**DEPARTMENT OF ECONOMICS
COLLEGE OF BUSINESS AND ECONOMICS
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND**

**THE CAUSES OF ORDER EFFECTS IN CONTINGENT
VALUATION SURVEYS: AN EXPERIMENTAL INVESTIGATION**

by Jeremy Clark and Lana Friesen

WORKING PAPER

No. 06/2006

**Department of Economics, College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand**

WORKING PAPER No. 06/2006

THE CAUSES OF ORDER EFFECTS IN CONTINGENT VALUATION SURVEYS: AN EXPERIMENTAL INVESTIGATION

by Jeremy Clark¹ and Lana Friesen²

January 16th, 2006

ABSTRACT. CV researchers have found that the hypothetical values respondents place on a nested sequence of environmental goods are sensitive to the order in which the goods are presented. Typically, the smallest bundle of goods is valued more highly if presented first than if following more comprehensive bundles. Such effects appear even when each bundle is valued from an “exclusive” list, or as an alternative to any other, so that income and substitution effects are controlled. Order of presentation has also affected the degree to which values are sensitive to scope. We conduct lab experiments where participants are asked to value sequences of nested goods for actual purchase from an exclusive list using the incentive compatible BDM mechanism. We test whether order effects occur in valuation for a) induced value goods, b) actual private goods, and c) identical private goods that are to be donated to charities. We find significant order effects when the goods are valued for own use, but not when they are valued for donation.

Keywords: Order effects, exclusive list, warm glow, contingent valuation

¹ Department of Economics, College of Business & Economics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. jeremy.clark@canterbury.ac.nz. Phone 011 643 364 2308 Fax: 011 643 364 2635

² Commerce Division, Lincoln University, PO Box 84, Canterbury, New Zealand. friesenl@lincoln.ac.nz. Phone: 011 643 325-3627 Fax: 011 643 325-3847

1. Introduction and Literature Review

In recent years arguments over the reliability of contingent valuation (CV) have come to focus on the method's construct validity, or the internal consistency of its answers according to economic theory. Striking valuation anomalies have been identified, prompting CV researchers to examine more closely what economic theory does and does not predict. This in turn has led to more precisely formulated validity tests and nuanced discussions of the circumstances under which CV responses can be, if not accepted by all, at least internally consistent.

A prime example of this winnowing process has been the “embedding” effect named by Kahneman and Knetsch (1992), and previously identified by Randall et al. (1981). These authors found that the value respondents place on an environmental good seems to depend on whether it is valued in isolation, or as part of a more inclusive good. Others such as Brookshire (1981) found that the value respondents assigned to a good (air quality) could vary according to the order in which that quantity appeared relative to others. In response, researchers such as Randall and Hoehn (1996) identified conditions under which substitution effects and constrained endowments could rationally result in embedding effects. Carson, Flores and Hanemann (1998) showed that if rationed public goods are strict Hicksian substitutes and normal goods, willingness to pay (WTP) for them should decrease, the further out in a sequence they are valued. As the debate progressed, Carson and Mitchell (1995) proposed that “embedding” effects be divided into the more precise “scope” and “sequence” effects. With regard to scope, the values respondents place on an environmental project should be weakly increasing in its scope or size (though not strictly increasing if satiation is a possibility). Sensitivity to scope can

be tested internally, by asking the same subjects to value a nested sequence of goods, or externally, by asking different subjects to value different levels of the sequence. With regard to sequence, the value respondents place on an environmental (public) good presented in a nested sequence should depend on the order of presentation because of substitution and income effects. However the total valuation of all projects in the sequence should be independent of order.

Regarding empirical tests of scope, Carson (1997) surveys over 30 studies that use external split samples to test explicitly for scope insensitivity, and reject it. Carson cites an additional handful of well-known exceptions such as Kahneman and Knetsch (1992), Diamond et al. (1993) and Desvousges et al. (1992, 1993) that do not reject insensitivity to scope, but criticizes their survey design, sample size, or the low power of their statistical tests. More recently, however, Heberlein et al.(2005) have criticized conventional scope tests that compare average valuation results across samples for masking false positive and negatives at the individual level.

Regarding empirical tests of sequence effects, the ability of researchers to evaluate the validity of CV took a major step forward with Bateman et al.'s (2004) observation that multiple valuations could be made from an *inclusive* or *exclusive* list. In inclusive lists, each subsequent good is presented as an addition to goods already presented and valued. In contrast, in exclusive lists, each good is presented as a mutually exclusive alternative to any other good presented. While substitution and income effects imply that the valuation of a good from an inclusive list should depend on sequence, they should be controlled for in valuation from an exclusive list. Furthermore, exclusive lists should also maintain reference utility at the status quo from which each good is proposed

as an alternative, so that loss aversion or reference dependent preferences (Tversky and Kahneman (1991)) should not influence valuation according to order. Economic theory thus provides a clean prediction that the valuation of individual goods should not be affected by the sequence in which they are presented if they appear on a mutually exclusive list. Bateman et al. (2004) thus distinguish “sequence effects” in valuations from inclusive lists (which should be expected) from “order effects” in valuations from exclusive lists (which should not).

The evidence regarding order effects from exclusive lists in CV is sparse, in part because earlier studies do not make it clear if respondents valued multiple projects from an inclusive or exclusive list. Boyle et al. (1993) find that inexperienced white water boaters exhibited order effects in valuing alternative levels of water flow in the Grand Canyon, but that experienced boaters did not. Bateman and Langford (1996) and Payne et al. (2000) both find strong sequence/order effects in valuations of multiple payment vehicles or multiple environmental goods. The latter authors find that WTP is much larger for the first project evaluated than for those evaluated subsequently. However neither study makes it clear whether respondents knew they were valuing each proposal as alternatives or incremental additions.

The clearest evidence concerning order effects from exclusive list CV comes from Powe and Bateman (2003) and Bateman et al. (2004). Powe and Bateman find significant order effects in WTP for the prevention of salt water flooding in the Norfolk Broads (wetlands) in eastern England. In particular, Powe and Bateman used a dichotomous choice CV survey to ask visitors to the Norfolk Broads their WTP to prevent saline flooding for a smaller (nested) and larger part of the wetlands. WTP to

prevent saline flooding in the smaller part was higher if it was elicited prior to that for the larger part, than if it was elicited after. To a lesser but still significant degree, WTP to prevent saline flooding in the larger area was larger if it followed the elicitation of WTP for a part, rather than preceding it. Powe and Bateman attribute the order effects either to strategic misrepresentation of WTP, or “non preference based” explanations such as the purchase of moral satisfaction (Kahneman and Knetsch (1992)) or anchoring and adjustment heuristics (Kahneman et al. (1982)).

Bateman et al. (2004) conduct an analogous exercise as Powe and Bateman, but using open ended contingent valuation, and with the added distinction of using “stepwise” or “advanced” disclosure to respondents of the sequence of goods they would be asked to value. Bateman et al. (2004) first conduct a CV survey of a student population for three nested levels of environmental improvement for a lake adjacent to the University of East Anglia. They then conduct two field CV tests, the first using advanced disclosure concerning three levels of water quality in the River Tame at Birmingham. The second field CV test used stepwise disclosure, again concerning the prevention of saline flooding at two levels of geographic extent in the Norfolk Broads.

As with Powe and Bateman, Bateman et al. find that even with exclusive valuation, WTP for a specified good depends on the order in which it is presented. As before, the WTP for the smallest scope of a project was greater when that level was presented first, rather than following projects of greater scope. This was true both of the student sample, and of the sample of visitors to the Norfolk Broads. Interestingly, however, Bateman et al. find that all order effects were eliminated in those treatments where respondents were given advanced descriptions of all levels of the projects before

being asked to provide values for any. In related findings, the degree of sensitivity of WTP to scope was dramatically affected by the sequence of presentation in stepwise disclosure treatments, but not in advanced disclosure treatments.

In discussing the order effects under stepwise disclosure, Bateman et al. (2004) concede they could be due to the purchase of moral satisfaction, but query why such effects do not then persist under advanced disclosure. Alternatively, the authors note that strategic incentives to misreport WTP could differ during the stepwise treatment but remain constant under advanced disclosure, or that perhaps loss aversion could be playing a role (despite reference utility being held constant). Finally, the authors speculate that there is a psychological element of surprise in stepwise disclosure absent in advanced disclosure, because respondents are suddenly asked their WTP for a greater (or lesser) project than was originally described. Disgruntled (relieved) respondents could respond to the implication that they are expected to pay yet more (less) by raising their WTP for projects of greater scope by less than they otherwise would, and lowering their WTP for projects of lesser scope by less than they otherwise would.

While limited in number, the above studies suggest that the construct validity of contingent valuation remains open to doubt regarding order effects. In cases where economic theory clearly predicts invariance, researchers have found that the valuation given for a specific project seems dependent on the order in which it is presented relative to projects of greater or lesser scope. This occurs even when alternative projects are clearly presented as mutually exclusive alternatives to the status quo.

This paper's contribution is to use laboratory experiments to shed light on the possible explanations that have been offered for order effects observed in CV. In

particular, we shall test whether order effects persist in the valuation of nested goods from an exclusive list even when 1) the goods are private, 2) strategic considerations are eliminated, 3) decisions are non-hypothetical, and 4) the potential for warm glow/moral satisfaction is controlled for. We ask laboratory subjects their maximum WTP for three levels of a nested good, using the incentive compatible Becker-DeGroot-Marschak (BDM) mechanism (Becker et al. (1964)). Using stepwise disclosure, half the subjects see and value the good bundles from smallest to largest (“bottom up”), and half from largest to smallest (“top down”). A control treatment elicits maximum WTP for private, induced value goods, to ensure that subjects understand the incentive compatible properties of the BDM. A second treatment elicits bids for private actual goods. A third treatment elicits bids for the same actual goods, but in order for them to be donated to specific charities.

We find that 1) order effects persist even for private market goods that are actually purchased, and that 2) the introduction of strong warm glow considerations in the formation of WTP eliminates these effects.

2. Experimental Design

2.1 Eliciting Valuations

A growing number of researchers are using incentive compatible auction mechanisms in experiments to elicit valuation responses in order to shed light on CV preference anomalies (Noussair et al. (2004), Bateman et al. (1997a), Bateman et al. (1997b), Coursey, Hovis and Schulze (1987), Shogren et al. (1994)). Economists hope that valuations provided with strategic incentives to be truthful will be more reliable than

those asked hypothetically, with ambiguous strategic incentives. Better incentives are purchased at a cost, however, as these experimental auctions are carried out for excludable (private) goods far removed from typical CV applications, and are typically run with paid student volunteers rather than random samples from the general population.

Popular mechanisms used have included second (or n^{th}) price sealed bid auctions (Vickrey (1961)), the sequential English auction, which is demand revealing for all but the highest bidder, and the Becker-DeGroot-Marschak (BDM) mechanism (Becker et al. (1964)). Under the BDM, subjects simultaneously submit bids for a good, and a sale price is then drawn at random from a pre-specified distribution. All those whose bids equal or exceed the sale price receive the good, and pay the sale price. The rest pay nothing and do not purchase the good. Bidders have a dominant strategy to bid an amount equal to their true maximum WTP, regardless of risk preference or beliefs about others' strategies.

The BDM has come to be favoured for applications where researchers want subjects to make sequential valuations without being influenced by others' valuations as conveyed by the winning bid in previous rounds (Bateman et al. (1997a), (1997b), Noussair et al. (2004)). A sale price determined at random conveys no such information. One drawback of the BDM is its counter-intuitive nature; subjects must be convinced that it is truly in their interests to bid their maximum WTP. Combining approaches from previous researchers (Davis and Holt (1993, p. 490), Noussair et al. (2004)), we spend considerable time asserting, explaining, and demonstrating the incentive properties of the BDM to subjects during four practice rounds that preceded valuation decisions in all treatments of the experiment. These will be described in greater detail below. Three of

the four practice rounds used induced value goods, and our control treatment used induced value goods throughout the valuation sequence. In this way, we can verify that the BDM is eliciting subjects' maximum WTP in practice as well as in theory.

2.2 Order effects in nested good valuation

Our experimental design testing for order effects follows closely the CV design used by Bateman et al. (2004). In the main decision rounds, subjects were endowed with (Can.) \$22, and asked their “maximum purchase price” for a nested sequence of three bundles of goods. The valuations were from an exclusive list, in that the experimental instructions emphasized that a market would be implemented only for *one* of the three bundles valued, to be chosen at random at the end of the experiment. The sequence of goods was described and valued in stepwise fashion.¹ Subjects were given prior experience with this exclusive list procedure, in that only one of the four practice rounds, chosen at random, was implemented before the main decision rounds. Random draws concerning round selection and sale price were carried out with a bingo cage.

Half of subjects were asked to value a nested sequence of goods from smallest to greatest, or “bottom up”, and half from greatest to smallest, or “top down.” For those subjects in the control Treatment I, the goods in question all had induced values. In bottom up order, they were 1) a green slip of paper, 2) a green and a yellow slip of paper, and finally 3) a green and a yellow and a blue slip of paper. The induced value of each coloured slip differed across subjects, and each was drawn from the uniform distribution

¹ To avoid any element of unethical “surprise,” we followed Payne et al. (2000) in informing subjects at the outset that they would be making three valuation decisions, but without giving any advanced indication as to the nature of the goods or bundles to be valued. This actually moves our design a small step towards advanced disclosure, which from Bateman et al. (2004) should make order effects *less* likely to occur.

[\$0, \$7.00] in ten cent increments.² Thus the combined value of the sequence of goods in bottom up order was contained in the ranges [\$0, \$7], [\$0, \$14] and [\$0, \$21], respectively. Subjects knew that induced values varied across participants, but not the distribution or range from which they were determined.

For those subjects in Treatment II, induced value goods were replaced by actual market goods for the main decision rounds. In bottom up order, the nested sequence was

1. A 10 pack of 200 ml drink boxes of *Minute Maid* brand orange juice, or **J**.
2. **J** plus a stationary set comprised of a pack of five *Staples* brand retractable gel ink pens (two black, one blue, one red and one green) and one *BiC Wite-Out* brand correction tape, or **J,S**.
3. **J, S** plus a *Media* brand 27 exposure disposable 35 mm camera with built in flash, or **J,S,C**.

Prior to each valuation decision, the relevant bundle of goods was displayed on a tray and presented to each subject individually for his or her inspection.

For those subjects in Treatment III, the same goods were presented as in Treatment II. Here however, participants were asked their maximum purchase price from their \$22 endowments to donate the bundles of goods to specific charities. If the “market” for a particular bundle was implemented, those whose maximum purchase price for donation equaled or exceeded the sale price would pay the sale price to the experimenter, and that bundle of goods would be donated to the appropriate charities. If not, the person would keep the entire \$22 endowment, and the good(s) would not be donated on his or her behalf. The charities receiving the donations were as follows:

² The same induced values that were drawn randomly for the participants in a "bottom up" session were then also used in reverse order for participants in a "top down" session.

1. **J** went to the Calgary Inter-Faith Food Bank, the single largest food bank in Calgary, Alberta.
2. **S** went to students at the Salvation Army's Children's Village, an elementary school in Calgary for children with severe behavioural/emotional difficulties
3. **C** went to children at the Salvation Army's Pine Lake Camp, a summer camp near Red Deer, Alberta for children from disadvantaged families who could not otherwise afford summer camp.

Following Eckel and Grossman (2003), care was taken to ensure the credibility of the claim that donations of goods would in fact be carried out. Aside from general assurances at the outset that deception was not used and that the experiment passed human ethics review, letters from the appropriate charities acknowledging their participation in this research exercise were shown to subjects prior to making each valuation decision. Pamphlets describing each charity's work were also provided when the goods for donation were displayed.³

Sale prices for all bundles of goods were drawn from the uniform distribution, with ranges to be explained. For the four practice rounds that were common to all three treatments, the items valued respectively were a mini Mars Bar, and then three induced value coloured slips of paper. The induced values of these slips of paper varied between subject and between rounds, but (unbeknownst to subjects) were drawn from the uniform distribution between [\$0, \$2]. For each practice round subjects were informed that the potential sale price would be drawn from the range [\$.05, \$2.00] in five cent increments. For the three main decision rounds in all three treatments, subjects were informed that the

³ Subjects in Treatment III were also asked in a post experimental questionnaire "when making your decisions today, did you feel confident that any items you purchased for donation in the experiment really would go to the charities as described?" 92 of 95 subjects answered the question, of whom 96% said "yes."

sale price would be drawn from the [\$.10, \$7.00] range in ten cent increments for the smallest bundle, the [\$.20,\$14.00] range in twenty cent increments for the intermediate bundle, and the [\$.25, \$21] range in twenty-five cent increments for the largest bundle. Altering the range of sale prices for larger bundles has the potential to signal subjects that we expect bids to rise with the scope of the bundle (Bohm et al. (1997)), but ensured that subjects had a non-trivial probability of purchasing even the smallest scope of good. Endowments were set to exceed the maximum possible sale price in all cases, while the latter were set high enough that subjects would in all likelihood choose to bid below them.⁴ This ensured that subjects could secure the probability of purchasing a bundle with probability one, but would be unlikely to do so.

Sessions were run with 12 subjects at a time, and lasted roughly 90 minutes. Subjects were paid a \$3 endowment up front from which to make bids during the four practice rounds, and \$22 in advance from which to make bids during the three main valuation rounds.⁵ Final earnings were thus \$25 per subject, plus any goods purchased, less their sale price(s). 235 enrolled student subjects were recruited from the University of Calgary between February 16th and October 5th 2005. The allocation of subjects to treatments is given in Table 1.

⁴ While not announced to subjects, the retail prices of the juice, stationary and camera were (Canadian) \$3.67, \$7.00 and \$6.13 respectively. Transactions costs aside, the retail prices of the nested bundles (bottom up) were thus \$3.67, \$10.67 and \$16.80.

⁵ Subjects signed consent forms agreeing that they could keep the endowments only if they stayed for the duration of the experiment. Their student ID cards were held for the duration of the experiment as security.

Table 1
Experimental design and sub-sample labels

		List Direction	
		Bottom Up (B)	Top Down (T)
Treatment I		S _I B	S _I T
Self, Induced Value	(S _I)	n = 20	n = 24
Treatment II		SB	ST
Self, Actual	(S)	n = 48	n = 48
Treatment III		CB	CT
Charity, Actual	(C)	n = 47	n = 48

3. Results

3.1 Learning to report maximum valuation in the BDM

There is encouraging evidence that subjects mastered the incentives of the BDM mechanism under our comprehensive approach. The experimental instructions explicitly informed subjects in bold that “it will always be in your best interest to give your true *maximum* value,” and gave the reasoning why using two symmetric examples of over- and under-valuation, respectively. For each of the four practice rounds, each subject’s induced value and reported maximum purchase price were sorted by the former (if available) or the latter, and written on a large whiteboard. A sale price was then drawn and announced. Following Noussair et al. (2004), subjects were collectively asked 1) to identify who would receive the item if the market for that round were implemented, 2) what price the purchasers would pay, and 3) whether anyone would regret the purchase price that he or she reported. The experimenter pointed out examples of potential regret caused by over- or under-stated reports of maximum purchase price.

With this feedback provided, reported purchase prices converged towards induced value across the three practice rounds that used induced value goods. Pooling all subjects, the percentage who reported a maximum purchase price within $\pm \$0.20$ of induced value rose from 67% in the first practice round with induced values, to 75% in the second, to 86% in the third.^{6,7} The subsequent results from the main valuation decisions in Treatment I suggest that the practice experience with the BDM was fully exploited. With stakes raised, the percentage of Treatment I subjects who were within $\pm \$0.20$ of induced value was 84% for the first main bundle evaluated (whether the largest or smallest), 91% for the second, and 84% for the third. Alternatively, since a $\pm \$0.20$ difference between maximum purchase price and induced value seems proportionately less troubling at the greater scales of induced values used in the main rounds of Treatment I, we could instead ask whether subjects there reported a maximum purchase price within $\pm 10\%$ of induced value. 93% were within that band for the first main round, 96% were for the second, and 93% were for the third. Either measure increases our confidence that the three valuations of interest from subjects in Treatments II and III closely reflect their true maximum willingness to pay, to the best of their abilities.

3.2 Order Effects

Perhaps the most interesting results of this paper are that we find evidence of order effects similar to those found in CV, even in the non-hypothetical valuation of private market goods. Yet such effects are absent when the purchase of moral

⁶ The discrete five cent increments in the sales price range for the practice rounds meant that subjects could shave up to five cents off their bids or inflate by up to four cents without being strictly worse off. Analogous (bottom up) bands of $[-\$0.10, \$0.09]$, $[-\$0.20, \$0.19]$ and $[-\$0.25, \$0.24]$ exist in the three main rounds.

⁷ Appendix 1 provides a summary of the distribution of discrepancies between maximum purchase price and induced value for the final practice round.

satisfaction becomes a prominent consideration. Following the structure of analysis in Bateman et al. (2004), here is what we find.

Figure 1 illustrates the mean valuations of the three nested bundles of goods by order of presentation, first for induced value goods (a), then for actual goods for self (c), then for actual goods for charities (e). Because mean WTP can be skewed by outliers, corresponding medians are also shown in (b), (d) and (f). As we would expect from the success of the BDM, no order effects appear for induced value goods ((a) and (b)). In contrast, for actual goods (c) and (d) there is the familiar pattern that subjects on average are willing to pay more for smaller consumption bundles when considered before more inclusive bundles than when considered after, and more for larger consumption bundles when considered after less inclusive bundles than when considered before. In contrast, when the good(s) are being valued for donation to different charities, subjects on average appear willing to pay the same amount regardless of order of valuation. Also of interest is that (different) students on average were willing to pay almost as much to donate the particular goods to charities as they were to purchase them for themselves.⁸

Moving to numbers, descriptive statistics by means and medians are summarized in Table 2. As in Figure 1, mean and median valuations are very similar regardless of order in the induced value treatment. In contrast, when actual goods for own use are valued, there is a 43% mean WTP premium for **J** when it is valued first rather than last, a 32% premium for **J,S** when it valued after **J** rather than after **J,S,C**, and a 15% premium for **J,S,C** when it is valued after smaller bundles than when valued first. The

⁸ This may reflect the suitability of the goods for the charitable uses specified, and their lack of suitability to university students. Some commented in the post experiment questionnaire that they did not really need the goods on offer, so would not pay much for them. No such comments appeared for participants in Treatment III.

Mean WTP - Induced Value
(a)

Median WTP -- Induced Value
(b)

Mean WTP - Goods for Self
(c)

Median WTP -- Goods for Self
(d)

Mean WTP - Goods for Charities
(e)

Median WTP -- Goods for Charities
(f)

Fig. 1. Order Effects

Table 2
Mean and Median WTP for Induced, Private and Donated Bundles

Sample	Mean WTP (Can \$)			Median WTP (Can \$)		
	J ^a	J,S	J,S,C	J	J,S	J,S,C
S _I B N=20	3.18 (.430) ^b	5.87 (.475)	10.67 (.650)	2.90 (.539)	6.60 (.595)	11.90 (.814)
S _I T N=20 ^c	3.04 (.421)	5.91 (.478)	10.31 (.724)	2.73 (.528)	6.65 (.599)	11.90 (.907)
SB N=48	2.40 (.356)	4.58 (.463)	7.01 (.753)	2.00 (.446)	4.00 (.580)	5.68 (.943)
ST N=48	1.67 (.172)	3.48 (.354)	6.07 (.581)	1.50 (.216)	3.50 (.443)	5.00 (.729)
CB N=47	2.19 (.255)	3.85 (.416)	5.81 (.654)	2.00 (.320)	2.60 (.522)	5.00 (.820)
CT N=48	2.21 (.253)	3.67 (.442)	6.01 (.734)	1.73 (.317)	3.00 (.554)	5.00 (.920)

^a For the induced value Treatment, **J** refers to one slip of paper, **J,S** refer to two slips of paper, etc. ^b Standard errors. The standard error calculation procedure for medians assumes a large sample and normal distribution, and is given in Wonnacott and Wonnacott (1984, 677). ^c For purposes of fair comparison, the four valuation observations of subjects in S_IT with induced values set according to those of the four missing observations in S_IB are excluded.

corresponding median WTP premiums are 33%, 14%, and 14%. In contrast, order effects appear negligible and inconsistent for goods bundles being valued for donation. With regard to scope sensitivity, both the means and medians in Table 2 suggest that valuations are rising in scope, whether internally across rows, or externally comparing the diagonal first responses of (different) respondents for **J** and **J,S,C** within each treatment.

Turning to formal statistical analysis, we follow Bateman et al. (2004) in testing for order and scope effects in three different ways. We begin with univariate tests of differences in means using t tests, and in medians using Mann Whitney tests. We finish with multivariate analysis. For the effects of order on gross bundle valuation, we test hypotheses of the form:

$$WTP_{i,B}(X) = WTP_{i,T}(X) \quad i = S_I, S \text{ or } C, \quad X = \mathbf{J} \text{ or } \mathbf{J,S} \text{ or } \mathbf{J,S,C}. \quad (1)$$

To test for sensitivity to scope, we use external (split sample) tests of whether subjects provide identical valuations for the first bundle they see, regardless of size:

$$WTP_{i,B}(\mathbf{J}) = WTP_{i,T}(\mathbf{J,S,C}) \quad i = S_I, S \text{ or } C. \quad (2)$$

Finally, to test whether order influences the degree of scope sensitivity, we ask whether the marginal valuation of **S** or **C** or **S,C** together differs by order:

$$\begin{aligned} MWTP_{i,B}(\mathbf{S}) &\equiv WTP_{i,B}(\mathbf{J,S}) - WTP_{i,B}(\mathbf{J}) = \\ &\quad WTP_{i,T}(\mathbf{J,S}) - WTP_{i,T}(\mathbf{J}) \equiv MWTP_{i,T}(\mathbf{S}) \\ MWTP_{i,B}(\mathbf{C}) &\equiv WTP_{i,B}(\mathbf{J,S,C}) - WTP_{i,B}(\mathbf{J,S}) = \\ &\quad WTP_{i,T}(\mathbf{J,S,C}) - WTP_{i,T}(\mathbf{J,S}) \equiv MWTP_{i,T}(\mathbf{C}) \quad (3) \\ MWTP_{i,B}(\mathbf{S,C}) &\equiv WTP_{i,B}(\mathbf{J,S,C}) - WTP_{i,B}(\mathbf{J}) = \\ &\quad WTP_{i,T}(\mathbf{J,S,C}) - WTP_{i,T}(\mathbf{J}) \equiv MWTP_{i,T}(\mathbf{S,C}) \end{aligned}$$

where $i = S_I, S \text{ or } C$. Probability values of each of our test results are reported in Table 3.

From Table 3, simple difference in means tests do indeed find significant effects of order on WTP for actual bundles **J** and for **J,S** when valued for self (ρ values = .072, .061 respectively). These t tests do not find significant order effects for **J,S,C** when valued for self, nor for any size of bundle in the induced value or charity treatments. Reassuringly, they find that valuations are sensitive to scope in external tests for all three

Table 3
Significance of order and scope effects using *t*- , Median and Regression Analysis

Hypothesis	Difference in means* (ρ value)	Difference in medians (asymptotic ρ value)	Difference using regression analysis (ρ value)
<i>Tests of Order Effects</i>			
1. $WTP_{SIB}(J) = WTP_{SIT}(J)$	0.817	0.818	0.854
2. $WTP_{SIB}(J,S) = WTP_{SIT}(J,S)$	0.964	0.957	0.969
3. $WTP_{SIB}(J,S,C) = WTP_{SIT}(J,S,C)$	0.714	0.745	0.636
4. $WTP_{SB}(J) = WTP_{ST}(J)$	0.072	0.215	0.027
5. $WTP_{SB}(J,S) = WTP_{ST}(J,S)$	0.061	0.089	0.006
6. $WTP_{SB}(J,S,C) = WTP_{ST}(J,S,C)$	0.328	0.597	0.024
7. $WTP_{CB}(J) = WTP_{CT}(J)$	0.949	0.937	0.398
8. $WTP_{CB}(J,S) = WTP_{CT}(J,S)$	0.764	0.576	0.220
9. $WTP_{CB}(J,S,C) = WTP_{CT}(J,S,C)$	0.844	0.938	0.482
<i>External Tests of Scope Effects by First Response</i>			
10. $WTP_{SIB}(J) = WTP_{SIT}(J,S,C)$	0.000	0.000	0.000
11. $WTP_{SB}(J) = WTP_{ST}(J,S,C)$	0.000	0.000	0.000
12. $WTP_{CB}(J) = WTP_{CT}(J,S,C)$	0.000	0.000	0.000
<i>Consistency of Scope Effects by Order</i>			
13. $MWTP_{SIB}(S) = MWTP_{SIT}(S)$	0.789	0.892	0.793
14. $MWTP_{SIB}(C) = MWTP_{SIT}(C)$	0.561	0.978	0.547
15. $MWTP_{SIB}(S,C) = MWTP_{SIT}(S,C)$	0.810	0.892	0.811
16. $MWTP_{SB}(S) = MWTP_{ST}(S)$	0.290	0.340	0.181
17. $MWTP_{SB}(C) = MWTP_{ST}(C)$	0.730	0.656	0.939
18. $MWTP_{SB}(S,C) = MWTP_{ST}(S,C)$	0.766	0.974	0.432
19. $MWTP_{CB}(S) = MWTP_{CT}(S)$	0.538	0.700	0.627
20. $MWTP_{CB}(C) = MWTP_{CT}(C)$	0.452	0.283	0.374
21. $MWTP_{CB}(S,C) = MWTP_{CT}(S,C)$	0.820	0.841	0.820

All tests two tailed. P values in bold are significant at the 10% level. *t*- and Mann-Whitney medians tests run on SPSS 11.5. Multivariate analysis run on SPSS 7.0. For the details of the multivariate analysis, see Appendix 2. * Equality of variance not assumed if rejected by an F test at the 10% level.

treatments (hypotheses 10-12). Finally, t tests find that the order effects found in the valuation of bundles of private goods are not sufficient to affect the marginal valuation of **S**, **C**, or **S+C** in that or either of the other treatments (hypotheses 13-21). Moving to a comparison of medians, Mann-Whitney tests provide similar findings as the t tests, except that the median valuation for the bundle **J,S** when valued for self is no longer significantly different when bottom up (\$4.00) than when top down (\$3.50).

One drawback of both t- and median tests is that neither explicitly allow for the effect of respondents' characteristics on their valuations, or exploit the information that each individual provides three valuations. Thus, following Payne et al. (2000) and Bateman et al. (2004), we lastly test for order and scope effects using a random effects Tobit panel data model, which allows for left censoring of WTP at zero. As laid out in Appendix 2, we regress WTP on 5 of 6 possible bundle/order dummies and other covariates for each of the three treatments.⁹ The covariates were collected in a post experiment questionnaire, including age, year of university, gender, whether Economics courses had been completed, first language, and frequency of donation to charities. Only those covariates that were significant in a given treatment were retained. In particular, students for whom English was a second language bid significantly more for actual items for self in Treatment II, while students who had completed economics courses bid significantly less to donate items to charities in Treatment III. We then test for sensitivity to order or scope using Wald tests of equality between the coefficients on the appropriate bundle/order dummies. For example, the effect of order on WTP for actual good **J** for

⁹ The omitted dummy variables were Induced Self Bottom (**J,S**) in Treatment I, Self Bottom (**J,S**) in Treatment II, and Charity Bottom (**J,S**) in Treatment III. Other specifications were also tried, resulting in Wald test results with identical *p* values as those reported.

self compares the coefficients on the dummy variables “Self Bottom (**J**)” and “Self Top (**J**)” in the regression using data from Treatment II.

As seen in the final column of Table 3, the Wald test results are consistent with those of the univariate tests, with the evidence of order effects in Treatment II slightly strengthened. In particular, WTP for own purchase differs significantly by order of valuation for all three bundle levels (**J**, **J,S** and **J,S,C**), rather than just for two levels (**J** and **J,S**) using t tests, or at one level (**J,S**) using the Mann Whitney test. Otherwise, multivariate analysis corroborates the earlier findings that WTP is not sensitive to order effects for induced value goods or goods donated to charities, but that it is sensitive to scope for all three treatments.

Finally, to test for an effect of order on sensitivity to scope, we follow Bateman et al. (2004) in allowing that the marginal WTP (MWTP) across bundles can be negative or positive, and so use a random effects generalized least squares (GLS) panel regression. Because the MWTP(**C**) and MWTP(**S**) are not independent of the MWTP(**S,C**) only observations for the first two were used. Again as laid out in Appendix 2, we regress MWTP on 3 of 4 possible dummies for differences in bundle and order and on other covariates for each of the three treatments.¹⁰ Again, only those covariates that were significant for a given treatment were retained. It should be noted that in general the fit of the GLS regression is not as good as that for the Tobit analysis. We tested for the effect of order on the degree of scope sensitivity using Wald tests of equality of coefficients on the appropriate dummies. Again as seen in the final column of Table 3, multivariate analysis concurs with the univariate tests in finding that the order effects

¹⁰ The omitted dummy variable was Induced Self Bottom (**J,S**) – Induced Self Bottom (**J**) for Treatment I, Self Bottom (**J,S**) – Self Bottom (**J**) for Treatment II, and Charity Bottom (**J,S**) – Charity Bottom (**J**) for Treatment III.

found in the valuation of bundles of private goods are not sufficient to affect the degree of sensitivity to scope in Treatment II, or in either of the other two treatments.

4. Is Confusion the Cause?

Whether by univariate or multivariate tests, we have found evidence that the order in which nested private goods are presented can affect the valuation that experimental subjects assign to them. In particular, a given bundle of goods seems to suffer by relative comparison with greater bundles that have already been viewed. These order effects have emerged despite the facts that 1) the goods being valued are familiar market goods for actual purchase, 2) income and substitution effects should play no role, 3) subjects appear to understand that they have incentives to reveal their true willingness to pay, and 4) warm glow considerations should play no role. Indeed, when warm glow considerations are introduced as a predominant consideration in valuation, the order effects are eliminated. Subjects seem more "rational" in assigning maximum amounts to pay to donate goods to various charities than they are in valuing goods for their own use.

Could the order effects found here (as in Contingent Valuation) be attributed to a confused fringe of respondents who do not understand the valuation question being asked of them, or who do not care about the goods in question? We address these possibilities by constructing four filters by which confused or indifferent subjects could be identified and eliminated *ex post* from analysis. These are:

Filter I : exclude all subjects whose reported maximum purchase price in the final practice round differed by more than \$.20 from their induced value. Excludes 34 of 235 subjects.

- Filter II: exclude all subjects whose marginal WTP for a bundle of greater scope was negative in any of the three main decision rounds. Excludes 11 of 235 subjects.
- Filter III: exclude all subjects whose highest reported WTP in the three main rounds ensured a zero chance of purchasing any of the bundles. Excludes 5 of 235 subjects.
- Filter IV: exclude all subjects whose highest reported WTP in the three main rounds ensured a small (2.4% or less) chance of purchasing any of the bundles. Excludes 13 of 235 subjects.

In brief, we find that none of the four confusion or indifference filters eliminates the order effects in Treatment II, though the evidence is weakened slightly. Regression analysis analogous to that described above continues to find order effects for all three bundle levels (**J**, **J,S** and **J,S,C**) under all four filters. However t- tests now find evidence of order effects for **J** only under Filter I, with marginal evidence under Filters III and IV, and for **J,S** under Filters I, III and IV. Mann Whitney tests now find evidence of order effects for **J,S** only under Filter I. Appendix 3 provides greater detail.

5. Discussion and Conclusion

Carefully structured CV studies have found that the valuation of multiple nested environmental proposals can be influenced by the order of presentation, even when economic theory clearly predicts that order should not matter. In particular, goods of a relatively small scope receive a higher valuation if considered prior to goods of a greater

scope, than if they are considered after (Powe and Bateman 2003, Bateman et al. 2004). Such order effects have occurred even when the multiple proposals are valued as mutually exclusive alternatives to the status quo, so that income and substitution effects and reference utility are controlled. Researchers have speculated that order effects could be caused by the hypothetical nature of CV, with its unclear strategic incentives, inchoate preferences over unfamiliar public goods, and potential for free purchase of moral satisfaction unrelated to the specific good being valued.

We have attempted to explore these potential explanations for order effects in CV using an incentive compatible valuation experiment over mutually exclusive bundles of private market goods, with and without a strong potential for warm glow or “purchase of moral satisfaction.” We find two striking things. First, there is significant evidence that the order effect whereby smaller goods suffer when compared to larger goods may be a robust preference phenomenon unrelated to CV *per se*. For we find evidence of the same effect when subjects report their non-hypothetical maximum WTP for private market goods under incentive compatible conditions that they *understand* to be incentive compatible. Mean WTP for a ten pack of *Minute Maid* orange juice was \$2.40 when valued first, but \$1.67 when valued after a more comprehensive bundle -- a 43% difference. Similarly, mean marginal WTP for the juice combined with *Staples* gel ink pens and *BiC* correction tape was \$4.58 when valued after a smaller bundle, but \$3.48 when valued after a larger bundle -- a 32% difference. Such valuations were provided in a context where moral satisfaction arguably plays no role.

Second, we do not find evidence of order effects when subjects provide their maximum WTP to donate the same bundles of goods to charities. Sugden (1997) and

Nunes and Schokkaert (2003) have proposed that the WTP responses given in CV may be divided between “instrumental” and “expressive” values, or alternatively between “use,” “existence” and “warm glow” components. Contrary to the hypothesis of Kahneman and Knetsch (1992), we find that a higher relative importance of “warm glow” in overall WTP eliminates rather than strengthens order effects.

One could argue that lab experiments shares key characteristics with CV that are responsible for the order effects observed in both. For both exercises ask people to do something unfamiliar in a relatively short space of time (here 90 minutes) -- provide a maximum willingness to pay for a good. Exactly identifying own maximum WTP is not something people must commonly do when purchasing items in posted offer markets. However, we find that the order effects attenuate but do not disappear when we exclude subject responses using four criteria of confusion or indifference to the goods being valued. In addition, the limited time and experience that subjects had in providing valuations did not result in order effects when they provided the maximum they would pay to donate the same goods to 3 charities. We speculate that this could be because subjects have more experience identifying the maximum amount they will donate to charities (though not to donate specific goods to charities) than they have in identifying the maximum amount they will pay to purchase items for their own use. Future research could investigate whether order effects in CV could be eliminated by increasing the experience respondents have in providing valuations.

References

- I.J. Bateman, M. Cole, P. Cooper, S. Georgiou, D. Hadley, G. Poe, On visible choice sets and scope sensitivity, *Journal of Environmental Economics and Management*. 47 (2004) 71-93.
- I.J. Bateman, I.H. Langford, Budget-constraint, temporal, and question-ordering effects in contingent valuation studies, *Environment and Planning A*. 28 (1996) 1215-1228.
- I.J. Bateman, A. Munro, B. Rhodes, C. Starmer, R. Sugden, Does part-whole bias exist? An experimental investigation, *The Economic Journal*. 107, 441 (1997a) 322-332
- I.J. Bateman, A. Munro, B. Rhodes, C. Starmer and R. Sugden, A test of the theory of reference dependent preferences, *The Quarterly Journal of Economics*. 112, 2 (1997b) 479-505.
- G. Becker, M. DeGroot, J. Marschak, Measuring utility by a single-response sequential method, *Behavioural Science*. 9 (1964) 226-232.
- P. Bohm, J. Linden and J. Sonnegard, Eliciting reservation prices: Becker-DeGroot-Marschak mechanisms vs. markets, *The Economic Journal*. 107, 443, 1079-1089.
- K.J. Boyle, M.P. Welsh, R.C. Bishop, The role of question order and respondent experience in contingent valuation studies, *Journal of Environmental Economics and Management*. 25 (1993) S80 – S99.
- D.S. Brookshire, R.C. d'Arge, W.D. Schulze, M.A. Thayer, Experiments in valuing public goods, in V.K. Smith (Ed.) *Advances in Applied Microeconomics: Volume 1* (1981) Greenwich CT: JAI Press.
- R.T. Carson, N.E. Flores, W.M. Hanemann, Sequencing and valuing public goods, *Journal of Environmental Economics and Management*. 36 (1998) 314-323.
- R.T. Carson, Contingent valuation surveys and tests of insensitivity to scope, in R.J. Kopp, W. Pommerhene, N. Schwartz (Eds.) *Determining the Value of Non-Marketed Goods: Economic, Psychological and Policy Relevant Aspects of Contingent Valuation Methods*. (1997) Boston: Kluwer, Chapter 6.
- R.T. Carson, R.C. Mitchell, Sequencing and nesting in contingent valuation surveys, *Journal of Environmental Economics and Management*. 28 (1995) 155-173.
- D.L. Coursey, J.L. Hovis, W.D. Schulze, On the supposed disparity between willingness to accept and willingness to pay measures of value, *Quarterly Journal of Economics*. 102, (1987), 679-690.

D.D. Davis, C.A. Holt, *Experimental Economics*. (1993) Princeton New Jersey: Princeton University Press.

W.H. Desvousges, F.R. Johnson, R. W. Dunford, K.J. Boyle, S.P. Hudson and K.N. Wilson, *Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy*. (1992) Research Triangle Park North Carolina: Research Triangle Institute.

W.H. Desvousges, F.R. Johnson, R. W. Dunford, K.J. Boyle, S.P. Hudson and K.N. Wilson, *Measuring natural resource damages with contingent valuation: tests of validity and reliability*, in J.A. Hausman (Ed.) *Contingent Valuation: A Critical Assessment*. (1993) Amsterdam: North Holland Publishers.

P.A. Diamond, J.A. Hausman, G.K. Leonard, M.A. Denning, *Does contingent valuation measure preferences? Experimental Evidence*, in J. A. Hausman (Ed.) *Contingent Valuation: A Critical Assessment*. (1993) Amsterdam: North Holland Publishers.

C.C. Eckel, P.J. Grossman, *Rebate versus matching: does how we subsidize charitable contributions matter?*, *Journal of Public Economics*. 87 (2003), 681-701.

T. A. Heberlein, M.A. Wilson, R.C. Bishop and N.C. Schaeffer, *Rethinking the scope test as a criterion for validity in contingent valuation*, *Journal of Environmental Economics and Management*. 50 (2005) 1-22.

D. Kahneman, J.L. Knetsch, *Valuing public goods: the purchase of moral satisfaction*, *Journal of Environmental Economics and Management*. 22 (1992) 57-70.

D. Kahneman, P. Slovic, A. Tversky (Eds.), *Judgement Under Uncertainty: Heuristics and Biases*. (1982) New York: Cambridge University Press.

C. Noussair, S. Robin, B. Ruffieux, *Do consumers really refuse to buy genetically modified food?* *The Economic Journal*. 114, 492 (2004) 102-120.

P.A. Nunes, E. Schokkaert, *Identifying the warm glow effect in contingent valuation*, *Journal of Environmental Economics and Management*. 45 (2003) 231-245.

J.W. Payne, D.A. Schkade, W.H. Desvousges, C. Aultman, *Valuation of multiple environmental programs*, *Journal of Risk and Uncertainty*. 21, 1, (2000) 95-115.

N.A. Powe, I.J. Bateman, *Ordering effects in nested 'top-down' and 'bottom-up' contingent valuation designs*, *Ecological Economics*. 45 (2003) 255-270.

A. Randall, J.P. Hoehn, *Embedding in market demand systems*, *Journal of Environmental Economics and Management*. 30 (1996) 369-380.

A. Randall, J.P. Hoehn, G. Tolley, "The Structure of Contingent Markets: Some Results of a Recent Experiment." Presented to the annual meeting of the American Economics Association, Washington, D.C., (1981).

J.F. Shogren, S.Y. Shin, D.J. Hayes, J.B. Kliebenstein, Resolving differences in willingness to pay and willingness to accept, *The American Economic Review*. 84, 1, (1994) 255-270.

R. Sugden, Public goods and contingent valuation, in R.J. Kopp, W. Pommerhene, N. Schwartz (Eds.) *Determining the Value of Non-Marketed Goods: Economic, Psychological and Policy Relevant Aspects of Contingent Valuation Methods*. (1997) Boston: Kluwer, Chapter 5.

A. Tversky, D. Kahneman, Loss aversion in riskless choices: a reference-dependent model, *The Quarterly Journal of Economics*. 106, 4, (1991) 1039-1061.

W. Vickrey, Counterspeculation, auctions, and competitive sealed tenders, *Journal of Finance*. 16 (1961) 8-37.

T.H. Wonnacott, R.J. Wonnacott, *Introductory Statistics for Business and Economics*, 3rd Edition, (1984) New York: Wiley.

Appendix 1: The distribution of discrepancies between maximum purchase price reported and induced value for the final practice round (all subjects).

Appendix 2: Multivariate Analysis of Order and Scope Effects

Treatment I: Induced Value Private Goods

	(1) Random Effects Tobit Panel Regression	(2) Random Effects GLS Panel Regression	
	WTP		Marginal WTP
Constant	5.881*** (.539)	Constant	2.695*** (.458)
Induced Self Bottom (J)	-2.695*** (.516)	Induced Self Top (S)	.171 (.648)
Induced Self Top (J)	-2.835*** (.761)	Induced Self Bottom (C)	2.098*** (.648)
Induced Self Top (J,S)	0.030 (.761)	Induced Self Top (C)	1.708*** (.648)
Induced Self Bottom (J,S,C)	4.793*** (.516)	Induced Self Bottom (S,C)	dropped
Induced Self Top (J,S,C)	4.433*** (.761)	Induced Self Top (S,C)	dropped
σ_e	1.633	σ_e	2.067
σ_u	1.690	σ_u	0
N ^a	120	N ^a	80
uncensored	120		
left censored	0		
Log Likelihood	-258.0		
Wald statistic	416.82	Wald statistic	16.15
Prob > χ^2	.000	Prob > χ^2	.001

Standard errors in parentheses. ***, **, * refer to significance at the 1, 5 and 10% levels, respectively. Run on Stata 7.0. ^a For purposes of comparing valuations, the valuations provided by 4 subjects in a top down session with induced values set according to those of four missing subjects in a bottom up session are excluded.

Appendix 2: (Cont'd)

Treatment II: Actual Private Goods

	(3) Random Effects Tobit Panel Regression WTP	(4) Random Effects GLS Panel Regression Marginal WTP
Constant	6.471*** (.635)	Constant 3.262*** (.399)
Self Bottom (J)	-2.261*** (.380)	Self Top (S) -.559 (.417)
Self Top (J)	-3.600*** (.603)	Self Bottom (C) .203 (.354)
Self Top (J,S)	-1.660*** (.601)	Self Top (C) .235 (.417)
Self Bottom (J,S,C)	2.380*** (.378)	Self Bottom (S,C) dropped
Self Top (J,S,C)	1.022* (.599)	Self Top (S,C) dropped
English First Language	-2.623*** (.639)	English First Language -1.441*** (.360)
σ_e	1.832	σ_e 1.716
σ_u	2.704	σ_u 1.071
N ^a	285	N ^a 190
uncensored	272	
left censored	13	
Log Likelihood	-656.03	
Wald statistic	323.06	Wald statistic 21.54
Prob > χ^2	.000	Prob > χ^2 .0002

Standard errors in parentheses. ***, **, * refer to significance at the 1, 5 and 10% levels, respectively. Run on Stata 7.0. ^a One subject provided no demographic data, reducing N from 288 in (3) and from 192 in (2).

Appendix 2: (Cont'd)

Treatment III: Actual Goods for Donation to Charities

	(5) Random Effects Tobit Panel Regression WTP	(6) Random Effects GLS Panel Regression Marginal WTP	
Constant	5.565*** (.761)	Constant	1.665*** (.302)
Charity Bottom (J)	-1.743*** (.392)	Charity Top (S)	-.207 (.425)
Charity Top (J)	-2.422*** (.800)	Charity Bottom (C)	-.297 (.284)
Charity Top (J,S)	-.981 (.800)	Charity Top (C)	.675 (.425)
Charity Bottom (J,S,C)	1.984*** (.389)	Charity Bottom (S,C)	dropped
Charity Top (J,S,C)	1.423* (.798)	Charity Top (S,C)	dropped
Taken Economics	-2.033*** (.584)		
σ_e	1.876	σ_e	1.374
σ_u	2.822	σ_u	1.551
N	285	N	190
uncensored	277		
left censored	8		
Log Likelihood	-670.2		
Wald statistic	204.66	Wald statistic	11.02
Prob > χ^2	.0000	Prob > χ^2	.0116

Standard errors in parentheses. ***, **, * refer to significance at the 1, 5 and 10% levels, respectively. Run on Stata 7.0

Appendix 2: Cont'd

Multivariate Wald Tests Reported in Table 3:

1. $WTP_{SIB}(J) = WTP_{SIT}(J)$

A Wald test of the equality of the coefficients on the Induced Self Bottom (J) and Induced Self Top (J) dummies in regression (1). $\chi^2_1 = .03$, ρ value = .8536.

2. $WTP_{SIB}(J,S) = WTP_{SIT}(J,S)$

A Wald test of the equality of the coefficients on the Induced Self Bottom (J,S) and Induced Self Top (J,S). Because Induced Self Bottom (J,S) is the omitted dummy variable, this becomes a test of difference from the base intercept, or if the coefficient on Induced Self Top (J,S) equals zero in regression (1). $\chi^2_1 = .00$, ρ value = .9685.

3. $WTP_{SIB}(J,S,C) = WTP_{SIT}(J,S,C)$

A Wald test of the equality of the coefficients on the Induced Self Bottom (J,S,C) and Induced Self Top (J,S,C) dummies in regression (1). $\chi^2_1 = .22$, ρ value = .6362.

4. $WTP_{SB}(J) = WTP_{ST}(J)$

A Wald test of the equality of the coefficients on the Self Bottom (J) and Self Top (J) dummies in regression (3). $\chi^2_1 = 4.89$, ρ value = .0270.

5. $WTP_{SB}(J,S) = WTP_{ST}(J,S)$

A Wald test of the equality of the coefficients on the Self Bottom (J,S) and Self Top (J,S). Because Self Bottom (J,S) is the omitted dummy variable, this becomes a test of difference from the base intercept, or if the coefficient on Self Top (J,S) equals zero in regression (3). $\chi^2_1 = 7.64$, ρ value = .0057.

6. $WTP_{SB}(J,S,C) = WTP_{ST}(J,S,C)$

A Wald test of the equality of the coefficients on the Self Bottom (J,S,C) and Self Top (J,S,C) dummies in regression (3). $\chi^2_1 = 5.13$, ρ value = .0236.

7. $WTP_{CB}(J) = WTP_{CT}(J)$

A Wald test of the equality of the coefficients on the Charity Bottom (J) and Charity Top (J) dummies in regression (5). $\chi^2_1 = .71$, ρ value = .3978.

8. $WTP_{CB}(J,S) = WTP_{CT}(J,S)$

A Wald test of the equality of the coefficients on the Charity Bottom (J,S) and Charity Top (J,S). Because Charity Bottom (J,S) is the omitted dummy variable, this becomes a test of difference from the base intercept, or if the coefficient on Charity Top (J,S) equals zero in regression (5). $\chi^2_1 = 1.50$, ρ value = .2202.

9. $WTP_{CB}(J,S,C) = WTP_{CT}(J,S,C)$

A Wald test of the equality of the coefficients on the Charity Bottom (J,S,C) and Charity Top (J,S,C) dummies in regression (5). $\chi^2_1 = .49$, ρ value = .4824.

10. $WTP_{SIB}(J) = WTP_{SIT}(J,S,C)$

A Wald test of the equality of the coefficients on the Induced Self Bottom (J) and Induced Self Top (J,S,C) dummies in regression (1). $\chi^2_1 = 87.72$, ρ value = .0000.

11. $WTP_{SB}(J) = WTP_{ST}(J,S,C)$

A Wald test of the equality of the coefficients on the Self Bottom (J) and Self Top (J,S,C) dummies in regression (3). $\chi^2_1 = 29.80$, ρ value = .0000.

12. $WTP_{CB}(J) = WTP_{CT}(J,S,C)$

A Wald test of the equality of the coefficients on the Charity Bottom (J) and Charity Top (J,S,C) dummies in regression (5). $\chi^2_1 = 15.66$, ρ value = .0001.

13. $MWTP_{SIB}(S) = MWTP_{SIT}(S)$

A Wald test of the equality of the coefficients on the Induced Self Bottom (S) and Induced Self Top (S) dummies in regression (2). Because the former dummy is the omitted case, this becomes a test of difference from the base intercept, or if the coefficient on Induced Self Top (S) equals zero. $\chi^2_1 = .07$, ρ value = .7925.

14. $MWTP_{SIB}(C) = MWTP_{SIT}(C)$

A Wald test of the equality of the coefficients on the Induced Self Bottom (C) and Induced Self Top (C) dummies in regression (2). $\chi^2_1 = .36$, ρ value = .5474.

15. $MWTP_{SIB}(S,C) = MWTP_{SIT}(S,C)$

This test can be reformulated as a test of whether

$$MWTP_{SIB}(C) + MWTP_{SIB}(S) = MWTP_{SIT}(C) + MWTP_{SIT}(S).$$

Given that the second term of the left hand side is the omitted dummy variable in (2), this reduces to a test of whether

$$MWTP_{SB}(C) = MWTP_{ST}(C) + MWTP_{ST}(S),$$

or whether the coefficient on Induced Self Bottom (C) equals the sum of those on Induced Self Top (C) and Induced Self Top (S). $\chi^2_1 = .06, \rho \text{ value} = .8108$.

$$16. MWTP_{SB}(S) = MWTP_{ST}(S)$$

A Wald test of the equality of the coefficients on the Self Bottom (S) and Self Top (S) dummies in regression (4). Because the former dummy is the omitted case, this becomes a test of difference from the base intercept, or if the coefficient on Self Top (S) equals zero. $\chi^2_1 = 1.79, \rho \text{ value} = .1807$.

$$17. MWTP_{SB}(C) = MWTP_{ST}(C)$$

A Wald test of the equality of the coefficients on the Self Bottom (C) and Self Top (C) dummies in regression (4). $\chi^2_1 = .01, \rho \text{ value} = .9391$.

$$18. MWTP_{SB}(S,C) = MWTP_{ST}(S,C)$$

This test can be reformulated as a test of whether

$$MWTP_{SB}(C) + MWTP_{SB}(S) = MWTP_{ST}(C) + MWTP_{ST}(S)$$

Given that the second term of the left hand side is the omitted dummy variable in (4), this reduces to a test of whether

$$MWTP_{SB}(C) = MWTP_{ST}(C) + MWTP_{ST}(S)$$

or whether the coefficient on Self Bottom (C) equals the sum of those on Self Top (C) and Self Top (S). $\chi^2_1 = .62, \rho \text{ value} = .4316$.

$$19. MWTP_{CB}(S) = MWTP_{CT}(S)$$

A Wald test of the equality of the coefficients on the Charity Bottom (S) and Charity Top (S) dummies in regression (6). Because the former dummy is the omitted case, this becomes a test of difference from the base intercept, or if the coefficient on Charity Top (S) equals zero. $\chi^2_1 = .24, \rho \text{ value} = .6272$.

20. $MWTP_{CB}(C) = MWTP_{CT}(C)$

A Wald test of the equality of the coefficients on the Charity Bottom (C) and Charity Top (C) dummies in regression (6). $\chi^2_1 = .79, \rho \text{ value} = .3743$.

21. $MWTP_{CB}(S,C) = MWTP_{CT}(S,C)$

This test can be reformulated as a test of whether

$$MWTP_{CB}(C) + MWTP_{CB}(S) = MWTP_{CT}(C) + MWTP_{CT}(S)$$

Given that the second term of the left hand side is the omitted dummy variable in (6), this reduces to a test of whether

$$MWTP_{CB}(C) = MWTP_{CT}(C) + MWTP_{CT}(S)$$

or whether the coefficient on Charity Bottom (C) equals the sum of those on Charity Top (C) and Charity Top (S). $\chi^2_1 = .05, \rho \text{ value} = .8196$.

Appendix 3: Summary of Order and Scope Effect Test p Values Using Filters

- Filter I: Confusion on BDM as demonstrated on last practice round.
Excludes 7 subjects from Treatment I, 13 from II, and 14 from III.
- Filter II: Confusion or changed mind as demonstrated by negative MWTP.
Excludes 2 subjects from Treatment I, 4 from Treatment II, and 5 from III.
- Filter III: Indifference to goods as demonstrated by valuations yielding 0 chance of purchasing any items. Excludes 4 subjects from Treatment II, 1 from III.
- Filer IV: Indifference to goods as demonstrated by valuations yielding $\leq 2.4\%$ chance of purchasing any items. Excludes 4 subjects from Treatment II, and 9 from III.

Hypothesis (See Table 3)	Test	Filter I			Filter II			Filter III			Filter IV		
		t	MW	W	t	MW	W	t	MW	W	t	MW	W
		p values	p values	p values	p values	p values	p values	p values	p values	p values	p values	p values	p values
1.		.	.	.	.	.	.	.	.	.	.	.	.
2.		.	.	.	.	.	.	.	.	.	.	.	.
3.		.	.	.	.	.	.	.	.	.	.	.	.
4.		.086	.190	.053	.198	.409	.083	.102	.344	.077	.102	.344	.077
5.		.055	.062	.007	.119	.169	.011	.099	.154	.020	.099	.154	.020
6.		.143	.283	.003	.349	.541	.021	.477	.851	.076	.477	.851	.076
7.		.	.	.	.	.	.	.	.	.	.	.	.
8.		.	.	.042	.	.	.	.	.	.	.	.	.
9.		.	.	.078	.	.	.	.	.	.	.	.	.
10.		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
11.		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
12.		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
13.		.	.	.	.	.	.	.	.	.	.	.	.
14.		.	.	.	.	.	.	.	.	.	.	.	.
15.		.	.	.	.	.	.	.	.	.	.	.	.
16.		.	.	.	.	.	.081	.	.	.	.	.	.
17.		.	.	.	.	.	.	.	.	.	.	.	.
18.		.	.	.	.	.	.	.	.	.	.	.	.
19.		.081	.	.	.	.	.	.	.	.	.	.	.
20.		.	.	.	.	.	.	.	.	.	.	.	.
21.		.	.	.	.	.	.	.	.	.	.	.	.

All tests two tailed. With the exception of Treatment II test results, only p values that are significant at the 10% level are included, and are in bold. t and Mann-Whitney (MW) tests run on SPSS 11.5. Wald (W) tests and regressions run on STATA 7.0.