they could have either at no cost. The choice sets were drawn from a set of four private goods and six locally relevant public goods along with 11 monetary amounts[6]. Each respondent made 155 choices; 45 between items and 110 between an item and a dollar amount. Items are not paired with themselves and dollar amounts are not compared. Three hundred and thirty students from Colorado State University participated in the study. Three were dropped because of missing data leaving a total of 327 respondents, providing 50,685 individual observations.

[6]Monetary amounts included 1, 25, 50, 75, 100, 200, 300, 400, 500, 600 and 700.

| Raw Choice Matrix | | | | | |
|---|---|---|---|---|---|
| | | A | B | C | D |
| | A | - | 0 | 0 | 0 |
| | B | 1 | - | 1 | 1 |
| | C | 1 | 0 | - | 1 |
| | D | 1 | 0 | 0 | - |
| Preference Score | | 3 | 0 | 1 | 2 |

Table 2: Raw Choice Matrix

## 4.1 Identification of Inconsistent Choices

After all choices have been made preferences are described using choice matrices. One of these choice matrices, referred to as the raw choice matrix, summarizes the choices and preferences of a single individual. For example, consider a choice set containing only four items A, B, C and D. A choice is made between each pair, in this example each respondent makes 6 choices. A 1 in the matrix implies that the column item was chosen over the row item. Consider Table 2, it can be seen that this respondent's preferences are as follows: A $\succ$ B, A $\succ$ C and A $\succ$ D; D $\succ$ B D $\succ$ C and C $\succ$ B. Summation of the columns provides each items *preference score*. The preference score is simply the number of times the item was chosen over another item in the choice set and provides an ordinal measure of the item's value. An important measure is the *preference score difference* (PSD) which provides an approximate measure of the difference in value placed on the items. For example, the PSD between items A and B is 3 while between A and D it is 1, implying that this respondent prefers A over B more than they prefer A over D.

Next the double sorted choice matrix is formed. This matrix is formed by ranking the items by increasing preference score from left to right, similarly the rows are ranked by increasing row sum bottom to top. Table 3 displays two double sorted matrices. The top half is obtained from the raw choice ma-

| Consistent Choice Matrix | | | | | |
|---|---|---|---|---|---|
| | | B | C | D | A |
| | B | - | 1 | 1 | 1 |
| | C | 0 | - | 1 | 1 |
| | D | 0 | 0 | - | 1 |
| | A | 0 | 0 | 0 | - |
| Preference Score | | 0 | 1 | 2 | 3 |
| **Inconsistent Choice Matrix** | | | | | |
| | | B | C | D | A |
| | B | - | 1 | 1 | 0 |
| | C | 0 | - | 1 | 1 |
| | D | 0 | 0 | - | 1 |
| | A | 1 | 0 | 0 | - |
| Preference Score | | 1 | 1 | 2 | 2 |

Table 3: Double Sorted Choice Matrices

trix in Table 2. Note that the 1's all appear above the diagonal and that each integer between 0 and $t-1$ appears as a preference score. This choice matrix represents perfect choice consistency. The bottom half of Table 3 shows the double sorted matrix if the respondent had chosen B $\succ$ A rather than A $\succ$ B. This inconsistent choice causes a violation of the transitivity axiom: A $\succ$ C, C $\succ$ B but B $\succ$ A. Notice that the preference scores are no longer unique integers with both 1 and 2 being repeated. The important change is the 1 that appears below the principal diagonal. The double sorted choice matrix identifies inconsistent choices by isolating 1's below the diagonal[7]. In this example the choice B $\succ$ A would be identified as inconsistent.

[7] The effectiveness of the double sort algorithm has been extensively examined using simulations (Peterson & Brown, 2006). The details of these simulations are beyond the scope of the current paper. However, it is shown that given a choice set including 21 items (e.g. Peterson and Brown (1998)) and assuming normal error terms approximately 72% of the choices identified as inconsistent are indeed inconsistent. The Peterson and Brown (2006) paper is available at the following web address: http://www.fs.fed.us/rm/value/discpaper.html.

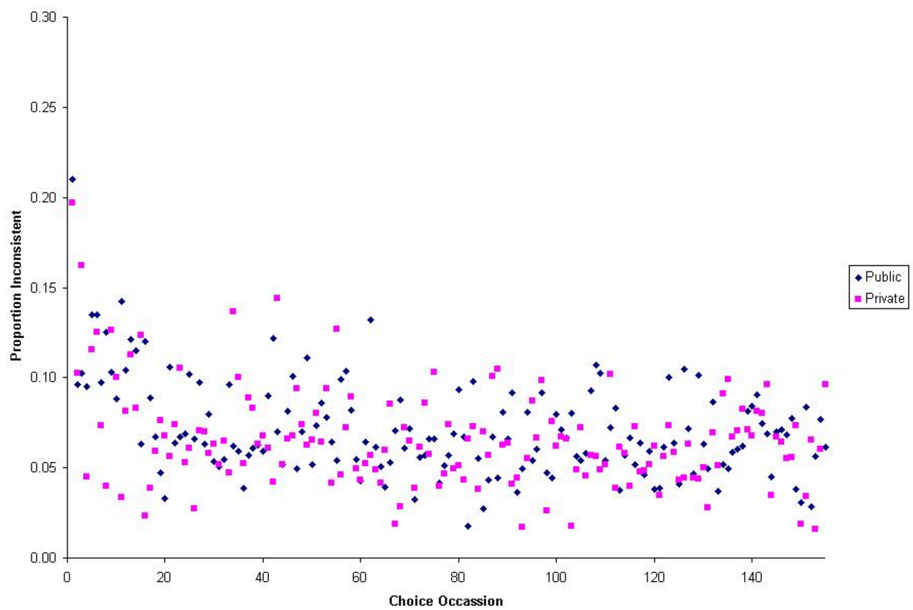


Figure 2: Proportion of Choices Identified as Inconsistent

## 4.2 Descriptive Results

The Peterson and Brown 1998 experiment, yields several important results. First, the proportion of choices identified as inconsistent drops as the respondents progress through the experiment. Recall that 155 choices were made by each respondent. In Figure 2 *Public* (*Private*) refers to choices between either two public (private) items or a public (private) item and a dollar amount. For both sets of choices a downward trend exists. This process, by which respondents become more consistent has been referred to as *fine-tuning* (Brown et al., 2006).

Furthermore the experimental design included retesting 10 consistent choices and all inconsistent choices made by an individual after the initial 155 choices were made. The respondents did not know that this portion of the experiment had started. Of the 3270 consistent choices retested 290 or

| Type of Choice | Inconsistent | Switched | Proportion |
|---|---|---|---|
| Public v. Public | 368 | 216 | .59 |
| Public v. Money | 1498 | 974 | .65 |
| Private v. Private | 153 | 95 | .62 |
| Private v. Money | 922 | 518 | .56 |
| Public v. Private | 747 | 453 | .61 |
| Total | 3688 | 2256 | .61 |
| | Consistent | Switched | Proportion |
| Public v. Public | 257 | 26 | .1 |
| Public v. Money | 1313 | 121 | .09 |
| Private v. Private | 157 | 18 | .11 |
| Private v. Money | 1060 | 83 | .08 |
| Public v. Private | 483 | 42 | .09 |
| Total | 3270 | 290 | .09 |

Table 4: Proportion of Choices Switched

8.9% were switched while 2256 of 3688 inconsistent choices or 61.2% were switched (see Table 4). This implies respondents are not simply being consistent with their previous choices but, rather, attempting to express their true preferences.

# 5 Methodology and Results

The above analysis provides several testable predictions. First, the asserted inverse relationship between choice consistency and the scale of a random utility model is verified. This result is interpreted as preference refinement. Second, the model of preference uncertainty predicts the probability of an inconsistent choice declines with both a greater utility difference between the involved items and a narrowing of the scale.

Further, preference reversals can be investigated with the retested data. First, inconsistent choices that were switched are recycled into the original data set to reflect refined preferences. This data set is again tested for refinement with a heteroscedastic probit. Second, the paper considers the likelihood of a preference reversal for both inconsistent and consistent choices when retested.

## 5.1 Choice Consistency and Scale

Results from the paired comparison choice experiment suggest that as respondents gain experience expressing their preferences they are less likely to commit an inconsistent choice. A heteroscedastic probit model is proposed to test the conjecture that increased choice consistency is associated with a reduction in the scale of the random utility model. Again the valuation function is represented as follows.

$$u_{ijk} = \alpha_k + \epsilon_{ijk} \tag{14}$$

$$\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2(j)) \tag{15}$$

Where the index $i$ denotes the individual and $j$ denotes the choice occasion. Note that the data is set up in rows and columns, as such the item index will be $k = r, c$ for row or column. The row contains only the ten items while the column contains the ten items along with the 11 monetary amounts. The monetary amounts which only appear in the column are denoted by $t_{ijc}$. Note that the stochastic component of utility, $\epsilon_{ijk}$, as specified is not a composite error term including both researcher and respondent error. Unlike Li and Mattsson (1995) this experiment did not directly measure uncertainty. Therefore a composite error term cannot be identified. Instead the scale of the model, $\sigma_\epsilon(j)$, will be a function of choice occasion $j$ in order to reflect preference uncertainty and preference refinement. The paper assumes that all respondents are identical and that they have the same valuation on each $\alpha_k$, the mean is assumed to be stable over choice occasion. The probability contribution to the likelihood function can now be constructed, $P_{rc}$ ($P_{cr}$) is the probability that the row (column) item is chosen over the column (row) item. Consider a choice between two items.

$$P_{rc} = Pr(u_{ijr} > u_{ijc}) = Pr(\alpha_r + \epsilon_{ijr} > \alpha_c + \epsilon_{ijc}) \quad (16)$$

$$P_{rc} = \Phi((\alpha_r - \alpha_c)/\sqrt{2}\sigma_\epsilon(j)) \quad (17)$$

The choice between an item and a dollar amount.

$$P_{rc} = Pr(u_{ijr} > t_{ijc}) = Pr(\alpha_r + \epsilon_{ijr} > t_{ijc}) \quad (18)$$

$$P_{rc} = 1 - \Phi((t_{ijc} - \alpha_r)/\sigma_\epsilon(j)) \quad (19)$$

As before $\Phi$ represents the cumulative distribution function of a normal random variable and $\sqrt{2}\sigma_\epsilon(j)$ is the standard deviation of $\epsilon_{ijc}-\epsilon_{ijr}$. Particular attention must be paid to the functional form of the scale, $\sigma_\epsilon(j) = \lambda+\beta(1/j)$[8]. This function will either decrease to the level of $\lambda$ with choice occasion or increase depending on the sign of $\beta$. Interpretation of these parameters is as follows. Researcher error, $\lambda$, is hypothesized to be constant. This parameter will also pick up any error generated by the respondent unrelated to choice occasion. A significant $\beta$ coefficient implies a significant magnitude of preference uncertainty as it represents a significant change in the scale of the model through choice sequence. As no other aspects of the experiment are changing a positive $\beta$ represents refinement while a negative $\beta$ represents fatigue or boredom. The hypothesis to be tested is $H_o : \beta = 0$ implying that choice sequence has no affect the scale of the model.

The items are grouped by type, either public or private, so that all items within type are assumed to have a single scale. Grouping the data smoothes the data across choice occasion. Although the data is randomized across

[8] This functional form is chosen by estimating the change and shape of the scale using two common methods. First the ratio of the standard deviation is estimated for different subsets of the data (Swait & Louviere, 1993) and second an equivalent method is used to estimate the change between groups of choices. Both methods suggest a function that decreases and levels off over the experiment.

choice occasion for each respondent the data tends to cluster reducing the observations across choice occasion. With this grouping there are five separate contributions to the likelihood function each representing a type of choice. These include public good versus public good, public good v. money, public v. private, private v. private, private v. money. Consider the choice between two public goods.

$$P_{rc} = \Phi((\alpha_r - \alpha_c)/\sqrt{2}(\lambda_b + \beta_b(1/j))) \tag{20}$$

The choice between a public good and a monetary amount.

$$P_{rc} = 1 - \Phi((t_{ijc} - \alpha_r)/(\lambda_b + \beta_b(1/j))) \tag{21}$$

A similar statement can be made for choices between two private goods and between a private good and a monetary amount. The difference being that the scale is defined as $\lambda_p + \beta_p(1/j)$. Note that the above assumes that the i.i.d. assumption holds within choice type so all public goods have the same standard deviation and that choices between public goods have no covariance. Finally, consider the choice between a public good and a private good, or as they will be referred to, mixed good choices. The scale here is $((\lambda_b + \beta_b(1/j))^2 + (\lambda_p + \beta_p(1/j))^2)^{1/2}$ which maintains the independence assumptions but relaxes the identical assumption across item type.

The sample is pooled over all individuals $i$ and all choice occasions $j$ so the likelihood function now takes form. The dependent variable $y_{ijk}$ equals 0 if the row item is chosen and 1 if the column item is chosen.

$$L(y_{ijk}; \alpha_k, \lambda_b, \beta_b, \lambda_p, \beta_p) = \prod_i^n \prod_j^J P_{rc}^{1-y_{ijk}} P_{cr}^{y_{ijk}} \tag{22}$$

| | Public | Private | Pooled |
|---|---|---|---|
| $\lambda$ | 370 | 203 | 296 |
| | (57.9) | (57.3) | (80.1) |
| $\beta$ | 334 | 305 | 340 |
| | (3.4) | (4.5) | (5.2) |

Table 5: Heteroscedastic Probit Parameters

The log likelihood function.

$$lnL(y_{ijk}; \alpha_k, \lambda_b, \beta_b, \lambda_p, \beta_p) = \sum_i^n \sum_j^J [(1 - y_{ijk}) lnP_{rc} + y_{ijk} lnP_{cr}] \tag{23}$$

The results in Table 5 are consistent with our expectations. First $\lambda_b = 370$ and $\beta_b = 334$ while $\lambda_p = 203$ and $\beta_p = 305$, all coefficients are significant (p < .01) t-statistics are in parentheses. These results were used to create Figure 3, the scale declines and levels off, similar to the proportion of choices which were inconsistent in Figure 1. The positive and significant $\beta$ in both the public and private subsets supports the interpretation of this reduction as preference refinement. 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.

### 5.1.1 Public Private Differential

Figure 3 shows a substantial difference between the scale of choices between public goods and those between private goods. It is reasonable to assume that there is more uncertainty when valuing public goods than private goods. This assertion is tested using a likelihood ratio test comparing a model which assumes only a single parametrization. This tests the restrictions $H_o : \lambda_b = \lambda_p = \lambda$ and $\beta_b = \beta_p = \beta$. This estimation is referred to as the pooled model, as shown in Table 5 $\lambda = 296$ and $\beta = 340$ again both are significant and

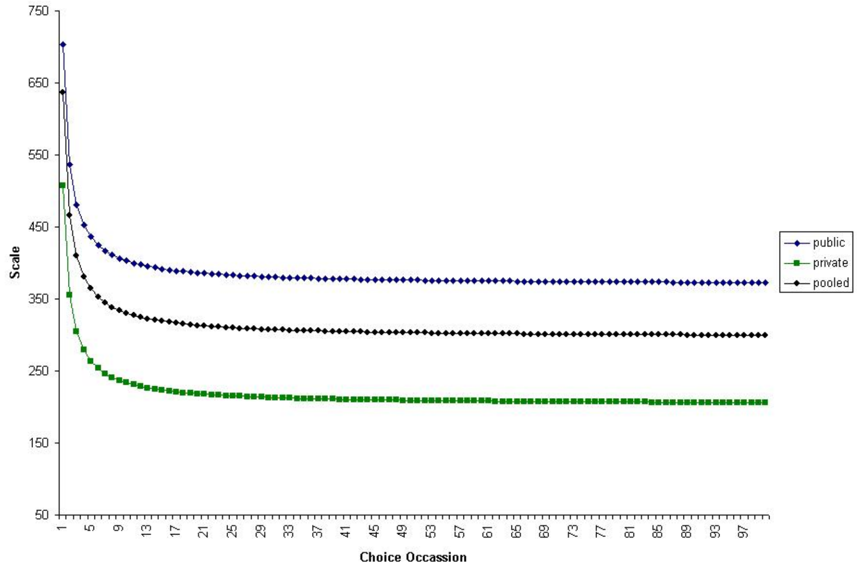


Figure 3: Graphical Representation of Preference Refinement

t-statistics appear in the parentheses. The restricted log likelihood equals $lnL_r = -26,544$ while the unrestricted log likelihood equals $lnL_u = -26,175$. Thus $-2(lnL_r - lnL_u) = 738$, this is compared to the 95% confidence measure for a $\chi^2_6 = 12.59$. Therefore we can reject the restrictions requiring a single parametrization. This suggests a significant difference between the scale of choices involving public goods than those involving private goods as expected.

## 5.2 Probability of an Inconsistent Choice

The model of preference uncertainty allowing for preference refinement predicts that an inconsistent choice becomes less likely as the scale narrows or the greater is the utility difference between the involved pair. This is tested using a probit model where the dependent variable, $y_i$, equals 1 if the choice is identified as inconsistent and 0 otherwise. Using the above result choice occasion serves as a proxy for the scale of the model. An increase in choice

| | (1) | (2) |
|---|---|---|
| Constant | -.36 | -.44 |
| | (-15.3) | (-16.6 |
| Choice Occasion | -.002 | -.002 |
| | (-8.8) | (-6.9) |
| Preference Score Difference | -.22 | -.22 |
| | (-56.9) | (-56.3) |
| Public | .05 | .05 |
| | (2.4) | (2.2) |
| Time | | .01 |
| | | (6.7) |

Table 6: Probability of an Inconsistent Choice

occasion represents a reduction in the scale. The preference score difference (PSD) between the items involved is used as an approximate measure of the utility difference. In addition a dummy variable is included indicating that the choice involves a public good. A probit model is developed to predict the probability of an inconsistent choice.

$$Pr(y_i = 1) = \Phi(x_i'\beta) \tag{24}$$

and

$$Pr(y_i = 0) = 1 - \Phi(x_i'\beta) \tag{25}$$

All 50,685 choices are pooled over individuals $i$ and choice occasion $j$ so the log likelihood becomes.

$$lnL(y_i;\beta) = \sum_i^n \sum_j^J [y_i ln\Phi(x_i'\beta) + (1 - y_i) ln(1 - \Phi(x_i'\beta))] \tag{26}$$

Results are shown in Table 6 and support the intuition of the model (t-statistics are in parentheses). Choice occasion is negative and significant reflecting that an inconsistent choice is less likely as the respondents progress

| | Public | Private | Pooled |
|---|---|---|---|
| $\lambda$ | 336 | 175 | 270 |
| | (66.3) | (64.7) | (91.1) |
| $\beta$ | 21 | 20 | 13 |
| | (.5) | (1.0) | (.5) |

Table 7: Retested Heteroscedastic Probit Parameters

through the experiment. A negative and significant preference score difference reflects that inconsistency is less likely the greater the utility difference of the involved items. Additionally choices including a public good, *Public*, are more likely to be inconsistent. The second column of Table 6 also controls for the amount of time taken for the choice, *Time*. The longer the choice took to make the more likely it is inconsistent. The amount of time taken reflects the difficulty of the choice, this may stem from greater uncertainty in the choice or indifference between the items.

### 5.3 Preference Reversals

Although the above results suggest a model of preference uncertainty allowing for preference refinement the data allow further investigation. All inconsistent choices are retested while 10 consistent choices per respondent are retested. A total of 3688 individual choices were identified as inconsistent (2256 or 61% were switched on retrial) while 3270 consistent choices were retested (290 or 9% were switched on retrial).

First, originally inconsistent choices that are switched when retested can be recycled into the original data set and the relationship between choice consistency and the scale of the random utility model can again be investigated. Recycling choices that were reversed with the benefit of preference refinement ought to reduce the magnitude of respondent error. In particular the scale is hypothesized to be constant across choice sequence. This is tested using a heteroscedastic probit but uses the data set obtained when the 2256

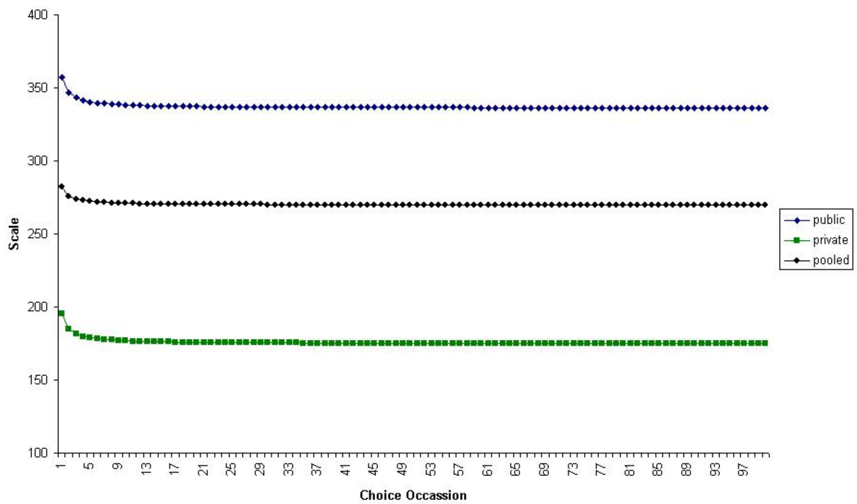


Figure 4: Graphical Representation of Retested Heteroscedastic Parameters

reversed inconsistent choices are in place.

Results are shown in Table 7 and further support preference refinement. Although respondent error is thought to exist within $\lambda$ it is no longer significantly effected by choice occasion, $\beta$ is no longer significant. Again Figure 4 depicts a graphical representation of these results.

Second, the data show substantial differences in the reversal rates of originally inconsistent and consistent choices. Originally inconsistent choices appear more likely to be switched when retested. To test this relationship two probit models are developed. For the inconsistent case the dependent variable, $y_i$, equals 1 if the inconsistent choice is switched and 0 otherwise. Similarly in the consistent subset the dependent variable equals 1 if the choice is reversed and 0 otherwise. Two exogenous variables are included, the PSD and the public good dummy. The expectation is that choices involving public goods are more likely to be reversed reflecting greater uncertainty. Greater

| | Inconsistent | Consistent |
|---|---|---|
| Constant | -.09 | -1.58 |
| | (-3.3) | (-33.8) |
| Preference Score Difference | -.11 | -.04 |
| | (-23.8) | (-5.7) |
| Public | .14 | .03 |
| | (4.4) | (.29) |

Table 8: Probability of a Preference Reversals

PSD is expected to increase the likelihood of reversal for inconsistent choices and decrease this likelihood for consistent choices.

Table 8 displays the results. The consistent subset results are as expected. The greater the preference score difference between the items the less likely an originally consistent choice is switched. However, the results from the inconsistent subset also imply that the PSD has a negative effect on the probability that an originally inconsistent choice is switched.

# 6 Conclusion and Discussion

Paired comparison choice experiments allow researchers to measure within individual choice consistency and to retest likely inconsistent choices. Using data from Peterson and Brown (1998) this paper supports the inverse relationship between the scale of a random utility model and choice consistency (Deshazo & Fermo, 2002). Within a model of preference uncertainty where preference are realizations from a valuation distribution this result is interpreted as a reduction in respondent error or preference refinement. Furthermore results suggest that choice inconsistency is less likely as the utility difference between the pair of items increases and as the standard deviation of the valuation distribution narrows. These results are consistent with the predictions of the model and intuition.

Preference Refinement is further supported by recycling inconsistent choices that were reversed when retested into the data set. Results, using this retested data, suggest that the scale is constant with choice occasion having no impact. Results predicting the occurrence of preference reversals are mixed. Both originally consistent and inconsistent choices are less likely to be switched the greater the utility difference between the pair. Intuition would suggest that greater utility difference would increase the probability of reversing an originally inconsistent choice.

The existence of preference uncertainty and preference refinement in choice experiments directly relates to recent research investigating the affect market experience and experimental design have on choice behavior (List, 2003; Cherry, Crocker, & Shogren, 2003). Under what conditions or choice environments can economists expect the rationality of preferences? Does the market reveal and reward underlying rationality or does it create rational agents? On-going work investigates the role that preference uncertainty and preference refinement may have on the common disparity between willingness to accept and willingness to pay. Results show that respondents participating in a paired comparison choice experiment prior to value elicitation do not report a significant disparity (Kingsley, 2006b).

# References


Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand* (M. L. Manheim, Ed.). The MIT Press.

Bock, D. R., & Jones, L. V. (1968). *The measurement and prediction of judgment and choice* (R. R. Bush, Ed.). Holden Day Series in Pyschology.

Bockstael, N. E., & Strand, I. E. (1987). The effect of common sources of regresion error on benefit estimates. *Land Economics*, *63*(1), 11-20.

Brown, T. C., Kingsley, D., Peterson, G. L., Flores, N., Clarke, A., & Birjulin, A. (2006). Reliability of individual valuations of public goods and private goods. *Working Paper, Rocky Mountain Research Station.*

Champ, P. A., Bishop, R. C., Brown, T. C., & McCollum, D. W. (1997). Using donation mechanisms to value nonuse benefits from public goods. *Journal of Environmental Economics and Management.*, *33*, 151-162.

Champ, P. A., & Loomis, J. B. (1998). Wta estimatesusing the method of paired comparison: Tests of robustness. *Environmental and Resource Economics*, *12*, 375-386.

Cherry, T. L., Crocker, T. D., & Shogren, J. F. (2003). Rationality spillovers. *Journal of Environmental Economics and Management*, *45*(1), 63-84.

Crocker, T. D., & Shogren, J. F. (1991). *Preference learning and contingent valuation methods in environmental policy and the economy* (F. Dietz, F. V. D. Ploeg, & J. V. D. Straaten, Eds.). Elsevier Science Publishers B.V.

David, H. (1969). *The method of paired comparisons* (A. Stuart, Ed.). Charles Griffin & Company Limited.

Deshazo, J., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*, 123-143.

Evans, M. F., Flores, N. E., & Boyle, K. J. (2003). Multiple-bounded uncertainty choice data as probabilistic intentions. *Land Economics*, *79*(4), 549-560.

Kingsley, D. C. (2006a). Multiple good valuation using piared comparison choice experiments. *University of Colorado at Boulder Working Paper.*

Kingsley, D. C. (2006b). The role of choice experience in the willingness to pay and willingness to accept disparity: An experimental test. *University of Colorado at Boulder Working Paper.*

Li, C.-Z., & Mattsson, L. (1995). Discrete choice under preference uncertainty: An improved structural model for contingent valuation. *Journal of Environmental Economics and Management*, *28*, 256-269.

List, J. A. (2003). Does market experience eliminate market anomalies? *Quarterly Journal of Economics*, *Volume 118*(Number 1), pp. 41-71.

Loomis, J., Peterson, G., Champ, P., Brown, T., & Lucero, B. (1998). Paired comparison estimates of willingness to accept versus contingent valuation estimates of willingness to pay. *Journal of Economic Behavior & Organization*, *35*, 501-515.

Manski, C. (1973). The analysis of qualitative choice. *Ph.D. Dissertation. Department of Economics, Massachusetts Institute of Technology.*

Marschak, J. (1960). Binary-choice constaints and random utility indicators. *Mathematical Methods in the Social Sciences*, *155*, 312-329.

McFadden, D. (2001). Economic choices. *The American Economic Review*, *91*(3), 351-378.

Peterson, G. L., & Brown, T. C. (1998). Economic valuation by the method of paired comparison, with emphasis on the evaluation of the transitivity axiom. *Land Economics*, *74*(2), 240-261.

Peterson, G. L., & Brown, T. C. (2006). An enquiry into the method of paired comparison: Reliability, scaling, and thurstones law of comparative judgment. *Discussion Paper, Rocky Mountain Research Station, U.S. Forest Service.*

Plott, C. R. (1996). Rational foundations of economic behavior. In K. Arrow, E. Colombatto, M. Perleman, & C. Schmidt (Eds.), (p. 225-250). London: Macmillan and NY: St. Martin's.

Savage, S. J., & Waldman, D. M. (2004). Learing and fatigue during choice experiments: A comparison of online and mail survey modes. *Working Paper*.

Shogren, J. F., List, J. A., & Hayes, D. J. (2000). Preference learning in consecutive experimental auctions. *American Journal of Agricultural Economics*, *82*(2), 1016-1021.

Swait, J., & Adamowicz, W. (1996). The effects of choice environment and task demands on consumer behavior: Discriminating between contribution and confusion. *Working Paper*, *96-09*.

Swait, J., & Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, *30*, 305-314.

Thurstone, L. (1927). A law of comparative judgment. *Pyschological Review*, *34*, 273-286.

Torgerson, W. S. (1958). *Theory and methods of scaling*. John Wiley and Sons, Inc.

Wang, H. (1997). Treatment of "don't know" responses in contingent valuation surveys: A random valuation model. *Journal of Environmental Economics and Mangement*, *32*, 219-232.

Welsh, M. P., & Poe, G. (1998). Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach. *Journal of Environmental Economics and Management*, *36*(2), 170-185.