Why Value Estimates Generated Using Choice Modelling Exceed Contingent Valuation: Further Experimental Evidence[1]

Kevin Boyle, Mark Morrison and Laura Taylor[2]

Abstract: Choice modelling is increasingly being used to generate estimates of the value of changes in environmental quality. This is partly because of the informational efficiencies of the technique, but also because of concern about the accuracy of contingent valuation. Experimental evidence has, however, demonstrated that choice modelling tends to produce much higher valuation estimates than contingent valuation. One possible explanation for the divergence between choice modelling and contingent valuation estimates is a lack of incentive compatibility with the former technique. This potentially has several sources, including having no provision rule (eg a referendum), respondents choosing between more than two alternatives, and repeated choices. We report on the results of a series of experiments involving over 2000 subjects designed to test whether a lack of incentive compatibility is responsible for divergences in value estimates.

1. Introduction

In recent years there has been a movement away from using contingent valuation to estimate non-use values towards using various forms of conjoint analysis. Conjoint analysis is becoming the technique of choice in major government sponsored valuation exercises both in the US and abroad (eg Breffle et al. 1990, Boyle et al. 2001, Bennett and Morrison 2001). This movement has occurred for several reasons. Conjoint analysis has the advantage of providing greater information about values. Specifically, it provides information about the value of attributes, so it can be used to value multiple project outcomes. However, its greater usage also reflects concerns about possible biases associated with contingent valuation that are *assumed* to be less prevalent in conjoint analysis (Hanley et al. 1998).

Several different conjoint approaches have been used in applications where the goal has been the estimation of non-market values, including choice modeling, contingent ratings and contingent rankings. The technique with the greatest number of applications in environmental, transportation and health economics is choice modeling, where respondents choose their preferred alternative rather than use rating scales or ranks to indicate their preferences.

Unlike conjoint analysis, contingent valuation has been subject to rigorous tests of predictive validity using experimental methods. For standard applications of contingent valuation, the evidence of convergence between real and hypothetical applications is mixed. Prince et al (1992) used a contribution game and find support for incentive compatibility using induced value

---

1 Paper presented at the Australian Agricultural and Resource Economics Society Conference, Melbourne, 11-13 February, 2004.
[2]The authors are, respectively: Libra Professor of Environmental Policy, Department of Agricultural and Resource Economics and Policy, University of Maine; Senior Lecturer, Charles Sturt University, Bathurst, Australia (mmorrison@csu.edu.au); and Associate Professor of Economics, Andrew Young School of Policy Studies, Georgia State University. Senior authorship is not assigned.

experiments. However, Cummings et al (1995) compared real and hypothetical choices for three private goods using the dichotomous choice CV format and student subjects – results not supportive of incentive compatibility. In another study, this time surveying community subjects rather than students, Cummings et al (1997) compared real and hypothetical choices for a single public good (citizens' groundwater quality guide). They found significant differences between real and hypothetical choices. However Haab, Huang and Whitehead (1999) re-analyzed the data from Cummings et al (1997) and found that when scale differences included, there is convergent validity between real and hypothetical choices. Finally, Taylor (1998) investigated incentive compatibility using a closed CV referendum, comparing real and hypothetical choices. This experiment involved a public good (the same good as was used in Cummings et al 97) and student subjects were again used. The hypothesis of predictive validity was rejected in this study.

For contingent valuation, there has however been more consistent success in demonstrating predictive validity using several different "demand revealing" designs. While these designs are not strictly incentive compatible, they have been affective at achieving predictive validity. The better known of these techniques is cheap talk. Cheap talk involves giving respondents an explanation about the problems associated with hypothetical bias prior to asking them about their willingness to pay. Cummings and Taylor (1999) found evidence that using a cheap talk design was demand revealing. Further studies by Lusk (2003) and Murphy, Stevens and Weatherhead (2003) have found this design to reduce overstatement of willingness to pay, but not eliminate it. Other demand revealing designs have also been identified. One example is the learning design suggested by Bjornstad, Cummings and Taylor (1997). With this design, respondents are initially given a hypothetical survey to answer, then a real survey, and then finally another hypothetical survey. The empirical evidence from this study demonstrated convergence between the results of the real survey and the second hypothetical survey. The implication is that sequencing of surveys can induce respondents to answer truthfully. Thus, it appears that contingent valuation designs exist that potentially can be used to generate accurate valuation estimates.

Less research of this kind has been undertaken to demonstrate the validity of conjoint analysis applications. In the only published study currently available, Carlsson and Martinnson (2001) compared hypothetical and actual choices using choice modeling (choice experiments). They find evidence of predictive validity, but in a specific context. They valued several different public goods using student subjects, they did not use a split sample test (the same respondents answered the real and hypothetical scenarios), it was not a multi-attribute experiment, and binary-repeated choices were used. This is different to a lot of standard choice modeling applications where there are more than two alternatives in a choice, goods have multiple attributes, and one good is valued in an experiment. Moreover, it is more common to use a split sample test in this type of testing, which arguably provides for a more robust test. A second study currently still in the grey literature was conducted by Whinstanley Newell and Swallow (2002), who compared real and hypothetical choices in a wetland valuation study. They found differences in preference functions, particularly amongst the non-monetary attributes. Some evidence of convergence of the marginal utility of income was identified; however equality of estimates of implicit prices or surplus estimates was not tested.

While few studies have tested the predictive validity of conjoint analysis, there have been a number of attempts to test the convergent validity of contingent valuation and conjoint analysis applications. Examples of these from both the published and grey literature are shown in Table 1. Of the ten studies reported, statistical tests were reported in eight of the studies and significant differences were identified in four of these comparisons. In each of these four comparisons, the value estimates from the conjoint approaches were larger, except for Foster and Mourato (2003) where for the less inclusive public good the converse is true. It is possible that in several of the tests where no significant difference was identified, that the test results were confounded by an inability to accurate derive value estimates. The confidence intervals in several papers were reported to be very large. Across the ten papers, the conjoint estimates are on average 193% larger, while when considering only the published estimates the conjoint estimates are on average 179% larger.

These results lead to the question of what lies behind this divergence between contingent valuation and conjoint estimates. A feature of conjoint analysis, where the objective is the estimation of non-market values, is that most applications have followed closely the methodology used in the marketing paradigm, where conjoint techniques were originally developed. A potential risk in adopting a technique developed in a different literature is that theoretical concerns that are fundamental to economics can be overlooked. An important part of the contingent valuation literature was the development of an incentive compatible provision rule that is made explicit to survey respondents. The most commonly used provision rule in these regards is the referendum voting mechanism (Hoehn and Randall, 1987; Arrow et al. 1993). With conjoint analysis applications, however, respondents are simply asked to reveal their preferences through various evaluation tasks. We are aware of no published conjoint application that explicitly or implicitly state a provision rule within the survey questionnaire. It is possible that the absence of provision rules provides at least part of the explanation for the divergence between contingent valuation and conjoint estimates.

We report the results of a series of experiments designed to identify the impact of including provision rules in conjoint surveys involving both public and private goods, and both hypothetical and real payment. We compare the results of choice modeling surveys which include: (1) an individual provision rule (IPR) that is incentive compatible, (2) a group or public provision rule (GPR) that is not incentive compatible, and (3) no provision rule (NPR). We also compare these treatments to a contingent valuation treatment involving hypothetical payment.

Preliminary results indicate that the inclusion or exclusion of provision rules has a substantial, significant impact on the choices made in conjoint surveys. We find that the exclusion of a provision rule causes marginal estimates to be significantly and substantially larger. Moreover, the inclusion of a provision rule in a choice experiment (IPR) leads to convergence with the estimates generated in the contingent valuation treatment. The inclusion of a provision rule was found to cause marginal value estimates to be much closer to the estimates generated using real payment, but did not totally eliminate hypothetical bias.

Table 1: Studies Comparing Contingent Valuation and Choice Modelling Estimates

Study	Value	Format of CV and CJ questions	Same bid vector	Estimates	Statistical tests
Published studies					
Smith and Desvouges 1986	Use	CV – OE, payment card, iterative bidding Contingent ranking	No	CV $15.95-36.88 CR $60.03-62.12	Inclusion of values in a regression equation (NS)[1]
Magat et al 1988	Use	CV – open ended Paired Comparison	No	CV/CR *Bleach:* Gassings $0.15/0.69 Child poisoning $0.21/0.25 *Drain opener* Hand burns $0.06/0.62 Child poisoning $0.18/0.41	Inclusion of values in a regression equation (S)[2]
Boxall et al 1996	Use	CV – referendum CM – choose one from	Yes	CV $85.59 CM $56.69	Not tested
Adamowicz et al 1998	Non-use	CV – referendum CM – choose one from three	Yes	CV $140.86 CM $217.83 (quadratic model results only for both)	LR tests, overlapping sd's (NS)
Hanley et al 1998	Use	CV – open ended (OE) dichotomous choice (DC) CM – choose one from three		CV (OE) £31.43 CV (DC) £98 CM £107.55	Overlapping CI's (NS)
Stevens et al 2000	Non-use	CV – referendum CJ – recoded as chosen if rated more highly than status quo	Yes	CV $86 CJ $116	Values were significantly different, but method of testing not reported (S)
Foster and Mourato 2003	Non-use	CV – double bounded DC CM – choose one from three	Yes	All charities CV $43.4-55.9 CJ $258.4 Housing charities only CV $31.6-46.7 CJ $2.9	Differences between two distributions (Poe et al 1997) (S)
Grey literature					
Lockwood and Carberry 1998	Non-use	CV – dichotomous choice CM – choose one from three	Yes	*NSW* CV $80.69-86.79 CM $37.75-$51.97 *VIC* CV $77.35-98.40 CM $43.15-$93.63	Overlapping CI's (NS)
Christie and Azevedo 2002	Non-use	CV – referendum CM – choose one from three	No	*PlanA:* CV -$658 CM -$2122 *Plan B:* CV $540 CM $616 *Plan C:* CV $821 CM $2921	t-tests (S)
Takatsuka et al 2002	Use and Non-use	CV1 – referendum CV2- modified referendum (explicit attributes like CM) CM – choose one from three	Yes	*Scenario 1* CV1 $17.14 CM $82.59 *Scenario 2* CV2 $27.41 CM $222.10	Not tested

1 – not significantly different, 2 – significantly different

2. Hypotheses and Experimental Design

Hypotheses and Provision Rule Design

We design surveys to test the hypothesis that there are no significant differences between the choices made in the following four conjoint survey treatments and a contingent valuation treatment:

- an individual provision rule that is incentive compatible with hypothetical payment (IPR),
- a public provision rule that is not incentive compatible with hypothetical payment (GPR),
- no provision rule with hypothetical payment (NPR).
- an individual provision rule that is incentive compatible with real payment (RPR)
- a contingent valuation treatment that is incentive compatible with hypothetical payment (CVM)

This hypothesis is tested using both a private good and a public good as the object of choice. While tests with private-goods are not the main focus of the paper, they are conducted because, as discussed below, the private-good IPR treatment is the only treatment that is known to be fully incentive-compatible (i.e., the subject's optimal strategy is to truthfully reveal their preferences).

All surveys follow the same format, only the discussion of the provision rules changes across treatment. All subjects are first given an introduction to the survey, then are made familiar with the goods they will be making decisions over, and finally shown a sample of the choice questions they will be answering (more detail on this is provided below). At this point, subjects either hear a discussion of the provision rule, or not (in the NPR treatment), and then there is a final review of the survey.

The provision rules, and their incentive properties, are most easily understood by first describing the IPR treatment when the private good is the object of choice. In the IPR treatment, subjects are given seven choice questions, each question asking them to choose one of two variants of a good (with varying costs), or to "opt-out" and not choose either good. Subjects are told in advance that one of the seven questions will be binding — that one of the seven questions will be randomly drawn at the end of the experiment and whatever good they chose in that question is the one they would have to purchase right then and there, if they chose to purchase at all. This treatment is the only one of the six treatments we conduct that is theoretically incentive compatible. By randomly choosing the binding question, the subject can do no better than to truthfully reveal his or her preferences in each question. There is no violation of incentive compatibility due to sequencing of choices.

In contrast to the private-good case, the IPR treatment conducted with a public-good as the object of choice is <u>not</u> incentive compatible. This treatment is identical to the private goods treatment: subjects are presented with choice questions, each with two variations of a public good (with varying costs) and asked to choose which they prefer. Subjects are also given an opt-out choice. However, because this treatment uses a deliverable field-public good (i.e., a "real-world" public good), the provision rule is not theoretically incentive compatible and individuals may not reveal their preferences truthfully. Subjects may realize that they can free-ride by relying on the contributions of the general public (and others in the experiment), and thus not reveal their true choices. The IPR provision rule in this setting is essentially a modified voluntary contribution mechanism for the field-public good, which is known to not be incentive compatible (e.g. Holt and Laury, 1997). It is important to note that the direction of the bias in the IPR treatment with public-goods is known *a-priori*: subjects will opt-out of the market more frequently.

The next provision rule we describe is the group-provision rule (GPR), which is designed to mimic the most typical provision of public goods. In the GPR treatment, subjects are given the same choice questions as in the IPR treatment, however the provision rule is as follows. At the end of the experiment, one question is randomly chosen to be "binding" just as in the IPR treatment. The option that receives the greatest support in the binding question is the option that <u>everyone in the group</u> has to purchase, regardless of how they personally responded to the question. This is akin to a referendum, however there are three options that can be "voted for": two variations of the good, and the option to choose neither good. This provision rule is not incentive compatible in either the private or public goods treatment because a three-choice referendum is not incentive compatible (Moulin, 1988).[3] The public goods treatment has the additional influence that subjects may choose to free-ride. Unfortunately, the direction of the bias in both the public and private-goods treatment is not known. It will depend on the distribution of preferences across the three alternatives in a manner that cannot be known *a priori*.

While GPR invites strategic behavior, it is important to understand how subjects behave in this treatment compared to the other treatments as it mimics the likely *inferred* provision rule in a conjoint survey eliciting preferences over "real-world" public goods. We use the term inferred as we cannot know what survey respondents have assumed for a provision rule in past conjoint surveys. However, it would seem reasonable that respondents might infer that the option that receives the greatest support would be the likely candidate to be implemented (if the survey was at all consequential).

While the GPR is not a 'natural' provision rule for private-goods markets, we conduct private-goods treatments because subjects do not have the incentive to free-ride in this treatment, only to strategize over which of the three options they prefer. By observing how GPR responses for the

[3]For example, a subject may rank the three options as follows: option 1 is preferred, option 2 is second preferred, and option 3 is least preferred. If the subject believes that option 1 has little chance of being the option that is most frequently chosen by the group, but that either option 2 and 3 are likely to be the most frequently chosen option by the entire group, he may choose to vote for option 2 (rather than his preferred option: option 1) to help ensure that at least his second-best alternative is chosen, rather than his least preferred option. Thus, he has not revealed his true preferences in the vote.

private good vary as compared to IPR for the private goods, we will better be able to understand the responses to the public goods experiments in which free-riding is an added strategic bias.

The third we conduct are surveys where no provision rule (NPR) is described. This treatment is designed to closely mimic a typical conjoint survey where no provision rule is described. Subjects are simply asked to answer the choice questions, carefully thinking about their responses. We do not know how the lack of provision rule will affect individual responses. To the extent no provision rule discussion makes the survey less consequential in the eyes of respondents, we would expect increased noise in the responses and possibly an upward bias in the number of subjects who opt-in to the market.

The fourth treatment (RPR) corresponds to the IPR treatment except that real payment is involved.

The fifth treatment (CVM) has similar incentive compatibility properties to the IPR treatment. . In the CVM treatment involving public goods, subjects are also given eight choice questions, each question asking them whether they would choose one variant of a good (with varying costs), or to "opt-out" and not choose this good. As with the other treatments, subjects are told in advance that one of the eight questions will be binding — that one of the eight questions will be randomly drawn at the end of the experiment and whatever alternative the majority of the group chose in that question (whether to purchase or not to purchase) will be binding for all. With the GPR treatment, the group decision rule was not incentive compatible because there were choices between three alternatives. But unlike the GPR treatment, because the CVM treatment involves a binary referendum, there is no loss of incentive compatibility due to a group decision rule being used. Nonetheless, similar to the IPR treatment, because this treatment uses a deliverable field-public good where free riding is possible, the provision rule is not theoretically incentive compatible and individuals may not reveal their preferences truthfully.

The "Goods"

Private Good: The private good used in the experiments are t-shirts. This good is familiar and acceptable to our subjects (students), and it was possible to define a parsimonious number of attributes for this good. The t-shirts are all heather-grey, and have two attributes (besides cost). The t-shirts could have long sleeves or short sleeves (attribute 1, with two levels) and have one of two Georgia State University logos embossed on the front, or no logo at all (attribute 2, with three levels). The number of attributes selected was deliberately kept small so that a relatively small experimental design could be used. This has several advantages. A smaller experimental design will mean that there are more repetitions for each block, increasing statistical power for some of our planned statistical tests. Second, a simpler design is likely to mean that evaluation tasks will be cognitively simpler for respondents, reducing noise in our experiment. Third, the standard protocol in designing experiments is to start with a relatively simple design, determine if there are effects occurring, and then use more complex designs.

Prices of the t-shirts vary from \$8 to \$16. A 14-alternative experimental design, split into two blocks of 7 questions each is created. Alternatives are based on orthogonal main effects design, with implausible alternatives deleted. Since the objective for these treatments was to demonstrate whether provision rules were having an effect on choices, rather than demonstrate the effect on value estimates of specific t-shirt attributes, subjects only answer questions based on the first block. This reduces the sample size needed for these treatments by 50%.

Public Good: The public good chosen must have the following characteristics. First, it must be "deliverable." If subjects were to choose to purchase "the good," it must be credible that it could actually be provided in return for their payment. Second, in the individual provision rule (IPR) treatments, the good must be divisible in provision. Subjects must be able to connect their specific payment with a specific "amount" of the public good provided (this rule may be relaxed, but for acceptability of the survey instruments we feel it is important). Third, we must be able to alter the attributes of the public good and still be able to provide or "deliver" any combination of the attributes we offer to subjects.

The public good we use is *Trees Atlanta*, a local non-profit organization that seeks to protect the Atlanta metropolitan environment by planting and conserving trees. *Trees Atlanta* focuses on planting and maintaining trees in downtown Atlanta, which is ranked among cities as having the least amount of downtown tree cover. The lack of trees, along with increased asphalt and reflective buildings has contributed to an 8-10 degree increase in downtown summer temperatures since 1970. *Trees Atlanta* seeks to alleviate this "urban heat-island" effect through the planting and maintaining of trees in the downtown area. The organization has a "gift trees" program in which it will plant a tree in the metro-area for a contribution. Each donation leads directly to the planting of a tree during the current planting cycle (from October through March each year).

The good we offer is to have *Trees Atlanta* plant a tree through the gift trees program. There are three possible trees that can be planted: an Oak, Dogwood, or Crepe Myrtle. Further, each tree can be planted in one of two sizes: a small tree (5-feet) or a medium tree (8-feet). The cost to the subject for having a tree planted could be one of the following size values: $8, $12, $15, $19, $22, $26. Thus, we have a design with two attributes (apart from cost), one with three levels (type of tree) and one with two levels (size of tree at planting). This resulted in 36 alternatives, of which 4 were implausible. Thus, the alternatives are split into four equal sized blocks of eight questions. Questionnaires are administered with all blocks to allow estimation of conditional logit models and computation of value estimates for the attributes of the public good. These models will allow us to test whether changes to provision rules cause differences in parameter vectors and attribute value estimates. This is in comparison to our private goods experiments in which we can only test for differences in the percentages of subjects who choose to enter the market (we are not interested in valuing t-shirt attributes).

Finally, the costs of planting trees as subjects saw them were allocated through the process of developing an orthogonal attribute design for the survey. Subjects did not know the actual cost of tree planting, and were only told:

> For *Trees Atlanta* to plant a tree, you will need to make a contribution. The amount of a contribution ranges from $8 to $26 per tree and varies by the size and type of tree. Each contribution leads to a tree being planted that would not have been planted otherwise.

Experimental Design

Subjects for all hypothetical-payment surveys were recruited through classes at Georgia State University (GSU). Surveys were conducted during regularly scheduled class-meetings. Professors agreed to end class early, and would introduce the experimenters. Subjects were then asked if they wished to participate in a short survey in return for which they could win a cash prize through a random drawing.[4] Subjects who wished to participate completed a consent form which was then placed in a "ballot box." One name was drawn at random from the box at the end of the survey to determine the winner of the cash prize. Class sizes varied from approximately 20 to 120 students. Group size was kept similar for the GPR treatments, varying from 11 to 15 subjects in the private-good treatments (with one larger group of 48, whose responses were not statistically different than in the smaller groups), and varying from 21 to 37 subjects in the public-goods treatments.

After completing the consent form, subjects then read the instructions for survey (referred to as a "decision making experiment") while the experimenter read them aloud. The instructions for both the private and public goods followed the same basic format. First, subjects were told that it it was a decision making experiment, and that they would be asked if they wish to purchase a good. They were told that the survey was hypothetical, and that they were <u>not</u> actually being given the opportunity to purchase the good, but that we ask that they answer all questions as if the survey were for real. Next, the "good" was described. In the private goods experiment, subjects were shown two sample t-shirts (held up in the front of the room: one long sleeve with a GSU logo on it, and one short sleeve with different GSU logo) and the t-shirt attributes were described to them (with additional photographs of the various attributes in the instructions).

In the public goods treatment, the issue of tree removal in downtown Atlanta was described, along with its environmental impacts. *Trees Atlanta* was then introduced to subjects and their efforts to increase tree cover in Atlanta was described. The "Gift Trees" program was introduced and a large table was shown to subjects which described each of the three trees that could be planted and a few of the important characteristics of each tree. The attributes of the trees that were described were: flowering potential, mature height, growth rate (years to reach maturity), and shade potential. Color photographs of each tree type (at maturity) were also shown to subjects.

After describing the good, the choice questions were introduced. Subjects were told that they would answer eight questions (seven in the private goods survey) in which they would choose one of three options: the option to purchase one of two variations of the good, or the option to not purchase either variation presented in the question. A sample question was then provided, and the experimenter described in detail what it would mean if they checked the first option, the

[4]The drawing amount varied from $40 to $75 depending on the size of the class. Larger classes were given larger amounts in the drawing because the survey would take a little longer (due to the time needed to pass out materials to large numbers of people) and because the expected value of the lottery should be similar across treatments.

second option, or the third option (which was always the option to not purchase either variants of the good).

Following the description of the choice questions, the provision rules were introduced in the IPR and GPR treatments. To make concrete the provision rule, the following is an excerpt from the instructions of the IPR public-good treatment:

> **Determining Which Question Counts**
> You will answer eight questions just like the above example, except the type of tree, size, and cost will change. To help you respond to these questions as if they were for real, we would like to describe how we would determine which tree would have been planted if this experiment were real. In other words, what would have happened if this experiment involved real decisions where you could actually contribute and have a tree planted as a result of your answers to these questions. Here's how it would have worked.
>
> I have here an 8-sided die. I would have rolled the die after you had answered all eight questions. The number that comes up would be the question that is "binding." In other words, the number that is rolled would have told us which of the eight questions we would use to determine which tree is planted, if any at all.
>
> To help you see how this would have worked, consider again the example question above. Suppose that this was the second question you answered, and 2 was the number that came up when the die was rolled.
>
> ➔ If you had checked the box for Option I, you would have paid $22 today and a medium sized oak tree would have been planted by *Trees Atlanta*.
>
> ➔ If you had checked the box for Option II, you would have paid $12 today and a small sized crepe myrtle would have been planted by *Trees Atlanta*.
>
> ➔ If instead, you had checked the last box, telling us that you do not want to contribute to plant either tree at the specified cost, you would have contributed nothing and neither tree would have been planted.
>
> **How Contributions Would have Worked**
> After the die was rolled and the binding question is determined, one of us would have come around and verified which tree you chose, if you chose one at all. If you chose a tree, you would have paid us the specified contribution amount, and we would have provided you with a receipt. We would have sent your contribution to *Trees Atlanta*, specifying which tree-type and size you had chosen. After *Trees Atlanta* had planted the tree, they would mail a confirmation of the planting directly to you.
>
> **Review**
> While this experiment is hypothetical and you **are not** actually being given the opportunity to contribute money, we would like you respond to the questions as if they were for real -- as if you could actually make a contribution to *Trees Atlanta* today.

Remember, if this were for real, you would be under no obligation to contribute. If you did not wish to contribute to the planting of either of the trees, you would simply mark the box "I would not choose either option, and neither tree will be planted" at the end of each question. If you would make the specified contribution to purchase a tree, you would mark the box for the tree that you would choose.

Is everyone clear on how this experiment works?

In the GPR treatment, the script is modified as follows. After describing how the binding question is chosen (the eight-sided die was held up and shown to subjects), subjects are then told that the option that the most people chose in the binding question is the option that <u>everyone</u> in the group would have to abide by, regardless of which option they personally chose.

Determining Which Question Counts
The actual tree that <u>each person</u> in the group would pay to have planted by *Trees Atlanta* is the one that is chosen most often in the binding question. In other words, the option that the group chooses most often is the option that <u>everyone</u> in the group would have to abide by, even if they were not one of the people who chose that option.

To make sure everyone understands, here's how it would have worked with this group. Let's say the box for Option I was chosen most often in the example on the previous page. That is, more people checked the box for Option I than the box for Option II or the box for the option not to contribute at all.

➔ In this case, every person in this group would contribute \$22 since this is the option that the group chose most often. A medium sized oak tree would have been planted by *Trees Atlanta* for each individual in the group. So, with ______ people in the group here today, a total of \$________ would be collected from the group (this is the number of people in the group times \$22), and _______ medium sized oak trees would be planted by Trees Atlanta.

➔ If you had checked the box for Option II, you would <u>not</u> be able to contribute \$12 today for a small sized crepe myrtle. Instead, you would have to contribute \$22 for a medium-sized oak tree, since that was the option chosen most often by the group. Everyone else would also be contributing \$22, regardless of which option they personally chose.

➔ If you had checked the last box, telling us that you do not want to contribute to plant either tree at the specified cost, you would still have to contribute \$22 and have the tree planted in Option I and abide by the group decision.

Now let's say that the group didn't choose Option I most often, but instead chose the option to <u>not</u> contribute to purchase either tree most often. In this case, <u>nobody</u> in the group would be able to contribute to *Trees Atlanta* today.

How Contributions Would have Worked
After the die was rolled and the binding question is determined, one of us would have come around and verified which tree you chose, if you chose one at all. We would then determine which of the three options was picked most often: Option I, Option II, or the

Option to not contribute at all. If Option I or Option II was chosen most often by the group (one of the options to plant trees), you would have paid us the specified contribution amount, and we would have provided you with a receipt. We would have sent the group's total contribution to *Trees Atlanta*, specifying which tree-type and size and how many trees are to be planted. After *Trees Atlanta* had planted this group of trees, they would mail a confirmation of the planting directly to you. Each person in the group would receive a personal confirmation that the trees had been planted.

Review
While this experiment is hypothetical and you **are not** actually being given the opportunity to contribute money, we would like you to respond to the questions as if they are for real -- as if you would actually make a contribution to *Trees Atlanta* today.
Remember, if this were for real, after you had answered all the questions, we would have chosen the binding question by rolling the die. One of us would have then come around to see what option you chose so we could determine which one was chosen most frequently by the group. The option chosen most often would have determined what <u>all members of the group</u> would have to pay, if anything at all. If one of the options to plant a tree is chosen most frequently, then everyone in the group would have paid the specified cost, <u>regardless of which option they personally chose</u>, and ________ trees would have been planted by *Trees Atlanta.* On the other hand, if the option to <u>not</u> contribute to either of the trees shown in the binding question is chosen most frequently by the group, then no one would be able to contribute today, regardless of which option they personally chose.

Is everyone clear on how this experiment works?

The experimenter would have counted the number of people in the group at the beginning of the experiment, and would read aloud the exact number of subjects and the total contribution amount where there are blank places in the script.

Following the description of the provision rule and the review, subjects were lastly told the following:

"Finally, because this is an experiment in decision making, you may see the same tree appearing with different contribution amounts. Please answer each question as if that choice was the only one available to you."

In the NPR treatment, all aspects of the survey are identical except that after reviewing the sample question, subjects go directly to the review (which itself is shorter as there is no review of the provision rule).

3. Results

Preliminary results from the private and public goods experiments are reported here. A total of 358 surveys were conducted for the private goods experiment, and 1500 surveys conducted for the public goods experiment[5]. Table 1 reports the number of surveys completed for each treatment. The demographic characteristics of each of the sub-samples are very similar as might be expected. GSU is an urban university in which many students live at home, or are working adults. The demographic characteristics of our sample reflects this. The students in our samples are slightly older than a typical college sample (almost half of the students report working at least part-time). Mean household income is approximately $50,000-$55,000, while the mean amount of income the student him/herself earns ("own-income") is generally less than $16,000. In the private-goods samples, t-tests indicate the characteristics of subjects in the GPR and NPR treatments are not statistically different from each other. However, the subjects in the IPR treatment are younger on average than subjects in either of the two other treatments. Subjects in the IPR treatment also have a lower mean own-income as compared to subjects in the NPR treatment (but are not statistically different than subjects in the GPR treatment). In the public-good samples, t-tests indicate no significant differences in any of the demographic characteristics of subjects in the IPR and GPR treatments.

5 Several hundred additional surveys have been conducted, but are yet to be coded and analysed.

Table 1: Summary of Private and Public-Good Treatments

	IPR	GPR	NPR	RPR	CVM
Private Goods Experiments (T-shirts)					
Number of Obs.	140	113	105		
% male	43.5	42.5	48.5		
% Caucasian	47.1	45.1	50.5		
Mean Age	22.4	24.0	23.8		
Mean HH-Income	$57,682	$54,071	$50,170		
Public Goods Experiments (Trees)					
Number of Obs.	225	267	198	244	108
% male	43.1%	46.8%	48.2%	48.4%	46.3%
% Caucasian	41.5%	44.2%	38.4%	41.4%	51.9%
Mean Age (years)	21.4	22.9	20.8	22.3	21.2
Mean HH-Income	$54,232	$54,490	$52,891	$52,818	$54,463

Table 2 reports subject responses for each of the seven t-shirt choice questions (C1 to C7) in which subjects chose between two different t-shirt options (Alt1 and Alt2) or the option to choose neither t-shirt (CN). Several results emerge from these frequencies. First, the percentage of subjects who "opt-out" of the market (i.e., choose neither t-shirt in a question) is highest in the treatment that describes the incentive compatible provision rule (IPR). Averaging across all seven questions, more respondents chose to opt-out of the market (chose the option for "choose neither t-shirt") with the incentive compatible IPR provision rule (53.1%) as compared to the NPR treatment (40.3%) or the GPR treatment (37.6%). This is robust across all seven questions. For example, in question three, 80% of subjects opted-out of the market in the IPR treatment. However, only 63% and 66% of subjects opted-out in the NPR and GPR treatments.

The frequencies at which Alt1, Alt2 and CN were chosen are significantly different in the IPR treatment as compared to the NPR for four of seven questions, and are significantly different as compared to the GPR treatment in five of the seven questions. If instead, we conduct chi-square tests just on the proportion of subjects who opt-in or opt-out (effectively summing the number of people who choose either Alt1 or Alt 2), there are significant differences in the proportion opting-in the market for six of the seven questions across IPR and NPR as well as across IPR and GPR.

While there are differences in the answers between IPR and the other two treatments, we find very little difference in the choices made in the GPR and NPR treatments. The average number of respondents opting-out of the market across all seven questions in the NPR and GPR treatments are not statistically different from each other. Furthermore, chi-square tests conducted for the individual questions indicate that the choices made in the NPR and GPR treatment are <u>not</u> statistically different in six of seven questions (see last column, Table 2).[6]

These results suggest that, for the private-good, introducing a discussion of an incentive compatible decision rule in the survey results in subjects choosing different alternatives than when no provision rule is discussed. Interestingly, there appears to be very little difference in the responses subjects give to surveys with no provision rule and surveys which describe a "group-provision" of the good. Thus, it would appear that no discussion of a provision rule results in behavior that is not truth-revealing -- i.e., results in behavior that is similar to that observed in the <u>non</u>-incentive compatible GPR treatment, and statistically different than that observed in the incentive-compatible IPR treatment. We now turn to the surveys using public-goods as the object of choice to further examine our hypotheses.

[6]These results are also confirmed with a multinomial logit regression analysis. These models indicated that subjects were more likely to participate in the market (i.e., choose a t-shirt) under both the NPR and GPR treatments relative to the incentive compatible IPR treatment. Similar to the frequencies in Table 2, the magnitudes of the regression coefficients also indicated that subjects had a marginally greater propensity to choose a t-shirt in the GPR treatment as compared to the NPR or IPR treatments. Separate regressions indicated, however, that there were no significant differences in the propensity to opt-in the market in the NPR treatment as compared to the GPR treatment.

Table 2: Frequencies and Chi-square Tests: Private Good Choice Surveys

		IPR	NPR	GPR	IPR vs. NPR	IPR vs. GPR	NPR vs. GPR
C1	*Alt1(%)* *Alt2(%)* *CN(%)*	3.6 44.3 52.1	3.8 53.3 42.9	3.5 65.5 31.0	χ^2=0.036, p-val=0.982	χ^2=23.627, p-val=0.000	χ^2=3.453, p-val=0.178
C2	*Alt1(%)* *Alt2(%)* *CN(%)*	40.7 8.6 50.7	44.8 19.0 36.2	54.0 11.5 34.5	χ^2=10.537, p-val=0.050	χ^2=9.361, p-val=0.009	χ^2=3.023, p-val=0.221
C3	*Alt1(%)* *Alt2(%)* *CN(%)*	15.0 5.0 80.0	25.7 11.4 62.9	25.7 8.0 66.4	χ^2=9.906, p-val=0.007	χ^2=3.448 p-val=0.178	χ^2=0.782 p-val=0.676
C4	*Alt1(%)* *Alt2(%)* *CN(%)*	2.9 76.4 20.7	1.9 78.1 20.0	3.5 86.7 9.7	χ^2=3.586, p-val=0.166	χ^2=20.074, p-val=0.000	χ^2=4.927, p-val=0.085
C5	*Alt1(%)* *Alt2(%)* *CN(%)*	28.6 9.3 62.1	46.7 9.5 43.8	53.1 8.8 38.1	χ^2=4.358, p-val=0.113	χ^2=20.454, p-val=0.000	χ^2=0.919, p-val=0.632
C6	*Alt1(%)* *Alt2(%)* *CN(%)*	50.0 15.0 35.0	52.4 26.7 21.0	54.0 28.3 17.7	χ^2=5.537, p-val=0.062	χ^2=16.401, p-val=0.000	χ^2=0.379, p-val=0.827
C7	*Alt1(%)* *Alt2(%)* *CN(%)*	20.0 7.9 72.1	26.7 18.1 55.2	19.5 15.0 65.5	χ^2=14.460, p-val=0.001	χ^2=1.806, p-val=0.405	χ^2=2.480, p-val=0.289
N		140	105	113			

Note: Alt1(%)=percentage of respondents choosing alternative 1, Alt2(%)=percentage of respondents choosing alternative 2, CN(%)=percentage of respondents choosing choose neither, N=number of respondents

In the public-good surveys, there are five treatments each with four blocks of eight questions. For the NPR, IPR and GPR treatments it was possible to estimate separate models, but for the RPR and CVM treatments this was not possible because of insufficient variation in the former and a small sample size in the latter (some of the data for this treatment is still being coded). Therefore we initially report the results from a pooled model for the five data sets. After the results from the pooled model have been reported, we report the results from the three individual models.

Pooled model

The pooled model was estimated using a random parameters or mixed logit model (Hensher and Greene 2002). The random parameters model has three advantages over the standard conditional logit. First, it does not have the independence of irrelevant alternatives (IIA) property that is often violated with stated preference data. Second, it allows for individual heterogeneity within the all or a subset of the parameters of the model. Third, it is also possible with this model to allow for panel or repeated data through the use of random effects. Random effects models explicitly allow for correlation among error terms in the responses of each individuals (Anderson 2003).

In the mixed logit model the error term is assumed to be additively separable into two parts: one part is correlated across individuals (η_{iq}), but the other is independently and identically distributed over alternatives and individuals (ε_{iq}) (Hensher and Greene 2002):

$$U_{iq} = \beta' X_{iq} + [\eta_{iq} + \varepsilon_{iq}] \qquad (1)$$

In the mixed logit model, η is assumed to have a general distribution (eg normal, log-normal) while ε is assumed to have an iid extreme value distribution.

For a given value of η, the mixed logit model can be expressed as follows:

$$P(i \mid A) = \exp(\lambda V_i + \eta_i)/\textstyle\sum_j \exp(V_j + \eta_i),\ \forall\ i, j \in A \qquad (2)$$

As η is not given in practice, the unconditional choice probability in the above formula are integrated over all values of η, weighted by the density of η. In other words, this means that the parameters are not estimated directly; rather the parameters for distribution are estimated. The distribution is then integrated to determine choice probabilities.

When estimating the mixed logit model, one or more parameters (β's) are specified as having a mean and standard deviation. The actual distribution is chosen by the researcher (eg normal, log-normal). By specifying a standard deviation, the presence of heterogeneity for the parameter is explicitly modelled. However, in practice determining which variables to distribute and which distribution to choose can be challenging. McFadden and Train (2000) propose a Lagrange Multiplier test for the presence of random parameters, however the test does not indicate which distribution should be chosen for variables that are identified as being random. Most applications have chosen either to distribute only price (eg Layton 2000) or non-price variables (see Anderson 2003).

Once statistical models have been estimated, it is possible to derive welfare estimates for use in cost-benefit analysis. For this study, we estimate implicit prices which are the value of a unit change in an attribute. Implicit prices are calculated as follows, if utility is a linear function of all attributes:

$$IP = \beta_A/\beta_M \tag{3}$$

where IP is the implicit price, β_A represents the coefficient of the Ath non-monetary attribute, and β_M represents a monetary attribute.

The results of the random parameters logit model for the pooled data set are presented in Table 3. The attributes of the good each are significant in determining the alternative chosen by subjects, except for DOGWOOD. This means that oak trees are preferred to both dogwoods and crepe myrtles, but that dogwoods are not more preferred than crepe myrtles. Cost is significant and negative, as expected. The model indicates that the respondents in the RPR treatment (see RPR * ASC) are less likely to opt into the market at the 1% significance level, and similar for the NPR treatment at the 10% significance level (although the value of this latter coefficient is relatively small). What is striking is the very large value for RPR * ASC, indicating a much lower propensity to opt into the market with the real treatment.

However, opting in or of the market does not necessarily affect the marginal valuations of subjects. As compared to the IPR treatment, subjects are significantly less responsive to price in both the GPR and NPR treatments, but more responsive to price in the RPR treatment. There is no difference in the responsiveness to price between the CVM and IPR treatments, indicating that inclusion of the incentive compatible individual decision rule has effectively controlled the differences between choice modeling and CVM estimates.

This is further reflected in the implicit price estimates reported in Table 4. Compared to the RPR treatment, the estimates for the NPR treatment are 46-59% larger, while the estimates for the GPR treatment are 39-49% larger. However, the implicit prices for the IPR and CVM treatments are much closer to the RPR results, being 20-25% and 25-30% larger respectively.

Table 3: Random Parameters Logit Results for the Pooled Data set

	Coeff.	t-ratio	P-value		
SIZE	1.190	3.419	0.001	Log-likelihood	-5008.941
OAK	1.018	15.417	0.000	Chi-square	8291.591
DOGWOOD	0.032	0.625	0.532	Rho-square adj	0.452
ASC_CM	3.468	15.296	0.000	Number iterations	28
ASC_CVM	-0.517	-1.360	0.174	Numb observations	8333
COST	-0.201	-11.517	0.000		
COST * SIZE	-0.047	-2.386	0.017		
NPR * ASC	-0.453	-1.766	0.077		
RPR * ASC	-4.766	-11.754	0.000		
NPR * COST	0.037	2.182	0.029		
GPR * COST	0.026	3.027	0.002		
RPR * COST	-0.061	-2.221	0.026		
CVM * COST	-0.009	-0.365	0.715		
σ SIZE	1.948	18.127	0.000		
σ OAK	1.504	21.156	0.000		
σ DOGWOOD	0.986	15.956	0.000		

Note: ASC_CM=alternative-specific constant for Alt1 and 2 for the choice modeling treatments; ASC_CVM=alternative specific constant for the CVM treatment.

Table 4: Implicit Prices

	Implict Prices				
	RPR	**CVM**	**IPR**	**NPR**	**GPR**
Size	$3.84	$4.62	$4.78	$5.61	$5.35
Oak	$3.88	$4.84	$5.05	$6.18	$5.82
Dogwood	$0.12	$0.15	$0.16	$0.20	$0.18
	% Differences				
Size	-	20.2%	24.5%	46.1%	39.3%
Oak	-	24.8%	30.3%	59.3%	49.9%
Dogwood	-	24.8%	30.3%	59.3%	49.9%

Individual models

The results from random parameters logit models estimated allowing for grouped data are presented in Table 5. The models were weighted to allow for differences in the distribution of income and sample size across the treatments. In each of the models there are three normally distributed variables: SIZE, OAK and DOGWOOD. These variables were normally distributed based on the results of testing using the approach recommended by McFadden and Train (2000). All of the variables are significant and have the expected sign, except for the cost * size interaction in the NPR and IPR models. The rho-squared adjusted figures indicates that the explanatory power of the models is very good.

Likelihood ratio tests were used to determine whether the parameter vectors for these models were different. Test results ($\chi^2_{\text{npr vs ipr}}$ = 811.54, $\chi^2_{\text{npr vs gpr}}$ = 913.80, $\chi^2_{\text{gpr vs ipr}}$ = 880.224) indicated that the models were statistically different from one another. Further testing will examine whether scale differences will change this result (Swait and Louviere 1993).

Table 5: RPL Models for NPR, IPR and GPR Treatments

Model	Coeff.	t-ratio	P-value		
NPR					
SIZE	1.29	1.97	0.05	Log-likelihood	-1449.64
OAK	2.58	10.05	0.00	Chi-square	1209.44
DOGWOOD	1.48	8.19	0.00	Rho-square adj	0.29
ASC	2.26	6.42	0.00	Number iterations	20
COST	-0.13	-4.86	0.00	Numb observations	5565
COST * SIZE	-0.06	-1.58	0.11		
σ SIZE	1.49	8.87	0.00		
σ OAK	3.10	10.54	0.00		
σ DOGWOOD	1.92	9.72	0.00		
IPR					
SIZE	1.81	2.64	0.01	Log-likelihood	-1331.98
OAK	2.66	9.92	0.00	Chi-square	1286.65
DOGWOOD	1.34	6.93	0.00	Rho-square adj	0.32
ASC	3.40	8.71	0.00	Number iterations	16
COST	-0.21	-6.96	0.00	Numb observations	5394
COST * SIZE	-0.06	-1.60	0.11		
σ SIZE	1.78	9.60	0.00		
σ OAK	2.89	9.29	0.00		
σ DOGWOOD	1.83	8.94	0.00		
GPR					
SIZE	2.89	4.01	0.00	Log-likelihood	-1335.8
OAK	2.65	8.25	0.00	Chi-square	1372.19
DOGWOOD	1.12	5.43	0.00	Rho-square adj	0.34
ASC	2.88	7.14	0.00	Number iterations	16
COST	-0.14	-4.40	0.00	Numb observations	6405
COST * SIZE	-0.14	-3.49	0.00		
σ SIZE	2.04	10.46	0.00		
σ OAK	3.37	10.06	0.00		
σ DOGWOOD	2.23	9.51	0.00		

The results from these models were used to generate estimates of implicit prices. These implicit prices are reported in Table 6. For the oak and dogwood trees, the implicit prices are substantially larger for the NPR and GPR treatments than the IPR treatments. For the size attribute, the implicit price is only substantially larger in the GPR treatment.

To test whether implicit prices were equivalent between metropolitan and rural locations, and between manufacturing and service industries; the Krinsky and Robb bootstrapping procedure was used to generate standard errors for the implicit prices. This test involves drawing a large number of random parameter vector estimates from a multivariate normal distribution with mean and variance equal to the β vector and a variance-covariance matrix from the estimated random parameters logit model (Park, Loomis, and Creel, 1991). However, the overlapping confidence intervals generated using this procedure may not correspond to the stated level of significance implied. Therefore, Poe, Giraud and Loomis (2003) proposed an alternate measure for testing the equality of means that involves taking differences between every element in the distributions generated using the Krinsky and Robb procedure.

Using this approach, significant differences at the 10% level for the oak attribute in both comparisons are identified, and significant at the 5% level for the dogwood attribute for the IPR/NPR comparison only. For the size attribute, the implicit price is significantly different at the 10% level in the GPR/IPR comparison.

Table 6: Implicit Prices for NPR, IPR and GPR RPL models

	IPR	**NPR**	**GPR**
SIZE	$6.70	$6.76	$10.43
OAK	$12.74	$19.32	$19.55
DOGWOOD	$6.41	$11.07	$8.26
		% difference	
SIZE		0.83%	55.62%
OAK		51.62%	53.42%
DOGWOOD		72.78%	28.90%
		P-value tests	
	IPR vs NPR	**IPR vs GPR**	**NPR vs GPR**
SIZE	0.495	0.091	0.118
OAK	0.076	0.080	0.472
DOGWOOD	0.038	0.246	0.226

4. Conclusions

This research reports preliminary results from choice surveys that indicate that explicit discussions of provision rules do have a significant impact on choices subjects make in conjoint surveys. We examine the impacts of provision rules in choice surveys using both public and private goods. Five survey treatments are conducted: one in which an incentive-compatible provision rule is described (IPR), one in which a non-incentive compatible decision rule is described (GPR – a non-incentive compatible decision rule), one in which no provision rule is described (NPR – which most closely corresponds to what is commonly done in most choice modeling surveys), a conjoint treatment involving real payment (RPR) and a hypothetical contingent valuation treatment (CVM).

Results indicate that in surveys using private goods, subjects opt to purchase the private good more often when either the non-incentive compatible group-provision rule is described (GPR), or when no provision rule is described (NPR). Results from surveys using a public good as the object of choice indicate that the provision rule did have a minor affect on the proportion of subjects who "opt-in" or "opt-out" of the market. However, this effect was relatively small and of marginal significance. However, whether a treatment is real or hypothetical did have a very substantial effect on whether subjects opt-in or opt-out of the market, with a much smaller proportion of respondents opting into the market with the real treatment.

Perhaps of greater importance is the effect of the provision rules on marginal values. Using the pooled model, we find that subjects were significantly less responsive to price in both the GPR and NPR treatments as compared to the IPR treatment. In addition, respondents in the RPR treatment were more responsive to price in the RPR treatment when compared to the IPR treatment. However, there was no difference in the price responsiveness of respondents in the IPR and CVM treatments.

Estimates of marginal values were computed to be 39-50% and 46-59% higher in surveys which discuss the group provision rule or no provision rule, respectively, as compared to the RPR treatment. However, the results in the IPR and CVM treatments were only 25-30% and 20-25% higher respectively. Moreover, no significant difference was identified between the IPR and CVM estimates (although the results may change when the remainder of the CVM data have been coded).

We find the results thus far to be provocative as they suggest that commonly employed methods for conducting choice surveys, which are akin to the NPR treatments conducted here, may lead to biased estimates of the marginal values of attributes. Moreover, the inclusion of a provision rule appears to eliminate the divergence between contingent valuation and conjoint estimates identified in previous studies. However, while the inclusion of a provision rule was found to substantially reduce the degree of overstatement with hypothetical estimates, they did not fully eliminate them.

Much work remains however before this conclusion is supported strongly. In addition to collecting and coding the remaining data, the next step in this research is to more fully develop the models used to describe the choices made in the public-goods surveys.

REFERENCES

Adamowicz, W.P. Boxall, P. Williams, M. and Louviere, J. (1998). Sated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation, *American Journal of Agricultural Economics*, 80: 64-75.

Anderson, C.M. (2003). Policy Implication of Alternative Economic Approaches to Population Heterogeneity in Stated Preference Surveys. Paper presented at the American Agricultural Economics Association Meetings.

Arrow, K., Solow, P., Portney, P., Leaner, E., Radner, R., and Schuman, H. (1993). Report of the NOAA Panel on Contingent Valuation. *Federal Register*, 58: 4601-4614.

Bearden, W.O. and Netemeyer, R.G. (1999). *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. Second Edition. Sage Publications, Thousand Oaks.

Ben-Akiva, M. and Morikawa, T. (1990). Estimation of Switching Models from Revealed Preferences and Stated Intentions. *Transportation Research A*, 24A (6): 485-495.

Bennett, J. and Blamey, R. (2001). The Strengths and Weaknesses of Environmental Choice Modelling. In Bennett, J. and Blamey, R. (Eds). *The Choice Modelling Approach to Environmental Valuation.* Edward Elgar, Cheltenham.

Bennett, J. and Morrison, M. (2001). Estimating the Environmental Values of New South Wales Rivers. *Proceedings of the Third Annual Stream Management Conference*, 1: 29-34.

Bjornstad, D., Cummings, R. and Taylor, L. (1997). A Learning Design for Reducing Hypothetical Bias in the Contingent Valuation Method. *Environmental and Resource Economics*, 10: 207-221.

Blamey. R. and Bennett, J. (2001). Yea-Saying and Validation of a Choice Model of Green Product Choice. In Bennett, J. and Blamey, R. (Eds). *The Choice Modelling Approach to Environmental Valuation*. Edward Elgar, Cheltenham.

Blamey, R.K., Bennett, J.W., Louviere, J.J., Morrison, M.D. and Rolfe, J. (2000). A Test of Policy Labels in Environmental Choice Modelling Studies. *Ecological Economics*, 32: 269-286.

Boxall, P.C., Adamowicz, W.L., Swait, J., Williams, M. and Louviere, J. (1996). A Comparison of Stated Preference Methods for Environmental Valuation. *Ecological Economics*, 18: 243-253.

Boyle, K.J., Holmes, T.P. , Teisl, M.F. and Roe, B. (2001). A Comparison of Conjoint Analysis Response Formats. *American Journal of Agricultural Economics*, 83(2): 441-454.

Breffle, W.S., Morey, E.R., Rowe, R.D., Waldman, D.M., Wytinck, S.M. (1999). Recreational Fishing Damages from Recreational Fishing Advisories in the Waters of Green Bay. Prepared by Status Consulting Inc for the US Fish and Wildlife Service, US Department of the Interior and the US Department of Justice.

Carlsson, F. and Martinnson, P. (2001). Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? *Journal of Environmental Economics and Management*, 41: 179-192.

Carson, R.T., Groves, T. and Machina, M.J. (2000). Incentive and Informational Properties of Preference Questions. Draft Working Paper, Department of Economics, University of California, La Jolla.

Christie, M. and Azevedo, C. (2002). Testing the Consistency in Benefit Estimates Across Contingent Valuation and Choice Experiments: A Multiple Policy Option Application. Paper presented at the Second World Congress of Environmental and Resource Economists, Monterey.

Cummings, R.G., Harrison, G.W. and E.E. Rutstrom (1995). Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?, *American Economic Review*, 85(1): 260-66.

Cummings, R.G., Elliott, S., Harrison, G.W. and Murphy, J. (1997). Are Hypothetical Referenda Incentive Compatible? *Journal of Political Economy*, 105(3): 609-621.

Cummings, R. and Taylor, L. (1999). Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method, *American Economic Review*, 89: 649-665.

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

Foster, V and Mourato, S. (2003). Elicitation Formats and Sensitivity to Scope. *Environmental and Resource Economics*, 24: 141-160.

Haab, T.C., Huang, J. and Whitehead, J.C. (1999). Are Hypothetical Referenda Incentive Compatible? A Comment. *Journal of Political Economy*, 107(1): 186-196.

Hanley, N., MacMillan, D., Wright, R.E., Bullock, C., Simpson, I., Parsisson, D. and B. Crabtree (1998). Contingent Valuation Versus Choice Experiments: Estimating the Benefits of Environmentally Sensitive Areas in Scotland. *Journal of Agricultural Economics*, 49(1): 1-15.

Hensher and Greene (2002). The Mixed Logit Model: The State of Practice and Warnings for the Unwary, *Transportation Research, B*

Hausman, J. and McFadden, D. (1984). Specification Tests for the Multinomial Logit Model. *Econometrica*, 52: 1219-1240.

Hoehn, J.P. and Randall, A. (1987). A Satisfactory Benefit-Cost Indicator from Contingent Valuation. *Journal of Environmental Economics and Management*, 14(3): 226-247.

Holt, Charles A., and Susan K. Laury, 1997, "Theoretical Explanations of Treatment Effects in Voluntary Contribution Games," forthcoming in the *Handbook of Experimental Economic Results*, C. Platt and V. Smith, eds., New York: Elsevier Press.

Layton, D. (2000). Random Coefficient Models for Stated Preference Surveys. *Journal of Environmental Economics and Management*, 40: 21-36.

Lockwood, M. and Carberry, D. (1998). Stated Preference Surveys of Remnant Native Vegetation Conservation. Report No.104, The Johnstone Centre, Charles Sturt University.

Lusk, J. (2003) Effects of Cheap Talk on Consumer Willingness-to-Pay for Golden Rice. American Journal of Agricultural Economics, 4: 840

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

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response, *Journal of Applied Econometrics*, 15: 447-470.

Morrison, M. and Bennett, J. (2000). Choice Modelling, Non-Use Values and Benefit Transfer. Economic Analysis and Policy, 30(1): 13-32.

Moulin, Hervé. Axioms of Cooperative Decision Making Cambridge: Cambridge University Press, 1988.

Murphy, J.J., Stevens, T. and Weatherhead, D. (2003). An Empirical Study of Hypothetical Bias in Voluntary Contribution Contingent Valuation: Does Cheap Talk Matter?, Worrking Paper No.2003-2, Department of Resource Economics, University of Massachuseets, Amherst.

Park, T., J.B. Loomis and M. Creel. “Confidence Intervals for Evaluating Benefits from Dichotomous Choice Contingent Valuation Studies.” *Land Economics*, 67(1991): 64-73.

Poe, G. L., K. L. Giraud, and J. B. Loomis. (2003)Computational methods for measuring the difference of empirical distributions. Forthcoming *American Journal of Agricultural Economics*.

Prince, R. McKee, M., Ben-David, S. and Bagnoli, M. (1992). Improving the Contingent Valuation Method: Implementing the Contribution Game. *Journal of Environmental Economics and Management*, 23: 78-90.

Smith, V.K. & Desvousges, W.H. (1986). *Measuring Water Quality Benefits*. Kluwer-Nojhoff Publishing, Boston.

Stevens, T.H., Belkner, R., Dennis, D. Kittredge, D. and Willis, C. (2000). Comparison of Contingent Valuation and Conjoint Analysis in Ecosystem Management. Ecological Economics, 32: 63-74.

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.

Takatsuka, Y., Kahn, J. and Stewart, S. (2002). Choice Model and Contingent Valuation Estimates of the Value of Ecosystem Attributes. Paper presented at the 72nd Annual Meeting of the Southern Economic Association, New Orleans, November.

Taylor, L.O. (1998).Incentive Compatible Referenda and the Valuation of Environmental Goods. *Agricultural and Resource Economics Review*, 132-139.

Whinstanley Newell, L. and Swallow, S. (2002). Are Stated Preferences Invariant to the Prospect of Real-Money Choice. Draft Working Paper, Department of Environmental and Natural Resource Economics, University of Rhode Island.