STATED PREFERENCE: WHAT DO WE KNOW? WHERE DO WE GO?

PROCEEDINGS
SESSION ONE

THEORY AND DESIGN OF STATED PREFERENCE METHODS

A WORKSHOP SPONSORED BY THE US ENVIRONMENTAL PROTECTION AGENCY'S NATIONAL CENTER FOR ENVIRONMENTAL ECONOMICS AND NATIONAL CENTER FOR ENVIRONMENTAL RESEARCH

October 12-13, 2000
Doubletree Hotel, Park Terrace
Washington, DC

Edited by Sylvan Environmental Consultants for the
Environmental Law Institute
1616 P Street NW, Washington, D.C. 20036

ACKNOWLEDGEMENTS

Sections of this report, indicated as "summarizations," were prepared by Sylvan Environmental Consultants for the Environmental Law Institute with funding from the National Center for Environmental Economics. ELI wishes to thank Matthew Clark of EPA's Office of Research and Development and Kelly Brown, Julie Hewitt, Nicole Owens and project officer Alan Carlin of National Center for Environmental Economics.

DISCLAIMER

These proceedings are being distributed in the interest of increasing public understanding and knowledge of the issues discussed at the Workshop and have been prepared independently of the Workshop. Although the proceedings have been funded in part by the United States Environmental Protection Agency under Cooperative Agreement CR-826755-01 to the Environmental Law Institute, it may not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Session I Proceedings

Table of Contents

Introduction to the Workshop

by Rick Farrell, Associate Administrator, US EPA Office of Policy, Economics, and Innovation

I'm happy to be here today to open the sixth workshop in the Environmental Policy and Economics workshop series. This series is cosponsored by the EPA Office of Research and Development's (ORD's) National Center for Environmental Research and the EPA Office of Policy, Economics and Innovation's (OPEI's) National Center for Environmental Economics.

The purpose of the series is to provide a forum for in-depth discussions on specific topics that further the use of economics as a tool for environmental decision-making. We also hope to showcase some of the research funded under the STAR (Science to Achieve Results) grants program. This workshop will highlight Stated Preference research and provide direction for further research in the future. Four-point-four million dollars has been spent on Stated Preference research through the STAR grants program — this is about one third of the joint NSF/EPA Environmental Social Science program budget. This program has funded some very notable researchers in the field, many of whom are in the room.

Economic analysis has played an important role in EPA's regulatory process and the role of economics continues to grow. In 1993, President Clinton signed Executive Order 12866 (replacing E.O. 12291) which requires benefit-cost analyses be conducted for all regulatory actions estimated to have an annual economic impact of more than $100 million. The 1996 amendments to the Safe Drinking Water Act allow, for the first time, the consideration of benefits and costs in setting maximum contaminant levels. The amendments even specify that EPA may measure benefits in terms of willingness to pay. The Small Business Regulatory Enforcement Fairness Act of 1996 gives Congress the opportunity to review and approve or disapprove environmental regulations based upon benefit-cost analyses, among other things. The Unfunded Mandates Reform Act of 1995 requires us to select the least costly, or least burdensome regulatory option or to provide an explanation of my we have not done so. Further legislative language requires the Office of Management and Budget to prepare the Thompson Report, providing estimates of the total annual costs and benefits associated with all federal regulations.

Because of the growing importance of economics in the regulatory process, the research and ideas to be presented today and tomorrow are extremely important. For many environmental goods and services, stated preference methods are the only available methods to assess the values, or benefits, associated with environmental goods.

As a testimony to the Agency's commitment to performing sound economic analyses, the National Center for Environmental Economics has recently revised the agency's economic guidelines. The new Guidelines for Preparing Economic Analyses will be released this month. It is worth noting that the new Guidelines include a much more detailed treatment of stated preference methods than did the previous version. This reflects the increased prominence and importance of these methods and the Agency's interest in them.

But we want to make sure that the numbers generated from these studies are appropriate for policy analysis and that the methods are sound and pass scientific muster in the world of environmental policy making, which can often be adversarial. We are asking you, the experts, to

evaluate the current state of stated preference methods and provide insight into how they can be further refined. We also hope that the presentations and discussions at this conference will help EPA and the other agencies present determine how to plan future research.

I'd like to thank you again for coming. You're all engaged in groundbreaking work and I hope that the lively discussion that will take place over the next two days will help us refine stated preference methods for use in policy analysis.

INCENTIVE AND INFORMATIONAL PROPERTIES OF PREFERENCE QUESTIONS[1]

Richard T. Carson
Theodore Groves
Mark J. Machina

Department of Economics, 0508
University of California, San Diego
La Jolla, CA 92093
rcarson@ucsd.edu

Draft
February 2000

[1] Earlier versions of this paper were presented in Oslo as a plenary address to the European Association of Environmental and Resource Economics, as an invited paper at the Japanese Form on Environmental Valuation meeting in Kobe, and at a NOAA conference on stated preference methods. Support of U.S. Environmental Protection Agency cooperative agreement R-824698 in carrying out the research reported on in this paper is gratefully acknowledged. The views expressed are those of the authors and not necessarily those of the U.S. Environmental Protection Agency.

Introduction

Businesses and governments frequently use surveys to help determine the relevant public's preferences toward different actions that could be taken. Applications are particularly common in environmental valuation (Mitchell and Carson, 1989), health care (McDowell and Newell, 1996), marketing (Louviere, 1994), political science (King, 1989) and transportation (Hensher, 1994). As long as the economic agents (hereafter, "agents") being surveyed believe that the survey responses might influence actions taken by businesses and governments (hereafter, "agency"), the standard economic framework suggests that the agents should respond to the survey in such a way as to maximize expected utility.

Given the billions of dollars spent annually on surveying and the frequently voiced concern that marketing surveys determine the fate of products and that major political decisions are largely poll-driven, the position of many economists that survey responses should be ignored as a source of information on preferences is somewhat surprising. These economists seem to regard survey responses as either completely meaningless because they are answers to hypothetical questions or as completely useless because agents will respond strategically. The first reason violates the standard rationality condition assumed of agents if agents believe that agency decisions are being made at least in part on the basis of the survey responses. The second reason stops short of the more relevant question: what are the strategic incentives and how should they influence responses?

In this paper, we systematically explore implications of the economic maximization framework for the behavior that one should expect to see from rational agents answering survey questions concerning preferences. The economic literature on neoclassical choice theory and mechanism design (Hurwitz, 1986; Groves, Radner and Reiter, 1987; Varian, 1992) provides the theoretical foundation for our work. This body of work can be contrasted with those who reject this framework in favor of other psychologically based theories (e.g., Kahneman, Slovic and Tversky, 1982; Sugden, 1999, McFadden, 1999). We believe that at least some of the evidence put forward in favor of those theories, particularly with respect to what differences should be expected with respect to asking questions using different response modes, has been incorrectly interpreted. We have endeavored here to put forth our results in an intuitive non-mathematical fashion as we hope fundamentally that our models represent a common sense approach to thinking about how agents should view preference questions. In the model informally presented here, agents are assumed to decide (1) whether they care about how the outcome might be influenced by the answers they provided, (2) whether the aspects of the scenario described are plausible, and (3) how the survey results are likely to be used. Judgements respecting these assumptions need not be elaborately or explicitly articulated any more than most judgements in life are. These three assumptions combined with the basic maximizing rationality assumption are capable of yielding a surprisingly rich picture of the manner in which agents should respond to survey questions.

A major reason that estimates of economic value from surveys are looked upon with suspicion by economists is a number of results that seem inconsistent with respect to economic intuition. These anomalous results have been interpreted by different researchers as evidence of (a)

the hypothetical nature of the question, (b) strategic behavior[2], or (c) preferences that are either ill defined or inconsistent with economic theory. In attempting to systematically categorize these anomalies it becomes immediately apparent that there is an antecedent question: does a survey question need to meet certain conditions before it should be expected to produce results that are consistent with economic theory?

This question turns out to be relatively easy to address from the standpoint of economic theory. First, the agent answering a preference survey question must perceive responses to the survey question as potentially influencing agency action. Second, the agent needs to care about what the outcome of that action is.[3] We will term surveys that meet these two basic criteria as *consequential* survey questions and those that don't as *inconsequential* survey questions. In more formal terms, we can state the following:

Consequential and Inconsequential Preference Survey Questions:

A. If the survey results *are* seen by the agent as *potentially* influencing agency actions and the agent cares about the outcome of that action, then the agent should treat the survey question as an opportunity to influence those actions. In the case of a *consequential* survey question, standard economic theory applies and *the response* to the question should be interpretable using mechanism design theory concerning incentive structures.

B. If the survey responses *are not* seen as having *any* influence on agency decisions or the agent is indifferent to all possible outcomes of the agency decision, then *all possible responses* by the agent will be perceived as having the same influence on the final outcome. In this case of an *inconsequential* survey question, economic theory makes no predictions about the nature of the responses to the survey given by the agent.

Most preference survey questions asked by businesses and governments meet the two basic criteria for being a *consequential* survey question, and hence, should be interpretable in economic terms.[4] There are, however, many preference survey questions that do not meet these criteria. While

[2] The possibility of strategic misrepresentation of preferences has long been seen as one of the central problems in public economics. Samuelson (1954) argued: "It is in the selfish interest of each person to give false signals to pretend to have less interest in a given collective activity than he really has." He made specific reference to the possibility of strategic behavior with respect to the use of surveys. Samuelson's admonition, repeated in many textbook discussions of public goods, had a profound effect on how many economists view the survey questions. The mistaken inference made by many from this admonition was to equate strategic behavior with lying. As the term is used in the modern mechanism design literature in economics, strategic behavior is merely synonymous with a rational agent maximizing (broadly defined) self-interest. Mechanism design theory has shown that the optimal strategic behavior for agents in many instances is to truthfully reveal their preferences. Whether this is the case or not depends upon the particular format of the preference question asked and other aspects of the scenario, including the type of good involved.

[3] For instance, a non-smoker may not care about the addition of a new type of cigarette with a much lower nicotine level and a higher price to the current cigarette choice set. Confusion often exists over the magnitude of the possible change in utility from agency action and the incentives the agent faces in the response given to the question. The size of the utility change generally does not influence the incentive structure of the question as long as there are differences in utility levels between different agency actions. The size of the utility change can influence agent participation in the survey.

[4] Marketing research firms, in particular, face a constant battle between asking questions to only those who are currently using a product category and trying to reach the larger and harder to identify population of all potential users. For public goods provided via taxation, the situation is generally easier. Even if the respondent does not care whether the good is provided at zero cost, the respondent does care about its provision if the tax cost is positive.

most of these *inconsequential* survey questions could be characterized as issuing from psychology lab exercises with undergraduates, there are plenty of real world examples.[5] It is pointless to try to explain apparent economic anomalies in *inconsequential* survey questions because any response to such a question has the same effect on the agent's utility. We are formally rejecting the notion sometimes advanced by proponents of the use of preference survey questions, which if a respondent perceives no gain or loss from how a preference survey is answered then that respondent will truthfully answer the question. While such an assumption may indeed be true, there is no basis in economic theory to either support or deny it.

Among questions meeting these two criteria for being *consequential* to the agent, we examine five key issues which should illustrate both the power and limitations of economic theory to explain a large body of empirical evidence related to the performance of survey questions under particular conditions. First, we look at the properties of binary discrete choice questions under different circumstances. In particular, we examine whether binary discrete choice questions are incentive compatible in the sense that truthful preference revelation represents an optimal (and the dominant) strategy for the agent. The empirical evidence suggests that such questions often work well: they predict actual behavior quite closely and they are sensitive to factors such as the scope of the good being valued. However, there are instances where such questions perform quite badly. Second, we consider the reasons responses to repeated binary discrete choice questions (*e.g.*, double-bounded dichotomous choice) by the same respondent are often inconsistent with each other. We also consider what information might be provided to the agent by the second choice question in this section. Third, we look at whether binary discrete choice questions and open-ended continuous response questions should produce similar estimates of statistics such as mean or median willingness to pay (WTP). In this section, we pay particular attention to the issue of what role, if any, information on cost should have on reported WTP values. Fourth, we consider the implication of moving from valuation of a single good to valuation of multiple goods, first in the context of the sequence of pair comparisons and then in the context of the increasingly popular multinomial choice questions. To begin to understand these issues, it is necessary to first confront what we have termed the face value dilemma.

The Face Value Dilemma

Economists tend to either reject preference survey results out of hand or treat the answers as truthful responses to the question asked. We term this latter behavior as taking the survey answers at *face value*. The two positions are not unrelated as both are result-oriented rather than process-oriented; many economists who reject the use of survey questions do so because the results are anomalous if taken at face value. We believe that either rejecting the usefulness of the preference survey answers or taking them at face value is likely to be wrong in many circumstances even when the two basic criteria for a *consequential* preference survey question have been met.

The *face value assumption* can be formally defined as "the assumption that respondents *always truthfully* answered *the specific* preference question intended to be asked". There are two key parts of this assumption: (a) respondents always truthfully reveal preferences, and (b) the specific question

[5] Inconsequential preference questions can most often be identified by having one or more of the following identifying characteristics: (a) being asked of a population or at a location that is unlikely from the perspective of an agency seeking input on a decision, (b) providing few if any details about the goods and how they would be provided, (c) asking about goods that are implausible to provide, and (d) providing prices for the goods that are implausible.

being asked is the one being answered. Note that (a) and (b) are both very strong assumptions. While the mainstream economic position is that (a) is dubious due to the strategic behavior, this assumption is routinely maintained in marketing research, political polling, psychology, sociology and other fields heavily dependent on survey research. In contrast, while economists who do use survey results routinely seem to believe (b), survey researchers have shown this to be a dubious assumption (Sudman, Bradburn, Schwarz. 1996).

Interpreting responses to survey questions appropriately requires consideration of the possibility that neither part of the face value assumption maybe be true. For truthfully revealing preferences, objections that agents may be responding strategically are insufficient to reject the use of *consequential* preference survey questions, as it may be in the respondent's strategic interest to truthfully reveal their preferences under some question formats in particular contexts.[6]

With respect to the decoupling of question and answer, the survey research community's usual rationale for the possibility that respondents may answer a different question than the one asked, is simply that respondents may not understand the question actually asked and instead answer the question that they think is being asked. Part of the survey designer's art lies in the crafting of language that elicits the answer to the question that the researcher intended to ask (Payne, 1951). From the perspective of preference survey questions for non-marketed goods or new consumer products, this issue needs to be taken particularly seriously since the development of questionnaires describing such goods is among the more difficult of survey design tasks; and most economists developing such surveys have little formal training in survey design. The pre-eminent issue here is that if survey responses are to be taken at face value, the question as written should elicit the answer to the question intended to be answered by the designer with all the conditions with which the designer wanted it answered. If this does not happen the results can easily be taken as implying violations of economic theory, when what has in reality happened is that agents have answered a different question.[7]

A further issue should be raised which concerns asking preference questions with implausible premises, for example, asking a binary discrete choice question with an implausibly high or low cost for providing the good. In such instances, respondents are likely to substitute what they consider to be the expected cost for the good and answer on that basis. Another easily recognized variant of this issue concerns implausible characteristics of the good provided, such as an assertion that a risk reduction program would be 100% successful, an assertion which is likely to

[6] Furthermore, under other question formats, the expected direction of the bias in responses can be theoretically predicted in some instances and empirically confirmed. In such cases, the results, even if biased, may be useful and often sufficient for agency decision making (Hoehn and Randall, 1987).

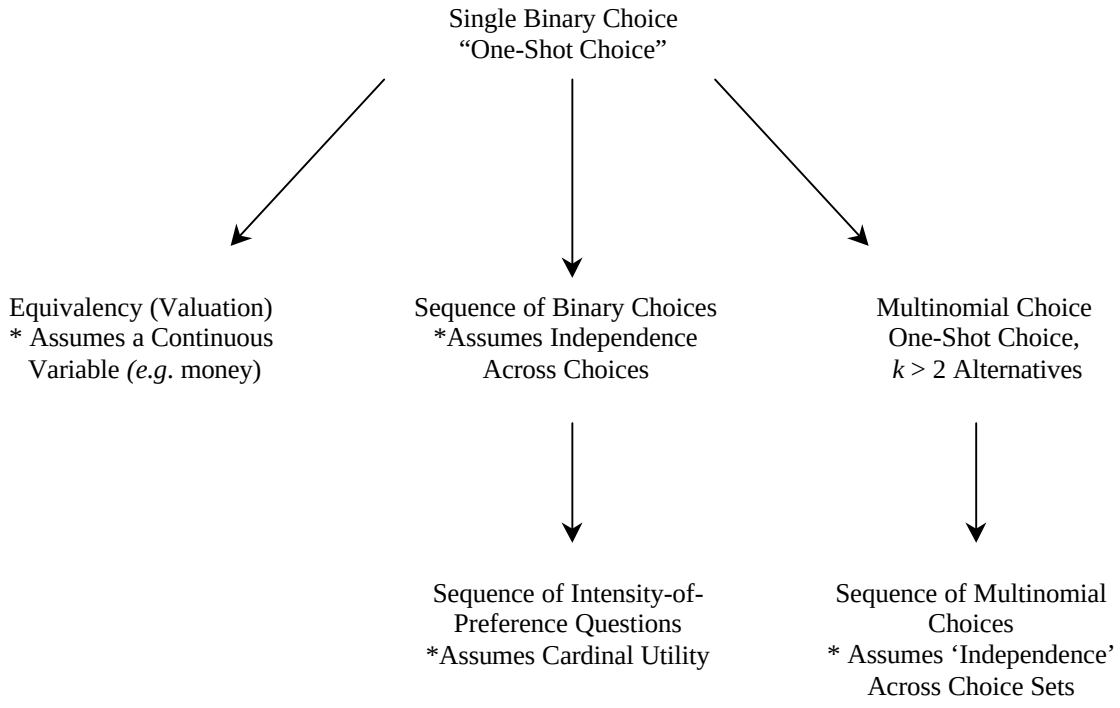
[7] For example, if a subset of agents does not believe that the good can be provided in any amount, these agents should be insensitive to the quantity (scope) of the good to be provided even though they may place a positive value on it. Divergences between the intended and answered question will always occur to a greater or lesser degree. The survey designer should endeavor to minimize them and the analyst should determine how they need be taken into account in order to arrive at reliable estimates. It should be noted that there is nothing unique about the use of stated preference data with respect to this issue. Most economic analyses of revealed preference data use objective indicators of good attributes to predict agent choices even though using agent perceptions of them can usually be shown to provide better predictions (Adamowicz *et al.*, 1997). Estimates of the value of a statistical life based upon hedonic wage equations have always been plagued by the need to make the demonstrably false assumption that agents were aware of the objective risk level used as a predictor variable in the equation.

be discounted by agents. More complicated variants of the issue manifest themselves when a respondent is given information at various points in a survey that is inconsistent. Examples include providing two different cost numbers in the double-bounded dichotomous-choice elicitation format and asking respondents about the provision of different levels of the same public good at different places in the survey. A key implication of this line of argument is that there are likely to be limits to the range of preference questions that a researcher can expect to have respondents answer. Survey questions can extend the range of goods and their attributes, including price, considerably beyond what agents have previously experienced; but any counterfactual scenarios must be credible portraits of possible future outcomes.

A Simple Typology of Elicitation Formats

The truthful preference revelation part of the face value assumption implies different conditions for different elicitation formats. This can most easily be seen by noting, that from a strategic perspective, all of the standard question formats can be shown to be generalizations of the single binary discrete choice format (Figure 1). Under this format, the respondent is told about two different alternatives and is assumed to pick which of the two alternatives provides the highest level of utility. As we discuss at length below, this assumption may be justified under some sets of conditions but not others.

FIGURE 1



There are three basic ways a single binary discrete choice question can be generalized. These are the open-ended matching type question, a sequence of binary choice choices, and the multinomial choice question. Matching questions, rather than ask for a choice between two alternatives, drops an attribute level (typically cost) of the second choice and asks the agent to provide the quantity of the attribute level that would make the agent see the two choices as equivalent in terms of utility to the agent. A sequence of binary choice questions adds the assumption that the agent answers each pair of choice independently to the assumption made of the single binary discrete choice. A number of different formats can be shown to be strategically manifestations of the sequence of binary discrete choice questions including the popular double bounded dichotomous choice format in contingent valuation (Hanemann, Loomis, and Kanninen, 1991) and the complete ranking of alternatives popular in marketing which can always be exploded into a set of binary paired comparisons if the independence assumption holds (Chapman and Staelin, 1982). Another commonly used variant of the sequence of binary choices asks agents to "rate" one choice relative to the other on a numeric scale (*e.g.*, 1 to 10) and exploits the information revealed about preference intensity (Johnson and Desvousges, 1997). This adds the assumption of cardinal utility. A multinomial choice question adds the assumption that the agent picks the most preferred out of $k > 2$ alternatives. A popular variant of this format, a sequence of multinomial choice sets (Louviere, 1994) adds the same assumption that a sequence of binary choice questions does, independence in responses across the choice sets.

For each of these questions formats it is possible to look at the divergence between the face value response and the strategic response. It is also possible to look at differences in the set of information conveyed by a particular elicitation format. Because the different elicitation formats are generalizations of the binary discrete choice format and because it can be shown that the binary discrete choice format has different strategic properties in different context we start with an examination of that format.

Binary Discrete Choice Preference Questions

A single binary discrete choice question between two alternatives, one typically being the status quo, is one of the most commonly used preference elicitation formats. It has a long history of use in survey research, and most other discrete choice and ranking formats can be easily shown to be generalizations of it. Bishop and Heberlein (1979) showed that this format could be used along with the random assignment of respondents to different monetary costs to recover the distribution of willingness to pay or willingness to accept compensation (WTA). Later papers by Hanemann (1984a, 1984b) formally worked out the utility theoretic approach from a random utility perspective (McFadden, 1974); and Cameron (1988) provided a purely statistical approach of tracing out the latent (unobserved) WTP or WTA variable in a manner similar to dose response experiments in biology or medicine. McConnell (1990), Kriström (1997), Haab and McConnell (1997; 1998) and Hanemann and Kanninen (1999) provide comprehensive examinations of the statistical issues involved in using the binary discrete choice format. While we will generally ignore the substantive issues raised in these papers with respect to the estimation process, we do note some of the implausible estimates that in the literature appear to be the result of failing to adequately model the data and incorporate sensible restrictions implied by economic theory.

Much of the attention focused on the binary discrete choice elicitation format in recent years is due to the NOAA Panel on Contingent Valuation's (Arrow *et al.*, 1993) recommendation for its use as a consequence of its well-known property of being incentive compatible in some circumstances. Indeed, one of the core results in mechanism design theory independently derived by Gibbard (1973) and Satterthwaite (1975) is that no response format that allows for more than a binary response can be incentive compatible without assuming restrictions on the realm of allowable agent preferences.

However, the Gibbard-Satterthwaite result is essentially a negative one—no response format with greater than a binary choice (including all multinomial and continuous response formats) can be incentive compatible without restrictions on preferences. This result does not say that *all* or *even any* binary discrete choice formats are incentive compatible; only that this is the only response format that is potentially incentive compatible.

It has long been known that in some settings that the binary discrete choice format is incentive compatible (Farquharson, 1969). The best-known examples are political races with only two candidates and binding (approve/disapprove) referendums with a plurality (usually majority or two-thirds approval) vote requirement. The binding referendum is a useful departure point for our discussion and the NOAA Panel references this mechanism before their recommendation to use a binary discrete choice format in contingent valuation (CV) surveys.

The first question is whether it is the binding nature of the referendum that makes it incentive compatible. Carson, Groves, and Machina (1997) consider an *advisory referendum* vote.[8] They show that replacing the binding plurality vote requirement with the weaker assumption that, over some range, the government is more likely to undertake an action the larger the percentage in favor.[9] The plurality vote requirement is a special case of this assumption with the knifed-edged decision rule that any vote less than the required plurality for the new ("yes") alternative results in the default ("no") alternative being implemented.

The second question is: does substituting an *advisory survey* for an advisory referendum alters the incentive properties of the mechanism? Green and Laffont (1978) have shown that any economic mechanism of the types being considered in this paper can be implemented using a sampling approach rather than complete participation. Thus, we come to the following:

> **Result:** It is possible to replace the binding nature of an incentive compatible referendum with the more general assumption that the agency is more likely to undertake the action the higher percent in favor. It is also possible to substitute a survey of the public for a vote of the public on the issue. Neither of these changes, alone or together alter the original incentive structure of the binding referendum.

---

[8] Many well-known referendums are technically advisory referendum. For example, Norway's vote on whether to join the European Union (EU) was an advisory referendum. Some observers believed that if the vote in favor were only a very slim majority, that the government would not join the EU, however, if a substantial majority favored joining then the government would join the EU.

[9] It is necessary to assume that agents believe they have only influence locally around the amount they are asked if this response function is considered to cover the case where the amounts agreed to are summed. We are indebted to Pere Riera for this observation.

A small number of CV studies (*e.g.*, Carson, Hanemann, and Mitchell, 1987; Polasky, Gainutdinova and Kerkvliet, 1996), which have compared survey estimates to the vote on actual binding referendums, have found the two to be quite close. A very large body of evidence from polling on referendum suggests that surveys taken close to an election generally provide quite good predictions of actual referendum votes.[10] It is important to note, however, that it is not casting the preference question as a referendum that provides its desirable incentive properties. It is the cast of the preference question in terms of being able to influence a government decision with a binary favor/not favor format.

Two key assumptions have been made in the discussion of the preceding sequence of mechanisms. The first assumption is that the agency (*i.e.*, government) can compel payment for a good if provided. The second assumption is that only a single issue is involved. Relaxing the first assumption destroys the incentive properties of what we will call the referendum—advisory referendum—advisory survey (RARAS) mechanism. To see this, consider the case where a charitable organization wants to provide a public good via voluntary contributions. A "yes" response to a binary discrete choice survey question of the form: "would you contribute $X to a fund to purchase the specified good if we started the fund?" will encourage the charitable organization to undertake the fundraising effort. Upon mounting the fundraising effort, the optimal strategic response of an agent who wants the public good will be to contribute less than her maximum willingness to pay for the good and in many instances to contribute nothing.[11] The preferred strategy is to sit back and wait to see if the good is provided without her contribution. This is the classic free riding behavior which economists have long seen as perhaps the fundamental problem with the provision of public goods. What is interesting in this case is that the same incentive structure which should cause free riding with respect to the actual contributions should induce respondents in a survey to over pledge because doing so helps to obtain the later opportunity to free ride. A number of empirical studies confirm the large predicted divergence between survey-based predictions of contributions and actual contributions (*e.g.*, Seip and Strand, 1992; Champ *et al.*, 1997).

Switching to the case of introducing a new private good does not improve the incentive situation. As long as there is any positive probability of wanting the new good at the stated price, the respondent should say, "yes—would purchase." The agent's logic is that such a response will encourage the company to produce the good, with the agent being able to decide later whether to purchase. Since increasing the agent's choice set in a desirable way increases utility, the optimal response is "yes." Folk wisdom from the marketing research literature supports the notion that consumers overstate their purchase proclivities for new products (Greenhalgh, 1986). Evidence

[10] Predicting an actual election vote from a survey involves two key difficulties unrelated to whether agents truthfully reveal their preferences in surveys. The first is that the information set the voter uses on election day may have changed from the one at the time of the survey due to activities such as political advertising and media coverage. It is this factor that makes surveys taken close to an election generally more accurate than surveys taken at some distance from the election. (The dynamics of the information process are such that the proponents of the measure are usually able to initially put out a largely unopposed positive message. As opponents slowly start their negative campaign, support for the measure falls over time.) The second is predicting who is going to actually vote. The characteristics of a good random sample of the public may be substantially different from the characteristics of the sample of the public that actually votes.

[11] In many charitable fundraising efforts, the quantity of the good provided is increasing in the amount of money raised. In such a case, it may be optimal for a (non-pivotal) agent who desires the good to contribute at a positive amount toward its provision (Blume, Bergstrom and Varian, 1986).

from experiments in economics (Cummings, Harrison, and Rustöm, 1995; Johannesson, Liljas, and Johansson, 1998) also supports this conclusion. The marketing research approach has tended to either shift to a different measurement scale such as the probability of purchasing (Inforsino, 1986) or to ask about more than one good (Louviere, 1994).

There is some irony in this result as it has so often been said that if standard CV elicitation formats did not work well for private goods then they would not work for pure public goods that are not bought and sold in the marketplace. The familiarity argument that is so often heard in support of doing experiments with private goods to learn about how CV is likely to work in the best case scenario (Neil *et al.,* 1994) is misguided. Examined in this light, the introduction of a new private good is one of the worst-case scenarios for a binary discrete choice question. It should not be surprising that the binary discrete choice format, which while initially seeing usage in marketing research, is now rarely used.

The ability of the agency to coercively collect payment for the good is the property that causes the agent to try to influence the agency's decision in the desired direction taking account of both the cost and the benefits of the action to the agent.[12] Voluntary contributions allow for the possibility that the survey response encourages the fund-raising effort to be undertaken, and hence, the possibility of free riding during the actual fund-raising effort. Thus, agents who want the good provided should say "yes" (would contribute) to the survey. In turn, it will be optimal for some of these agents to free ride in the expectation that other agents would contribute enough to provide the good. In this case, an initial survey "yes" response helps to set up the later opportunity to free ride with respect to the actual contribution. For the private goods case, a "yes" response (would purchase) to the survey encourages the production of the good while the agent gets to decide later whether to purchase the good. Thus, if the agent anticipates any positive probability of wanting to purchase the good, then a "yes" response is optimal. If the agent anticipates that the good will be offered irrespective of the responses given by agents but the agent perceives that the responses may influence the price of the good, then it is optimal for the agent to appear more price sensitive than is actually the case. This result is often seen in marketing research where agents have been found to more price elastic in surveys than in actual market purchases. The only problem with these cases from the perspective of economic theory is not whether there should be a divergence between actual behavior and the survey estimate, but rather, whether the magnitude of the divergences empirically observed should be even larger.

There are other interesting implications of the lack of incentive compatibility for binary discrete choice survey questions for voluntary contributions and the introduction of new private

[12] It is interesting to ask whether it is the two-step nature of a survey followed by a contribution/purchase that leads to the survey question not being incentive compatible. The answer is no. Consider the situation whereby the only way a public good can be provided is if it obtains the requisite plurality vote in a referendum and the legislature gets to decide whether to put the issue on the ballot for a vote. The legislature does not want to waste the public's time putting on propositions to vote on if they stand little chance of passing. The legislature (or the measure supporters) commissions a survey to determine the likely fraction of the public that would vote in favor of the measure. The only consistent responses (given no change in the information set) to the survey and actual referendum vote are "yes" to both the survey and the referendum or "no" to both the survey and referendum. For those in favor of the measure, the only way to get the good is to get the referendum put on the ballot and have the measure passed. "Yes" responses to both opportunities increase the chance of both. For those opposed to the measure, saying "yes" to the survey increases the chance that it will get put on the ballot, which in turn increases the chance that the agent will have to pay for the good, even though the good is not worth the cost to the agent if provided.

goods with respect to other anomalies such as insensitivity to the scope of the goods being valued. For instance, as long as the good is potentially desirable it is optimal to say "yes" to the survey question. The scope of the good and its cost do not influence this decision unless the good becomes so small that even if at a zero cost it is not desired or if the cost becomes so high that it would never be purchased. In both of these later instances, either a "yes" or a "no" response by the agent will have the same effect on the agent's utility.

If the binary choice is between two different forms of a quasi-public or private good, then desirable incentive properties can be restored as long as only potential users are interviewed.[13] To see this, consider the classic case of a campsite. At present the campsite is unimproved and has a low entrance fee (possibly $0). The alternative is to improve the campsite and increase the entrance fee. The agent should now choose the status quo campsite price/quality combination or the alternative campsite price/quantity combination to maximize utility. This binary choice can be shown to have identical properties to the RARAS survey mechanism. The property that the mechanism needs to be incentive compatible is the ability of the agency to force one of the alternatives on a particular agent irrespective of that agent's preferences in a situation where the agent's utility is influenced by the agency decision. Two important caveats should be kept in mind. First, in this situation the total number of times the good will be used under the alternative is endogenous. In our campsite example, if the higher quality-price campsite alternative provides more utility than the status quo, the anticipated number of visits to that campsite under that alternative may be larger or smaller than under the status quo. Second, for agents whose probability of use of the good does not differ between the two configurations, any response has the same impact on the agent's utility. This problem is not usually seen because most recreational surveys are either done on site or from lists of users. Marketing researchers typically screen out non-users of a product class before asking preference questions.[14] The risk in both instances is that focusing on current users of the good will miss those who would likely use the good if its quality/price attributes were changed.

This choice between two configurations of a good works for public goods and private goods too, irrespective of the nature of the payment obligation, as long as the agent desires the good at no cost. To see this, consider a private charity that wanted to build one of two different monuments in the center of town. The charity conducts a survey of the public to determine which monument is preferred and the higher the level of support for a particular monument the more likely that monument will be built. The agent should pick the preferred monument since this increases the agent's utility more than the alternative monument and neither monument imposes any cost on the agent. Our favorite example of a private good question is the bar owner that surveys patrons and asks whether they would prefer to have the bar's sole draft beer, currently a domestic brand priced at $1, switched to an imported brand at $2. The bar patron should pick the import only if having that

[13] Quasi-public goods are those provided by the government but for which it is possible to exclude members of the public from using. This exclusion can occur in terms of charging a price to use the resource, having the agent spend money or time to use the resource or by having the resource effectively bundled as an attribute of a privately purchased product. Common examples include government campgrounds and houses located on public lakes.

[14] There are exceptions. *Boxall, et al.* (1996), for instance, ask hunters in Alberta about two different management/cost regimes for a specific area that few currently hunted in and few were likely to hunt in with the alternative management scheme. In this instance, the contingent valuation estimate was dramatically larger than the travel cost estimate, something that is fairly unusual in comparisons between the two approaches for quasi-public goods (Carson *et al.*, 1996). When the estimate of the change in the probability of use is used to scale the CV estimate, the two approaches result in quite similar estimates.

alternative available provides more utility than the domestic. Note that the number of beers that will be purchased is not revealed by the agent's choice and could go up or down.

Table 1 summarizes the incentive properties of binary discrete choice questions by the type of good and the payment characteristics. In this case we have assumed that the agent would desire the good if there was no cost, otherwise the incentive properties of the question are undefined. What is striking is that anomalies with respect to a divergence between estimates based on stated preferences and estimates based on behavior are heavily concentrated in the two cases that are not incentive compatible.

Table 1: Incentive Properties of Binary Discrete Choice Questions

Type of Good	Incentive Property
New public good with coercive payment	Incentive compatible
New public good with voluntary payment	Not incentive compatible
Introduction of new private or quasi-public good	Not incentive compatible
Choice between which of two new public goods to provide	Incentive compatible
Change in an existing private or quasi-public good	Incentive compatible but choice does not reveal information about quantities

The second key assumption in the discussion of the RARAS mechanism is that is that of a single up-down vote on a single issue. It is also not possible to relax this condition and there are several common instances where it is violated. The best-known ones are the rules for school bond referendums in many areas (Romer and Rosenthal, 1978; Lankford 1985). The school board gets to propose the level of educational inputs and the tax rate. However, if the referendum is voted down, the school board can only bring up another referendum measure with a level of educational inputs and a tax rate that is lower than those voted down but higher than the default status quo. A respondent who prefers the initially offered bundle to the status quo may nonetheless have an incentive to vote against it in order to gain opportunity to vote in favor of an even more preferred provision/tax package. With respect to valuation of an environmental project, Richer (1995) shows that his CV WTP estimates are influenced by information about whether a different alternative plan for a national park in California's Mojave Desert was likely to be put forth if the current plan described in the survey was not approved. Another variant is where there is another party (*e.g.*, another government agency or private entity) who potentially can provide the good.[15] The general

[15] This problem appears to have influenced the Cummings *et al.* (1997) results. In that experiment, agents are randomly assigned to a "hypothetical" treatment and a "real" treatment in which the group votes on whether to contribute a specified amount per agent to provide the good. The estimate based upon the hypothetical treatment is higher than that of the real treatment, although Haab, Huang, and Whitehead (1999) show judgment of the significance of the difference depends upon how the larger variance in the "hypothetical" treatment is taken into account. We believe that to many of the agents interviewed in Georgia, the Cummings *et al.* hypothetical treatment should have appeared as an attempt to determine whether it was possible to mount a fundraising effort to provide printed information booklets on toxic hazards to poor people in New Mexico. As such, we would have expected the "hypothetical" treatment WTP to be

principle is that direct linkage between a decision on one issue and a decision on another issue can cause difficulty in interpreting the result, as the optimal response of the agent should generally take the sequence of decisions and options into account.

There is a further condition that is important for the interpretation of the results but not for the incentive properties of the RARAS mechanism. The agent needs to believe that if the agency implements a particular alternative: the specified good, Q, will be provided and the stated price, P, assessed. If instead the agent believes that Q* will be delivered and P* paid if this alternative is chosen by the agency, then the agent's optimal response should be based upon (Q*, P*) not the stated (Q, P). Note this condition holds for interpreting actual votes or actual consumer purchases as well as responses to preference survey questions.[16] An important implication of this condition though is if the goods and prices used in a preference survey go beyond what the agent finds plausible, the preference survey question is likely to be answered on the basis of the expected good and the expected price rather than the stated ones.

Introduction of Cost Uncertainty

Binary discrete choice preference surveys often provide a cost (in monetary or other terms) for each alternative and this cost information plays a key role in estimating welfare measures. What role should agent uncertainty over cost play in the answers given? The answer is obvious if the survey provides a cost estimate of $X and the agent thinks that since the government has a proclivity for cost overruns that the actual cost will be double the stated cost. The analysis should be performed with the cost as perceived by the agent.

The more interesting case is when the agent takes the survey and provides $X as the expected value with some type of distribution around $X. Here the key issues can be seen to revolve around whether the original status quo choice set will still be available and whether a commitment to the pay for the good is required ex ante before the cost uncertainty is resolved. These two conditions determine whether shifts from an original "yes" to a "no" and vice versa are possible given a mean preserving increase in cost uncertainty. Table 2 displays the possible outcomes.

---

higher than true WTP. However, uncertainty about why agents in Georgia should be asked about voluntary contributions to a New Mexico program may have lead to the larger variance found by Haab, Huang, and Whitehead (forthcoming). For the "real" treatment we would have expected an under-estimate of true WTP due to the possibility of having some other group pay to distribute the already printed booklets. A later experiment by Cummings and Osborne-Taylor (1998) effectively replicates this experiment but with additional treatments where there are different probabilities that the vote taken by the group is binding. The WTP estimate decreases from the "hypothetical" treatment to the "real" treatment as the probability that the group vote is binding goes from 0 to 1. This is the result that our model predicts if all treatments were perceived by agents as being consequential and that there are competing incentives to over pledge and free ride in all of the probabilistic treatments. The result that would be predicted theoretically if there was no incentive to over pledge in the "hypothetical" treatment and free ride in the "real" treatment would be that all of the treatments with a positive probability of the vote being binding should result in similar WTP estimates.

[16] Carson *et al.* (1994) show, for instance, in a recent CV study in California that respondents who do not currently pay taxes are willing to pay more than respondents with otherwise identical characteristics. That respondents who believe that the state government would assess the one time tax in multiple years are willing to pay less than respondents who think the fee will only be applied one time and that respondents who don't think that the plan will work completely are willing to pay less than those who think that it will work. See Randall (1994) for a discussion of this issue in the context of the travel cost model. There are large literatures in marketing and political science dealing with what are effectively the P's and Q's perceived by agents when they make decisions.

Table 2: Effect of Increased Cost Uncertainty upon Binary Choice

	Ex ante choice (i.e., commitment)	Ex post choice (i.e., no commitment)
Status Quo still available	**Can only shift Yes No**	**Can only shift No Yes**
Status Quo no longer available	**Can only shift Yes No**	**Can shift either Yes No or No Yes**

For the case of provision of a public good with a coercive payment mechanism, the status quo choice set is still available but one has to commit *ex ante* to paying the uncertain cost. This commitment translates into income uncertainty and hence is never preferred by risk adverse agents. Hence one would expect to see some shifts from "yes" to "no" responses. There should be no shifts in the opposite direction, so that the aggregate change is a decline in standard statistics of the WTP distribution like the mean and median relative to the case with no cost uncertainty. The other case where an *ex ante* commitment is required has the same result but may be of less practical relevance since most examples here require an ex ante commitment to purchase a fixed quantity of the alternative to the status quo before the actual cost of the alternative was observed.

The opposite phenomena, possible shifts from original an "no" to "yes" response with increases in cost uncertainty, should occur in the case where the choice can be made *ex post* after the cost is observed and the status quo choice set is always still available. The main examples of this case are provision of a pubic good via voluntary contributions and the introduction of a new private good. The basic logic in this case is that since the status quo choice set will still be available, all agents will either favor or be indifferent the addition of the new alternative. Increasing the level uncertainty can cause some agents who were indifferent to the addition of the alternative to the choice set to favor it. Changes the a "yes" to a "no" response cannot occur, even though it is possible that an increase in cost uncertainty can make some agents, who were already in favor, worse off.

The last of the four cases occurs where the only *ex post* commitment is required and the original status quo choice set will no longer exist, if the alternative to the status quo is provided. The main examples here are quasi-public goods and private goods where only one of two possible configurations of the good will be offered (*e.g.,* a low quality-low price recreation site could be transformed into is a high quality-higher priced version of the site) In this case, it obviously possible for increasing the degree of cost uncertainty to result in both shifts from "yes" to "no" and "no" to "yes."

There are a number of other informational issues that we do not explore here except to note that a formal analysis of the role of different types of uncertainty is likely to be more productive than

the all too frequently invoked vague concept of agent unfamiliarity with a good as a justification for all types of apparent aberrant behavior. Much of the richness of economic theory in recent years has come from the introduction of different types of uncertainty and asking how agents should optimize in the face of it (Varian, 1992). Particularly, relevant here is the rapidly developing literature on how agents process information in candidate and referendum elections (*e.g.*, Popkins, 1991; Lupia, 1994). This literature suggests ways in which agents make reasonably informed decisions based on imperfect information. Further, simply providing more information does not necessarily lead agents to make the decisions closer to that which they would make if fully informed (Lohmann, 1994).[17] This suggests that the informational content of a survey used for environmental valuation should be examined to see if agents were given a reasonably complete, comprehensible, and balance presentation of the alternatives offered.

Double-bounded Discrete Choice Questions

The inherent problem with a binary discrete choice question is the limited information the response to it provides about the agent's preferences.[18] Double-bounded discrete choice estimators have become popular in the environmental valuation literature because they tend to dramatically shrink the confidence intervals around point estimates of statistics of the willingness to pay distribution. The approach is straightforward. If the agent said "yes" to the initial cost amount asked, then ask the same question at a pre-chosen higher amount, and if the agent said "no" to the initial amount, ask the same question at a lower amount.[19] The initial presentations of the double-bounded format relied on double sampling/interval censoring statistical models (Carson, 1985; Carson and Steinberg, 1990; Hanemann, Loomis, and Kanninen, 1991). They assumed that agents have a single latent WTP value and that the responses to both the first and the second questions are based upon simply comparing this latent WTP value to the cost amount asked about in each question. Statistically, the implication of this assumption is that, with appropriate conditioning, there is perfect correlation between the WTP distributions implied by the responses to the two questions.

---

[17] For example, consider an agent who initially favored a project and saw both its benefits and costs as being small. The agent, if fully informed, would still favor the project but realized that both its benefits and its costs were large. The agent informed that the cost of the project is large, but not given the corresponding benefit information, will now oppose the project. Much advertising in marketing and political campaigns operates on this notion of providing selective "half-truths."

[18] The only information provided is whether the agent's WTP for the good is higher or lower than the single amount asked about in the survey question. It is possible to use parametric assumptions about the underlying WTP distribution to effectively overcome this sparse information, but such assumptions can play a large role in the estimates derived. Non-parametric approaches to the use of binary discrete choice data (*e.g.*, Kriström, 1990) exist that make the power of these assumptions abundantly clear.

[19] In some respects, the double bound model is similar to the iterative bidding game approach used in the early CV literature (Randall, Ives, and Eastman, 1974) that was usually found to suffer from a phenomena known as starting point bias whereby the amount initially provided the agent influences the agent's final WTP amount. There are some key differences though which make the two approaches fundamentally different. The initial cost amount in the iterative bidding game was never intended to reveal information about the goods actual cost and the iterative steps from that amount are usually quite small. In contrast, the statistical tools used to analyze data from both the binary discrete choice and the double bound discrete choice formats exploit the agent's conditioning on the cost number explicitly provided and the interval formed by the first and second price is fairly large. Most good studies using a double-bounded format go to some effort to provide a rationale to the agent as to why the cost number used in the second question is different from that of the first. An interesting variation on the double-bounded format is a single binary discrete choice format with a follow-up open-ended question. Farmer and Randall (1996) analyze this format from a theoretical and empirical perspective and show results similar to those described here for the double-double-bounded estimator: the second responses tend to be biased downward.

Following Cameron and Quiggan's (1994) pioneering examination of this assumption, several stylized facts have emerged concerning the comparison of the WTP estimates based on the first binary discrete choice question and both binary discrete questions: (a) the WTP distributions implied by the first and second questions are not perfectly correlated, (b) the WTP estimate based upon just the first estimate is higher than the WTP estimate based upon both questions, and (c) the number of negative responses to the second question is higher than would be expected based upon the WTP distribution estimate from the first question alone. Alberini, Kanninen, and Carson (1997) have put forth a general error-components model, and McLeod and Bergland (1999) have put forth a Bayesian preference-updating model to handle these issues.

What sort of effects should the asking of a second binary discrete choice question have on the latent WTP distribution? The key property of this format from our perspective is that the agent has been told that the same Q was available at two different prices. The best-case scenario here is that the agent takes the second price as the expected price but now considers the price to have some uncertainty surrounding it.[20] Consistent with the discussion in the previous section, statistics such as mean or median WTP will be shifted downward in the second question for risk adverse agents and public goods even though preferences for it have not changed.

There are, however, several other plausible alternatives for what the act of asking the second price should signal to agents. One of these is that the agency is willing, in some sense, to bargain over the price. For agents who originally answered "no" and got asked a lower price, the optimal response may be to answer "no" again in hopes of getting offered an even lower price.[21] This should result in the second WTP response being "no" for some of these agents even though had this amount been asked at the first question the response would have been "yes." A similar effect can be found with respect to those whose original answer was "yes". Since the good was originally offered at a lower price, it can presumably be provided with some positive probability at the initial price. As such, some agents will find it in their self-interest to risk not getting the good by holding to the lower price and saying "no" to the second higher price, even though the agent's WTP for the good exceeds the second price. The effect of this type of behavior would be to shift the WTP distribution implied by the second question to the left, and hence, reduce estimates of mean and median WTP.

Another plausible assumption is that the actual cost to an agent will be some type of weighted average between the two prices. If this assumption is made, the second question should be answered on the basis of this weighted average of the two prices. It is straightforward to see that for an initial "no" response, that any weighted average of the first and second prices is higher than the second price. For an initial "yes" response, any weighted average of the first and second prices is lower than the second price.[22]

---

[20] Alternatively, if the agent thought the first price had some uncertainty surrounding it, asking the second price should increase the original level of uncertainty since for the double-bounded estimator the first and second prices are typically fairly far apart.

[21] It is of course possible to expand the double-bounded concept to asking a third question. See Bateman *et al.* (1995) for an example.

[22] Note that this assumption is not inconsistent with the arguments concerning uncertainty and the two may be combined. For initial "no" responses, this effect of adding uncertainty is reinforcing in a downward direction. For initial "yes" responses, the effect is in the opposite direction and mitigates the upward effect of price averaging.

The last plausible assumption we consider is that the agent might interpret the signal given by the second price as implying that the quantity has changed to match the changed price in a consistent manner. For an initial "no" response, the shift in quantity that is consistent with the reduction in price is to reduce the perceived quantity/quality of the good that would be provided. The implication of this is to shift the WTP distribution implied by the second response to the right for these respondents. This is a commonly voiced concern in focus groups and debriefing questions. For agents who initially said "yes", the shift in perceived quantity is upward. There does not appear to be any collaborating evidence to support the proposition that this is a common phenomenon.

What should be grasped from this discussion is that to a rational agent the second price must signal that something is going on. All of the plausible assumptions lead to the correlation between the WTP distributions implied by the two questions being less than 1. All of these assumptions also shift the WTP distribution implied by the second question to the right for agents who initially gave a "no" response, and hence, produce an "excess" number of no-no responses. For agents initially giving a "yes" response, it is possible for the WTP distribution implied by the second question to shift either to the left or the right. However, only the price averaging assumption has much credence in terms of the possibility of producing an upward shift in the standard WTP statistics. On balance, we would expect that WTP estimates from a double-bounded format to be smaller than those from a single-bounded format. All of these hypotheses tend to be strongly supported by the empirical evidence. It may be desirable to use the double-bounded format in CV studies; however, this desirability rests on the analyst's tradeoff between the likely downward bias and the tighter confidence interval (Alberini, 1995).

Continuous Response Formats

Ideally one would like to have the agent's actual WTP or WTA, not a discrete indicator of it. So it is not surprising that many early CV studies used an open-ended direct question.[23] Many economists thought that these early efforts would fail because agents would give the extremely high WTP answers. This did not happen (*e.g.*, Brookshire, Ives, and Schulze, 1976), and interest in the survey based valuation methods grew in part due to this anomaly.

The early problem that researchers did find with the direct question was that agents always wanted to know what the project would cost them. Agents did not understand why they were not provided the cost information if the agency had worked out the details of how the good would be provided. Further, many agents appeared to have great difficulty formulating a (continuous) WTP response. This led to very high non-response rates and a large number of so-called "protest zeros" which were typically dropped from the analysis. This lead to speculation that survey respondents did not have "well-defined" preferences in an economic sense.

Three different directions were tried to overcome this problem. The binary discrete choice format (Bishop and Heberlein, 1979) discussed earlier gets around one of the key problems by giving agents the cost number they want and then uses a statistical analysis that "appropriately" conditions on agents reacting (favor/not favor) to that cost number. The earlier iterative bidding game method suggested an initial amount and iterates up or down from that amount in small increments (Randall, Ives, and Eastman, 1974). The payment card approach asks agents to pick a

[23] The continuous response format is known as a matching question in the psychology literature and is a special type of open-ended question in the survey research literature.

number (or any number in between) on a card (Mitchell and Carson, 1986; Cameron and Huppert, 1991). The latter two methods can come close to achieving a WTP response in continuous terms; and, except when these formats have special properties, the discussion of the continuous response format will apply to these formats as well.

With different elicitation formats came the inevitable urge to compare their results (*e.g.*, Smith and Desvousges, 1986). Researchers were dismayed to find that different response formats lead to different WTP estimates and the divergence between these estimates is frequently cited as one of the major reasons why estimates based on stated preference questions should be rejected (Hausman, 1993; McFadden, 1994).[24] The stylized fact here is that discrete choice formats produce higher WTP estimates than do continuous response formats (*e.g.*, Boyle *et al*., 1996).

Should the divergence in estimates from different formats be surprising?[25] No. Given the Gibbard-Satterthwaite result, it is impossible to formulate a continuous response question that has the same incentive and informational properties as an incentive-compatible binary discrete-choice question. Many researchers looking at the results, however, have been misled by the face-value dilemma. The divergence between the estimates from the different formats suggested that either agents were not truthfully revealing their preferences to one or more of the elicitation formats or that they did not have well-defined preferences in the sense used by economists.

As noted earlier in this discussion, the expectation of many economists was that most agents would provide very large WTP responses when asked an open-ended WTP question if agents were acting strategically but not truthfully revealing their preferences. However, the opposite phenomenon was observed: estimates from binary discrete-choice questions were higher than those from continuous response CV questions and continuous response CV questions contained lots of zero responses.

Faced with an open-ended question, a very large WTP response does turn out to be the optimal strategy for an agent who believes (a) the cost of the public good to the agent is fixed, (b) her true willingness to pay for the good is larger than the cost if provided, and (c) the good is more likely to be supplied the larger the sum of the willingness to pay responses given by agents. Note that only the subset of agents whose WTP is greater than their cost should be giving a positive WTP response, so one should never have expected all agents to engage in this behavior.

Condition (c) corresponds to the benefit-cost criteria, but it is hard to find a single instance where an agency decision has been made based purely on that criteria. There is little evidence to suggest that agents believe that the agency is simply summing their WTP responses. As such, we believe it useful to consider a variety of other beliefs that agents may hold.

---

[24] The irony in this position is that estimates of other economic quantities based upon substantially different econometric techniques have typically differed even though data on actual behavior was being used. The usually recommended approach in this situation has not been to discard economic theory and econometric methods but rather to understand the source of the differences.

[25] From the critique by cognitive psychologists, the divergence between framing provided by the (binary discrete) choice and open-ended matching question is at the heart of problems with microeconomic theory (Tversky, Slovic, and Kahneman, 1990).

Let's first consider the optimal response of an agent whose perceived cost of the public good is greater than the agent's willingness to pay. Maintaining the previous assumptions, this agent's optimal response is "zero". This result turns out to be fairly robust to plausible alternatives to (c) that we discuss below and, as such, may help to explain the large number of zero responses received to open-ended type questions. The intuition behind this result is that the agent's utility is reduced if the public good is provided and the cost assessed against the agent. The response that adds the least amount to the sum of the benefits (given the usual non-negativity constraint in the open-ended format) is "zero."

Step back for a moment from the benefit-cost criteria that has dominated economic thinking on the incentive structure of the open-ended question and recognize that the simple act of asking an open-ended question is likely to signal to agents that the cost allocation among agents for providing the good is not fixed. Once the agency is prepared to shift the vector of costs facing agents, changing condition (a) above, toward increasing the cost to agents having (relatively) high WTP for the good and decreasing it to those who do not, the incentives for agents whose WTP is greater than the initially perceived cost change substantially. These agents now have to balance the increased probability that the good will be supplied with a high WTP response against the potential upward shift in the cost they will pay if the good is provided. For agents having WTP less than the initially perceived cost, the optimal response is still zero.

Since the government rarely if ever uses a pure benefit-cost criteria, it may be plausible for agents to assume that the agency is simply trying to determine what percentage of the relevant population has a WTP higher than the cost which may or may not be assumed to be known to the agency at the time of the survey. Combined with the potential to reallocate the cost burden, the optimal response of an agent whose WTP is greater than the initially perceived cost is now equal to the cost while the optimal response of an agent whose WTP is less than the initially perceived cost the optimal response is still zero.

In all of these cases, the optimal response depends strongly on the agent's perception of the agency's cost of providing the good. The agent should first compare her actual WTP to the expected cost. The optimal response for agents whose WTP is less than the perceived cost, under most plausible uses of the information provide, is zero. Such an agent should additionally "protest" in any other way possible, as the change from the status quo will negatively impact the agent's utility. The optimal response for an agent whose actual WTP is greater than expected cost depends upon her belief about how the agency will use the stated WTP. In this case, the optimal response typically is further conditioned on expected cost. The difficulty in interpreting the positive WTP response is that different agents may have different beliefs about agency use of the response.

Agents, however, don't know the cost with certainty. They can formulate priors about the cost and should incorporate any information provided in the survey that they believe is related to cost. This type of behavior would give rise to starting point bias in iterative bidding games (Boyle, Bishop, and Welsh, 1985) and range/placement effects in studies using payments cards (Rowe, Schulze, and Breffle, 1996) to the extent that agents think that the "extra" information provided in these formats is correlated with cost.

On occasion, a variety of different open-ended formats that are said to be incentive compatible are used in a survey context such as the Becker-DeGroot-Marshack mechanism or the Vickery auction.[26] Both of these mechanisms elicit a continuous WTP response. There are two things to remember about such mechanisms. First, they do not get around the Gibbard-Satterthwaite result. Holt (1986) and Karni and Safra (1987) (hereafter HKS) independently showed such mechanisms depend crucially on the preferences obeying the expected utility assumption. Many researchers are willing to maintain the expected utility assumption, and many key economic results on risk are locally robust to most non-expected utility alternatives (Machina, 1995). However, when trying to implement either of these two mechanisms in a survey context, the difficulty lies much deeper. Both of these mechanisms rely on the ability to condition the agent's response on an "exogenous" random element. We have shown that it is impossible to formulate a simple open-ended matching question that is informationally and strategically equivalent to an incentive compatible binary discrete choice question in a survey context. This result is a companion of the HKS theorem. To make the matching question equivalent strategically to the binary discrete choice, the agency has to pre-commit either to the cost or to an exogenous device to provide the cost. Doing so prevents the agency from exploiting the extra information that the agent provides in the matching format but not in the choice format. To get the agent to reveal the matching answer, the agent cannot know the cost. The need for the agent's uncertainty about the cost puts one back in the HKS world where expected utility is required. The need for agency pre-commitment not to exploit the extra information contained in the continuous WTP response effectively prevents its being used in a survey context.

Sequence of Paired Comparisons

In addition to wanting tighter confidence intervals on the WTP distribution for a single good, decision makers often want information on the WTP distributions for a variety of related but different goods so that they can pick the best option. There are two popular approaches in the literature for doing this. The first is to offer agents a sequence of paired comparisons. The second is to ask agents to pick between or rank order a set of $k > 2$ alternatives. Here we discuss the strategic issues that arise with these choice formats and do not deal with issues related to the adequacy of the information set on each distinct good provided in the survey.

In an ideal world in which the objective involves valuing public goods, the agent treats each paired comparison independently and the desirable properties of a single binary discrete choice question with a coercive payment requirement can be repeatedly invoked. There is a very simple question, however, that illustrates the fundamental difficulty with a sequence of paired comparisons. Consider the case of air pollution levels in a city. The agent is asked to pick between different pairs of air pollution levels that involve different costs and different health effects and visibility levels. Since air pollution in the city is a public good, however, all agents will eventually face the same air pollution level. If *k* different air pollution levels are described to the agent in the course of the sequence of paired comparison, the agency must have some method of choosing among the *k* different levels. Any particular method that the agent perceives that the agency is using to incorporate agent preferences into its choice of an air pollution level generally will provides incentive for non-truthful preference revelation. In some instances, it will even be optimal for the agent to reject his or her most preferred level (out of the *k)* in a particular paired comparison. Once this is

[26] Other mechanisms eliciting a continuous response like the Groves mechanism (Groves, 1973) require stronger restrictions on preferences (*e.g.*, quasi-linearity in income) and the possibility of side payments.

possible, the standard methods of inferring value from choices no longer work. The essential problem is that an agent's optimal choice depends both upon the agent's preferences, expectations about what the other agents will do, and the perceived rule for aggregating the results of each paired comparison. This result has long been established in the literature on the properties of voting rules (Moulin, 1994).

With quasi-public and private goods, the difficulties noted for public goods still exist, with the exception that it may be possible for more than one of the *k* goods to be provided. This possibility tends to reduce the likelihood that an agent will make a choice that is not her favorite; and in the next section on multinomial choice, we discuss the aggregation issue further.[27]

Multinomial Choice Questions

Many of the issues raised in the previous section on a sequence of paired comparisons are relevant to multinomial choice questions. The strategic issue that an agent faces when answering a multinomial choice question (pick the most preferred out of $k > 2$ alternatives) is how the agency translates the responses into actions. The simplest case consists of generalizing the decision rule used in the binary discrete choice format by assuming that the agency will provide only one of the *k* goods and that the higher the percentage of the sample picking any particular alternative, the more likely that alternative will be provided. The well-known result from the voting literature on multi-candidate races with a simple plurality winner is that the race from an agent's strategic perspective reduces to a binary choice between the two alternatives that the agent believes will receive the largest votes independent of the agent's vote. The rationale behind this result is straightforward: only the top two alternatives have a chance of winning; picking the most preferred alternative among these two will maximize the utility of the agent's final outcome.[28] The agent is truthfully revealing her preferences, but such truthful preference revelation is, as it should be, conditional on the expectations about the choices of the other agents. However, the agent is not answering the question of interest to the analyst. It will be optimal in many instances for the agent to pick an alternative other than the (unconditionally) most preferred one.

Let us now consider perhaps the opposite case, one of particular relevance to private and quasi-public goods, by changing one of the key assumptions. Now instead of only one of the *k* goods being supplied, let *k-1* of the goods be supplied. To keep matters simple, assume further that the agent only uses at most one of the goods. Examples of such a choice context might be a

[27] There are further issues related to a sequence of paired comparisons that need to be addressed in any particular analysis. The first of these is the strong possibility that the scale term associated with each paired comparison is different. If this is the case then much of the gain in precision and the ability to deal with changes in attributes associated with asking the sequence of paired comparisons may be an illusion. The second is that most rules for combining information from different paired comparisons implicitly require that the independence of irrelevant alternatives assumption to hold. This property is routinely rejected in paired comparison data. The third involves the common use of pairs where both alternatives are off the agent's current utility frontier and neither represents the status quo. This practice requires much stronger assumptions about the nature of the agent's utility function than typically assumed in order to combine the data from different paired comparisons.

[28] With a richer model of agent expectations, it may be optimal for the agent to vote for an alternative that is not one of the top two if there is enough uncertainty over the expected finish of alternatives and the utility differences between the alternatives is large enough. The manifestation of this proposition can be seen in the behavior of fringe political candidates in plurality winner elections. Such candidates try to convince voters that they have a non-trivial chance of winning, that the difference in positions between the two front-runners is extremely small, and that they are much closer to the voter's ideal point.

government agency that had to close four out of five recreational fishing lakes or a computer company that was going to offer four out of five configurations of a particular computer model. In this case, it is optimal for the agent to pick the most preferred alternative out of those offered. Formally, it can be shown that this case collapses to a binary discrete choice of the agent's most preferred alternative against another stochastically chosen alternative. To see this, note that the worst possible outcome for the respondent is that the agent's first choice is not made available. Because all of the other alternatives are provided, the agent's second choice will be available. Effectively, this is a determination of what alternative will not be provided and in pairing the agent's favorite alternative against any of the other alternatives, the agent's optimal response is to pick her most preferred.

The general result we have shown is that, if all but *j* of the alternatives are to be provided, then the alternative chosen by the agent should be one of the agent's *j* favorites. Often the number of alternatives that will be provided is unknown to the agent at the time of making the multinomial choice. A stochastic version of this result has the agent trading off the utility of sets of alternatives with different maximum elements against the agent's prior on *j* and the agent's priors on the choices made by the other agents. Doing so reveals that agents will pick either their (unconditionally without considering the responses of other agents) favorite alternative or close to it, as long as one of three conditions holds: the expectation of *j* is fairly small, the utility difference between the agent's most favorite alternatives and the other alternatives is large, or the prior on the choices by the other agents is fairly uninformative. The implication of this is that agents will appear to make mistakes or optimization errors more often. If they don't pick their favorite, they should pick an alternative close to it.

The statistical manifestation of this type of behavior is a violation of the error term properties associated with the independence of irrelevant alternatives (IIA) assumption. In empirical applications of this elicitation format, the IIA assumption is usually violated. While there are a number of other good reasons for this assumption being violated, such as the rationale behind classic red bus-blue bus problem, it is usually impossible to separately identify the reason for an IIA violation. From a purely statistical viewpoint, it is possible to deal with the problem by introducing one or more scale/variance terms (Swait and Louviere, 1993). This is often sufficient for looking at marginal tradeoffs between attributes. To uniquely recover the latent WTP distribution, it is necessary to have an estimate of *the* correct scale factor.[29] The optimal strategic behavior in this case is often observationally equivalent to direct manipulation of the scale parameter and making recovery of the correct scale factor impossible.[30] Tests for whether data from stated preference surveys and revealed preference observations are consistent with each other and can be combined after (potentially) allowing for a difference in the scale factor (Adamowicz, Louviere and Williams, 1994) are tests against random responses in the stated preference data, not tests against strategic behavior.

[29] This correct scale factor can often be obtained in studies involving quasi-public or private goods from a model estimated on the more limited set of choices current available in the market. This will not typically be the case for public goods.

[30] The scale parameter is typically the negative inverse of the price coefficient. The agent's optimal strategy is to induce the agency to supply the good with the most desired set of attributes at the lowest price. The simple way to do this is to pick the favorite anytime the price is low and otherwise pick something close to it with a low price. Formulated in terms of the expected minimum cost that the agent believes the agency would provide the good at, the agent wants to appear to have an infinite demand elasticity at this cost and to be uninterested above that cost.

With either subadditivity or superadditivity of the utility of the different alternatives and *k-j* (*j* > 1) alternatives to be provided, it is possible to find conditions where the agent should indicate her unconditionally least preferred alternative. The rationale here is that the agent's outcome utility is defined on the set of goods to be provided, not the individual goods taken independently. This is a hopeless situation for learning anything reliable about agent preferences for individual goods.

An alternative to asking agents to pick their single most preferred alternative out of *k*, is to ask them to rank order all *k* alternatives. This exercise could potentially provide considerably more information, but an analysis of the agent's strategic incentives becomes considerably more difficult. The same issue for the agent still exists: how does the agency translate the ranks into a choice of which of the *k* alternatives to provide. Methods for dealing with rank data in a manner consistent with economic theory effectively require the IIA assumption to hold for all possible subsets of the ranked data. This implies that it is possible to explode the data to form sets of multinomial choice questions (Chapman and Staelin, 1982). The IIA assumption can be tested but it does not appear to generally hold for contingent ranking data, and welfare estimates can be substantially impacted if the IIA assumption does not hold (*e.g.*, Hausman and Ruud, 1987).[31]

Concluding Remarks

We have argued that serious consideration should be paid to the incentive and informational properties of preference questions. Much of the difficulty with interpreting the apparent anomalies[32] associated with the estimates based on preference survey questions revolves around what we call the face-value dilemma: either agents always truthfully reveal their preferences to survey question as stated or they do not. This is a false dilemma.

Simple common sense economic models predict large divergences between what agents say they will voluntarily contribute to provide a public good and what they actually contribute. There are now many studies that demonstrate this prediction empirically. The difficulty lies not in the theory or the experimental demonstration but rather in the interpretation that is often placed on these results. Rather than be taken as evidence that respondents don't have well-defined preferences, differences between the estimates obtained using different elicitation formats, if predicted by economic theory, should be taken as evidence supporting the proposition that respondents are taking the scenario posed seriously.

---

[31] A major problem occurs when there are a group of respondents who do not appear to want to trade-off one of the attributes against money. The appearance of such lexicographic preferences can lead to infinite WTP estimates. A subtler problem occurs in that the variance of the error term appears to be substantially larger for "middle" ranks than the most and least preferred alternatives.

[32] The term anomaly is often loosely used. It is possible to have results that represent anomalous behavior from the perspective of economic theory and it is also possible to have such behavior occur in a survey. The most interesting anomalies from the perspective of this paper are those that only occur in surveys. The first step to take with such an anomaly is to see if it can be observed in settings not involving surveys. A number of anomalies first alleged to be survey specific have been shown to be easily replicable in experimental contexts and examples readily identifiable in common market transactions. These include preference reversals (Grether and Plott, 1979), large divergences between WTP and WTA (Bishop and Heberlein, 1990), and part-whole bias (Bateman *et al*., 1997). In some of these instances, such as the often-noted WTP-WTA divergence, models predicting such divergences consistent with standard neoclassical economic theory have been proposed (Hanemann, 1990).

Divergences between binary discrete choice and double-bounded formats or between binary discrete choice and open-ended formats are likewise consistent with theory. Optimal response strategies in most cases are fairly simple, and in many instances, such as the zero responses to open-ended type questions, fairly robust to alternative assumptions made about agent beliefs. In some situations, particular elicitation formats should be avoided altogether, while in others one faces a classic bias versus variance trade-off. The researcher should understand the trade-off being made in the choice of an elicitation format.

Claims about the specific incentive and informational properties of a particular elicitation format should not be made in the abstract. Careful attention needs be paid to the type of good being offered, the nature of the payment obligation for the good, and other aspects of the context in which the good is offered in order to clearly determine incentive and informational properties. For the binary discrete choice format, the introduction of a new private good turns out to be one of the worst cases for truthful preference revelation. The other bad case is to compare survey indications of willingness to voluntary contribute to provide a public good to actual contributions. Here neither estimate should approximate the true underlying WTP. One need not cast a binary discrete choice question explicitly as a vote in a referendum to get an incentive compatible question; it is sufficient to structure the question as advice to the government on the issue, a result that should be of use to researchers in areas where referenda are not frequently held.

None of our analysis has relied on agent experience or familiarity with the good. While these may influence the agent's true WTP for the good, they do not influence the incentive properties of question format in the context in which it is being used. Nor have we relied on any notion that agents learn about preferences and update them. Informational and incentive properties of formats do play a role in updating of optimal response strategies. Indeed, it is possible to recast some Bayesian models, such as the recent work of McLeod and Bergland (1999), as Bayesian updating not with respect to preferences, but rather, with respect to determining the optimal strategic response.

A number of elicitation formats commonly used in marketing research are currently attracting considerable attention in environmental valuation, both for the hope that more information can be collected from each agent (than can be collected with the binary discrete choice format) and for the hope that these newer formats will have fewer problems than does a binary discrete choice format. From an incentive perspective, this latter hope is likely to be misplaced with respect to the two most common valuation situations in environmental economics, the provision of public goods provided by the government and changes in a single quasi-public good provided by the government. The generalization of the binary discrete choice format in the directions used by marketing researchers causes it to lose its desirable incentive properties. Further, as the number of goods that must be described in a survey increases the time available to describe each good shrinks. For the introduction of new private goods, the multinomial choice format may be close to incentive compatible from the perspective of estimating marginal trade-offs between attributes as long as the perceived number of goods that are likely to be provided is sufficiently large. This is because deviations from truthful preference revelation are most likely to impact the scale parameter that drops out of marginal comparisons. This fortunate occurrence is less likely to be true for estimating the total value of a good since that calculation requires a consistent estimate of the true scale parameter.

Our work suggests that there are different natural underlying economic structures to different valuation problems. The typical problem in environmental valuation is the determination of the total value of a single good or a non-marginal change in a single good. The strategic incentives facing agents confronting this problem may be the major force that has moved researchers away from the open-ended and ranking formats toward binary discrete choice formats. The typical problem in marketing is the determination of a trade-off between different attributes of goods when many competing goods will be offered. Researchers in this area have moved from open-ended and discrete choice questions, to ranking questions, and then to the multinomial choice format.

A shift to the paired comparison or multinomial choice format has sometimes been recommended as a means of reducing or eliminating the sensitivity of the estimate of the value of a particular good to the sequence in which it was valued. However, this sensitivity is not a problem of elicitation format. Attempts to get sequence effects to go away by shifting to a question format that explicitly involves multiple goods are misguided. One of the major differences between private goods and public goods is that, for the former, agents themselves largely determine the order in which they obtain information about goods and make purchases of them. For public goods, the government through its control of the agenda determines the order in which projects are considered. Sequence effects (Carson, Flores, and Hanemann, 1998) are inherent to sequential decision making and because substitution effects enter WTP calculations much differently than they do demand calculations, sequence effects are apt to be large. Flores (1994) shows that the classic agenda control problem can be rewritten in terms of WTP and WTA sequences.

In closing, a remark on the term *hypothetical*, frequently affixed as an adjective in front of the word survey, is in order. In a famous and often cited remark on the early use of surveys for environmental valuation, Scott (1965) bluntly states: "Ask a hypothetical question and you get a hypothetical answer." *Hypothetical* as used here seems to imply that the responses are to an "imaginary" *inconsequential* situation, and, as such, the responses will have no influence on any relevant decision. From an economic perspective nothing can be inferred about respondent preferences from asking such a question.

The term *hypothetical*, however, also means conjecture, counterfactual, and contingent. This is the context usually used by researchers who ask preference questions. It is consistent with our definition of a *consequential* survey but an incomplete one because we require the agent to care about the alternatives and the agent to perceive that the agency will take the survey responses into account in its decision making. Our suggestion is to eschew the use of the word *hypothetical* in discussing preference questions in favor of *consequential* and *inconsequential* to emphasize the conditions requisite for the application of economic theory.

References

Adamowicz, W. J. Louviere, and M. Williams (1994), "Combining Revealed and Stated Preference Measures for Valuing Environmental Amenities," *Journal of Environmental Economics and Management*, 26, 271-292.

Adamowicz, W., J. Swait, P. Boxall, and J. Louviere (1997), "Perceptions Versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation," *Journal of Environmental Economics and Management*, 32, 65-84.

Alberini, A. (1995), "Efficiency vs Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models," *Journal of Environmental Economics and Management*, 29, 169-180.

Alberini, A., B. Kanninen, R.T. Carson (1997), "Modeling Response Incentive Effects in Dichotomous Choice Contingent Valuation Data," *Land Economics* 73, 309-324.

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

Bateman, I.J., I.H. Langford, R.K. Turner, and K.G. Willis (1995), Elicitation and Truncation Effects in Contingent Valuation Studies," *Ecological Economics*, 12, 161-179.

Bateman, I.J., A. Munro, B. Rhodes, C. Starmer, and R. Sugden (1997) "Does Part-Whole Bias Exist?: An Experimental Investigation," *Economic Journal*, 107, 322-332.

Bishop, R.C. and T.A. Heberlein (1979), "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased," *American Journal of Agricultural Economics*, 61, 926-930.

Bishop, R.C. and T.A. Heberlein (1990), "The Contingent Valuation Method," in *Economic Valuation of Natural Resources: Issues, Theory, and Applications*, R.L. Johnson and G.V. Johnson, eds. (Boulder, CO: Westview Press).

Bergstrom, T., L. Blume, and H. Varian (1986), "On the Private Provision of Public Goods," *Journal of Public Economics*, 29, 2-49.

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

Boyle, K.J., R.C. Bishop and M.P. Welsh (1985), "Starting Point Bias in Contingent Valuation Bidding Games, " *Land Economics*, 61, 188-194.

Boyle, K.J, F.R. Johnson, D. McCollum, and W.H. Desvousges, R.W. Dunford, and S.P. Hudson (1996), "Valuing Public Goods: Discrete Versus Continuous Contingent-Valuation Responses," *Land Economics*, 72, 381-396.

Brookshire, D.S., B.C. Ives, and W.D. Schulze (1976), "The Valuation of Aesthetic Preferences," *Journal of Environmental Economics and Management*, 3, 325-346.

Cameron, T.A., (1988), "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression," *Journal of Environmental Economics and Management*, 15, 355-379.

Cameron, T.A. and D.D. Huppert (1991), "OLS versus ML Estimation of Non-Market Resources Values with Payment Card Interval Data," *Journal of Environmental Economics and Management*, 17, 230-246.

Cameron, T.A. and J. Quiggin (1994), "Estimation Using Contingent Valuation Data from a "Dichotomous Choice with Follow-up" Questionnaire," *Journal of Environmental Economics and Management*, 27, 218-234.

Carson, R.T. (1985), "Three Essays on Contingent Valuation," unpublished dissertation, University of California, Berkeley.

Carson, R.T., N.E. Flores, K.M. Martin and J.L. Wright (1996), "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods," *Land Economics*, 72, 80-99.

Carson, R.T., T. Groves, and M.J. Machina (1997), "Stated Preference Questions: Context and Optimal Response," paper presented at the National Science Foundation Preference Elicitation Symposium, University of California, Berkeley.

Carson, R.T., N.E. Flores and W.M. Hanemann (1998), "Sequencing and Valuing Public Goods," *Journal of Environmental Economics and Management*, 36, 314-323.

Carson, R.T., W.M. Hanemann and R.C. Mitchell (1987). "The Use of Simulated Political Markets to Value Public Goods." Discussion paper 87-7, Department of Economics, University of California, San Diego, March.

Carson, R.T., W. M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud, and V.K. Smith (1994), "Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight," report to National Oceanic and Atmospheric Administration, September.

Carson, R.T. and D. Steinberg (1990), "Experimental Design for Discrete Choice Voter Preference Surveys," in *1989 Proceeding of the Survey Methodology Section of the American Statistical Association*, (Washington: American Statistical Association, 1990).

Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum (1997), "Using Donation Mechanism To Value Nonuse Benefits From Public Goods," *Journal of Environmental Economics and Management*, 33, 151-162.

Chapman, R.G. nd R. Staelin (1982), "Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model," *Journal of Marketing Research*, 19, 281-299.

Cronin, F.J. (1982), "Valuing Nonmarket Goods Through Contingent Markets," report to U.S. Environmental Protection Agency by Battelle Memorial Institute, Richmond Washington.

Cummings, R.G., G.W. Harrison and E.E. Rutström (1995), "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?", *American Economic Review* 85, 260-266.

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

Cummings, R.G. and L. Osborne. (1998), "Does Realism Matter in Contingent Valuation Surveys?" *Land Economics*, 74, 203-215.

Farmer, M.C. and A. Randall (1996), "Referendum Voting Strategies and Implications for Follow-Up Open-Ended Responses," paper presented at the annual U.S.D.A. W- 133 meeting, Jekyll Island, GA, February.

Farquharson, R. (1969), *Theory of Voting.* (New Haven: Yale University Press).

Flores, N.E. (1994), "The Importance of Agenda and Willingness to Pay." Paper presented at the 1994 Public Choice Society Meetings, Austin, TX, April.

Gibbard, A. (1973), "Manipulation of Voting Schemes: A General Result," *Econometrica* 41, 587-601.

Green, J.R. and J.J. Laffont (1978), "A Sampling Approach to the Free Rider Problem," in Agnar Sandmo, ed., *Essays in Public Economics* (Lexington, MA: Lexington Books).

Greenhalgh, C. (1986), "Research for New Product Development," in *Consumer Market Research Handbook*, 3rd ed., R.M. Worcester and J. Downham, eds. (Amsterdam: North-Holland).

Grether, D.M and C.R. Plott (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69, 623-638.

Groves, T. (1973), "Incentives in Teams," *Econometrica*, 41, 617- 631.

Groves, T., R. Radner, and S. Reiter, eds. (1987), *Information, Incentives, and Economic Mechanisms: Essays in Honor of Leonid Hurwicz* (Minneapolis: University of Minnesota Press).

Haab, T.C. and K.E. McConnell (1998), "Referendum Models and Economic Values: Theoretical, Intuitive, and Practical Bounds on Willingness to Pay," *Land Economics*, 74, 216-229.

Haab, T.C. and K.E. McConnell (1997), "Referendum Models and Negative Willingness to Pay: Alternative Solutions," *Journal of Environmental Economics and Management*, 32, 251-270.

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

Hanemann, W.M. (1984a), "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52, 541-561.

Hanemann, W.M. (1984b), "Welfare Evaluations in Contingent Valuation: Experiments with Discrete Responses," *American Journal of Agricultural Economics*, 66, 335-379.

Hanemann, W.M., and B. Kanninen (1999), "The Statistical Analysis of Discrete-Response." In *Valuing the Environment Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC and Developing Countries*, edited by Ian Bateman and Ken Willis (Oxford: Oxford University).

Hanemann, W.M., J. Loomis, and B. Kanninen (1991), "Statistical Efficiency of Double Bounded Dichotomous Choice Contingent Valuation," *American Journal of Agricultural Economics*, 73, 1255-1263.

Hausman, J.A., ed. (1993), *Contingent Valuation: A Critical Assessment* (Amsterdam: North-Holland).

Hausman, J.A. and P. Ruud (1987), "Specifying and Testing Econometric Models for Rank-Ordered Data," *Journal of Econometrics*, 34, 83-104.

Hensher, D.A. (1994), "Stated Preference Analysis of Travel Choice—The State of Practice," *Transportation*, 21, 107-133.

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

Holt, Charles A. (1986), "Preference Reversals and the Independence Axiom," *American Economic Review*, 76, 508-515.

Hurwicz, L. (1986), "Incentive Aspects of Decentralization," in *Handbook of Mathematical Economics, vol. III*, K.J. Arrow and M.D. Intrilligator, eds. (Amsterdam: North-Holland).

Inforsino, W.J. (1986), "Forecasting New Product Sales From Likelihood of Purchase Ratings (with discussion)," *Marketing Science*, 5, 372-390.

Johannesson, M., B. Liljas, P.O. Johansson (1998), "An Experimental Comparison of Dichotomous Choice Contingent Valuation Questions and Real Purchase Decisions, *Applied Economics*, 30, 643-647.

Johnson, F.R. and W.H. Desvousges (1997), "Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs," *Journal of Environmental Economics and Management*, 34, 79-99.

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

Karni, E. and Z. Safra (1987), "Preference Reversal and the Observability of Preferences by Experimental Methods," *Econometrica*, 55 675-685.

King, G. (1989), *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*, (New York: Cambridge University Press).

Kriström, B. (1990), "A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies," *Land Economics*, 66, 135-139.

Kriström, B. (1997), "The Practical Problems of Contingent Valuation, in R.J. Kopp, W.W. Pommerhene, and N. Schwartz, eds., *Determining the Value of Non-Marketed Goods* (Boston: Kluwer).

Lankford, R.H. (1985), "Preferences of Citizens for Public Expenditures on Elementary and Secondary Education," *Journal of Econometrics*, 27, 1-20.

Louviere, J.J. (1994), "Conjoint Analysis," in *Handbook of Marketing Research*, R. Bagozzi, ed. (Oxford: Oxford University Press).

Lohmann, S. (1994), "Information Aggregation Through Costly Political Action," *American Economic Review* 84, 518- 530.

Lunander, A. (1998), "Inducing Incentives to Understate and to Overstate Willingness to Pay within the Open-Ended and the Dichotomous-Choice Elicitation Format," *Journal of Environmental Economics and Management*, 35, 88-102.

Lupia, A. (1994), "Shortcuts Versus Encyclopedias—Information and Voting Behavior in California Insurance Reform Elections," *American Political Science Review*, 88, 63-76.

Machina, M.J. (1995), "Non-Expected Utility and the Robustness of the Classical Insurance Paradigm," *Geneva Papers on Risk and Insurance Theory*, 20, 9-50.

McConnell, K. E. (1990), "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation," *Journal of Environmental Economics and Management*, 18, 19-34.

McDowell, I. and C. Newell (1996), *Measuring Health: A Guide to Rating Scales and Questionnaires, 2nd ed.* (New York: Oxford University Press).

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, P. Zarembka, ed., (New York: Academic Press).

McFadden, D. (1994), "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics* 76, 689-708.

McFadden, D. (1999) "Rationality for Economists?," *Journal of Risk and Uncertainty*, 19, 73-105.

McLeod, D.M; Bergland, O. (1999), "Willingness-to-pay Estimates Using the Double-Bounded Dichotomous-Choice Contingent Valuation Format: A Test for Validity and Precision in a Bayesian Framework, *Land Economics*, 75, 115-125.

Mitchell, R.C. and R.T. Carson (1986), "The Use of Contingent Valuation Data for Benefit-Cost Analysis in Water Pollution Control," report to the U.S. Environmental Protection Agency, September 1986.

Mitchell, R.C. and R.T. Carson (1989), *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Baltimore: Johns Hopkins University Press.

Moulin, H. (1994), "Social Choice," in *Handbook of Game Theory with Economic Applications*, R.J. Aumann and S. Hart, eds. (Amsterdam: North- Holland).

Neil, H., R.G. Cummings, P.T. Ganderton, and G.W. Harrison (1994), "Hypothetical Surveys and Real Economic Commitments," *Land Economics* 70, 145-154.

Payne, S. (1951), *The Art of Asking Questions* (Princeton: Princeton University Press).

Polasky, S., O. Gainutdinova, and J. Kerkvliet (1996), "Comparing CV Responses with Voting Behavior: Open Space Survey and Referendum in Corvallis Oregon," paper presented at the annual U.S.D.A. W- 133 meeting, Jekyll Island, GA, February.

Popkins, S. (1991), *The Reasoning Voter* (Chicago: University of Chicago Press).

Posavac, S.S. (1998), "Strategic overbidding in contingent valuation: Stated economic value of public goods varies according to consumers expectations of funding source," *Journal of Economic Psychology*,19, 205-214.

Randall, A. (1994), "A Difficulty with the Travel Cost Method," *Land Economics*, 70, 88-96.

Randall, A. (1996), "Calibration of CV Responses: Discussion," in *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs*, D.J. Bjornstad and J. R. Kahn (Brookfield, VT: Edward Elgar).

Randall, A., B.C. Ives, and C. Eastman (1974), "Bidding Games for Valuation of Aesthetic Environmental Improvements," *Journal of Environmental Economics and Management* 1, 132-149.

Richer, J. (1995), "Willingness to Pay for Desert Protection," *Contemporary Economic Policy*, 13, 93-104.

Romer, T. and H. Rosenthal (1978), "Political Resource Allocation, Controlled Agendas, and the Status Quo," *Public Choice*, 33, 27-43.

Rowe, R.D., W.D. Schulze, and W. Breffle (1996). "A Test for Payment Card Biases," *Journal of Environmental Economics and Management* 31, 178-185.

Samuelson, P.A. (1954), "The Pure Theory of Public Expenditure," *Review of Economics and Statistics* 36, 387- 389.

Satterthwaite, M. (1975), "Strategy-Proofness and Arrow Conditions: Existence and Correspondence Theorems for Voting Procedures and Welfare Functions," *Journal of Economic Theory* 10, 187-217.

Swait, J. and J.J. Louviere (1993), "The Role of the Scale Parameter in the Estimation and Use of Generalized Extreme Value Models," *Journal of Marketing Research*, 30, 305-314.

Scott, A. (1965), "The Valuation of Game Resources: Some Theoretical Aspects," *Canadian Fisheries Report* 4, 27-47.

Seip, K. and J. Strand (1992), "Willingness to Pay for Environmental Goods in Norway: A CV Study with Real Payment," *Environmental and Resource Economics*, 2, 91-106.

Smith, V.K. and W.H. Desvousges (1986), *Measuring Water Quality Benefits* (Boston: Kluwer).

Sudman, S., N.M. Bradburn, and N. Schwarz (1996), *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology* (San Francisco: Jossey-Bass Publishers).

Sugden, R. (1999) "Alternatives to Neo-classical Theory of Choice," in *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the USA, EC, and Developing Countries*, edited by I.J. Bateman and K.G. Willis. (Oxford: Oxford University Press).

Tversky, A., P. Slovic, and D. Kahneman (1990), "The Causes of Preference Reversal," *American Economic Review*, 80, 204-217.

Varian, H. (1992), *Microeconomic Analysis*, 3rd. ed. (New York: Norton)

OPTIMAL DESIGN OF CHOICE EXPERIMENTS FOR NONMARKET VALUATION

Barbara J. Kanninen
Hubert H. Humphrey Institute of Public Affairs
University of Minnesota

Please send correspondence to:

2431 N. Nottingham St.
Arlington, VA 22207

Phone: (703) 536-6949
FAX: (703) 536-6905
e-mail: barbkann@aol.com

This research has been funded by NSF Grant NSF/SBR-9613045 and EPA Grant 9975726

Abstract

This paper derives optimal designs for linear, multi-attribute, binary choice experiments. The purpose of optimal design is to improve model estimation, and obtain the equivalent effects of a larger sample size, by improving the informational content of the data collected. The two optimal design criteria that are addressed are "D-optimality" and "C-optimality." D-optimality is the maximization of the determinant of the Fisher information matrix. The criterion seeks to jointly maximize the efficiency of the parameter estimates. For the valuation context, C-optimality is the minimization of the variance of either total or marginal willingness to pay. Both criteria are developed in stages within the paper, starting with the univariate linear model and building toward the multi-attribute, binary model. This presentation allows the reader to see, exactly, where and how the different aspects of the optimal designs come to be.

With the linear model, D-optimality implies that attribute levels should be placed at their extreme values according to a main effects, orthogonal array. This result is tempered when discrete choices are introduced. With the binary model, all attributes but one should be placed orthogonally at their extreme values, with the base alternative being generated by taking the foldover of the first alternative. The remaining attribute is used as a balancing variable to obtain optimal response rates. The optimal response rates vary depending on the number of attributes in the model, ranging from .82/.18 for a one-attribute binary model to .67/.33 for an eight-attribute model.

C-optimal design emphasizes the estimation of marginal or total willingness to pay. With both the linear and binary models, the design solution requires that each attribute within each observation be balanced at exactly its marginal value. Unfortunately, this solution causes multicollinearity and prevents model estimation.

The author concludes that the lesson learned from the C-optimal design solution is that the approach to estimating willingness to pay, as a ratio of estimated parameters, is inherently inefficient. Despite the fact that our primary interest is willingness to pay, it seems that the D-optimal design approach is the most appropriate for practical purposes.

I. Introduction

To assess the total value, including use and nonuse values, of nonmarket goods such as environmental amenities, researchers often employ choice experiments that allow them to estimate willingness to pay (WTP) for hypothetical goods or services. Until recently, the standard technique for this purpose has been the contingent valuation (CV) method (Bateman and Willis, 1999, Mitchell and Carson, 1989). CV questions generally provide a detailed description of the goods or services being valued, describe the hypothetical circumstances under which they would be made available to respondents, and elicit WTP responses for these goods or services. Recently, a similar but more complex approach to choice experiments, sometimes referred to as conjoint analysis, has been used in several environmental contexts (Magat et al., 1988; Opaluch et al., 1993; Adamowicz et al., 1994). Conjoint analysis is a marketing technique that can be used to assess values for *attributes* of market or nonmarket goods based on experimental respondents' willingness to trade-off different bundles of these attributes (Carson et al. 1994, Louviere 1988).

In these choice experiments, respondents are presented with a set of alternative scenarios that differ in terms of a series of attributes (which generally include price) and are asked to choose their most preferred alternative. The scenarios in the choice set differ by the *levels* of the various attributes. For example, a respondent might be asked to choose among different beach experiences that vary by their congestion levels, beach aesthetics and water quality. Further, there might be an admission fee that varies across beach scenarios. Congestion, beach aesthetics, water quality and price are attributes of each beach alternative, and the particular amounts assigned to the attributes are the attribute levels. The researcher can use the experimental responses to estimate a model of choice behavior that allows the estimation of separate marginal values for each attribute, or a WTP measure for any particular beach experience, as described by a specific set of attribute levels.

Discrete response CV questions are simple versions of experimental choices. With discrete response CV, there is a choice between a status quo situation and a single scenario with fixed attribute levels offered at a particular price. Respondents are asked whether or not they would be willing to pay the offered price for the described scenario. This approach allows the researcher to estimate WTP for the alternative scenario but not for the individual attributes associated with that scenario, as they do not vary over the sample set.

There are several advantages to using more complex choice experiments instead of CV for valuing environment amenities. Choice experiments allow more flexibility for valuing scenarios. Because scenarios are presented with different combinations of attribute levels, the researcher can use the responses to construct values for several different scenarios. This is particularly advantageous when the researcher is not sure, a priori, what particular scenario will be of most interest, for example, when conducting a benefit-cost analysis under uncertainty.

The researcher can also assess the trade-offs respondents are willing to make between any two attributes. With CV, the only trade-off respondents are asked to make is between dollars and the amenity of interest. With larger choice experiments, respondents are asked to trade a variety of different attributes simultaneously. Acceptable trade-offs between any two attributes can be teased out of the response data using the econometric choice model. This information is particularly useful for "resource compensation," a method that is used in natural resource damage assessment to assess

compensation for the loss of resource amenities in terms of other resource amenities (Jones and Hanemann, 1996).

The advantages of these choice experiments come at a cost though. As the numbers of attributes and levels to be included in an experiment increase, the number of observations required to estimate the choice model increases *exponentially*. For example, a "full factorial" experimental design for three attributes, each taking two levels, requires at least 2^3, or 8 distinct observations to identify the complete set of parameters (including higher order terms). Increasing the number of attributes to four requires 2^4, or 16, distinct observations to estimate all the parameters. For studies conducted under conditions of uncertainty or for resource compensation, it is quite plausible that the number of attributes and levels to be considered will be large. Since survey administration costs are directly proportional to the sample size, it is important to develop techniques for eliciting as much information as possible from each observation so that survey costs can be kept as low as possible for any given problem.

This paper derives optimal experimental designs for main effects, multi-attribute, binary choice experiments. The idea behind optimal design is that the researcher has the opportunity to design his or her own data by specifying the content of the choice experiments. The number of attributes, the levels they take, and how they combine into choice sets, all affect the amount and nature of the statistical information that an experiment will provide after responses are collected. By employing optimal design results, either exactly or approximately, a researcher can improve the efficiency of model estimates and, effectively, obtain the equivalent effects of a larger sample size.

Optimal design recommendations are, likely, going to be of most interest to researchers working under limited budgets. They, of course, have the greatest need to maximize the information they collect from each observation, but, also, they might be more likely to be able to manipulate their designs during the data collection process. It turns out that the optimal designs derived here rest on the notion of obtaining particular response rates for each choice set. By manipulating the design during the process, a researcher can improve the quality of the data as the experiment goes on.

To implement optimal design, a specific research goal must be stated in terms of a "design criterion."[1] When the goal is to estimate the overall model as well as possible, the researcher will probably focus on the criterion called "D-optimality," which is the maximization of the determinant of the Fisher information matrix. This is a criterion that, in a sense, seeks joint statistical efficiency of all model parameters.

Environmental valuation problems often are more focused, however, on estimation of one or more specific measures. In particular, researchers typically need to estimate total or marginal WTP, both of which are nonlinear functions of the choice model parameters. In cases of resource compensation, the main goal might be to estimate a marginal rate of substitution between two particular attributes. A more appropriate design criterion for nonmarket valuation, therefore, would focus on these statistical measures. The optimal design criterion that optimizes estimated functions of the model parameters is called "C-optimality." This is the second design model addressed in this paper.

[1] See Federov (1972) and Silvery (1980) for descriptions of optimal design criteria and methodology.

To optimize these criteria, the number of attribute levels, the levels themselves, and the make-up of the choice sets are assumed to be design parameters. In other words, it is assumed that any of these factors can be manipulated to improve estimation efficiency. This is done by assuming, a priori, that all attributes are continuous variables that can be bounded above and below. The optimal design solutions, then, specifically describe where and how the various attribute levels should be placed to obtain the most information as specified by the design criterion employed . This approach is consistent with the optimal design literature for dose-response models (Abdelbasit and Plackett 1980, Minkin 1987, Wu 1988) and the literature on optimal design for CV (Alberini 1995, Alberini and Carson 1993, Cooper 1993, Kanninen 1993a and 1993b).

The paper is organized as follows. Section 2 reviews the binary choice model and briefly discusses the standard approach to experimental design for choice models. Section 3 introduces the D-optimality criterion and steps through a process that describes D-optimal designs for the linear and binary choice models. D-optimal designs for the linear and one-attribute binary models are already well-known. They are described in detail here to give the reader an understanding of the principles of optimal design and to show the sources of specific aspects of the later optimal designs. Section 4 provides the same approach for C-optimal designs. Section 5 offers concluding comments and thoughts about the course of future research.

2. The Logit Model for Choice Experiments

The utility-theoretic approach to modeling discrete choices was developed by McFadden (1974) and is discussed in detail by Ben-Akiva and Lerman (1985). When consumer *i* is presented with a binary choice set that differ by a particular set of *K* attributes, designated $z_i = \{z_{1i}^q, z_{2i}^q, ..., z_{Ki}^q\}$, for $q=\{0,1\}$, he or she will choose the alternative that offers the greatest utility. Specifying consumer *i*'s utility for alternative *q* to be linear with a fixed component, $\beta_1 z_{1i}^q + ... + \beta_K z_{Ki}^q$,[2] and an additive random component, ε_i^q, that follows an extreme value distribution, the probability that consumer *i* prefers choice 1 over alternative 0 is:

$$P(\theta_i) = \frac{Exp(\theta_i)}{1 + Exp(\theta_i)} \tag{1}$$

where:

$$\theta_i = \sum_{k=1}^{K} \beta_k (z_{ki}^1 - z_{ki}^0)$$

For the remainder of this paper, alternative 0 will be referred to as the "base alternative."

[2] This model specification does not include demographic characteristics or alternative-specific constants. These are excluded to keep notation manageable and because they are generally not aspects of the design that can be manipulated to improve design efficiency.

Letting y_i^q equal 1 when consumer *i* prefers alternative *q* and 0 otherwise, the individual log-likelihood is:

$$\log L(q_i; y_i) = \sum_{q=0}^{Q} y_i^q \log p(q_i^q) \tag{2}$$

The log-likelihood function is the sum of all individual log-likelihoods for $i = \{1,...,N\}$.
An important aspect of the design problem is that the log-likelihood function is a function only of the differences between attribute level vectors, $z_i^q - z_i^0$. For notational convenience, in the remainder of the paper, let $x_i^q = z_i^q - z_i^0$ for all *i* and *q*. Further, let x_i^q be continuous and bounded: $x_i^q \in [-1, 1,]$. These bounds are chosen, without loss of generality, to allow the *x's* to correspond with the {-1,1} notation often used in the experimental design literature. For actual experiments, these bounds should be translated to levels the researcher deems practical for the particular attributes being considered.

Once maximum likelihood estimation is performed on the above model, a number of analyses may be performed, for example, total willingness to pay (WTP) for alternative *q* may be estimated as:

$$\hat{WTP}_i^q = \frac{-(\hat{b}_1^q x_{1i}^q + \hat{b}_i^q x_{2i}^q + \ldots + \hat{b}_k^q x_{ki}^q)}{\hat{b}_1^q} \tag{3}$$

where b_1^q is arbitrarily specified to be the coefficient on the price attribute and the levels of the attributes are defined by the researcher as the levels in the package to be valued. Further, the marginal rate of substitution of attribute *m* for *l* may be estimated as:

$$\hat{MRS}_{ml} = \frac{\hat{b}_l^q}{\hat{b}_m^q} \tag{4}$$

When attribute *m* is the price attribute, this measure is equal to the marginal WTP for attribute *l*.

Discrete response CV is the special case of a binary choice model with one attribute. With CV, there is a choice between a status quo situation and a single scenario with fixed attribute levels offered at a particular price. Respondents are asked whether or not they would be willing to pay the offered price for the described scenario. This approach allows the researcher to estimate WTP for the alternative scenario but not for the individual attributes associated with that scenario, as they do not vary over the sample set. For this case, q_i is equal to $a + bx_i$ where x_i is the offered price.

CV experiments are performed principally to estimate WTP. Making no further assumptions on the model, mean or median WTP can be estimated as:

$$\hat{WTP} = \frac{-\hat{a}}{\hat{b}} \tag{5}$$

Louviere (1988) and Louviere and Woodworth (1983) summarize the traditional approach to experimental design for choice experiments.[3] The principal consideration in these discussions is model identification rather than statistical optimality. The approach assumes the researcher has specified the attribute levels to be used in the choice experiments in advance of the design stage.

Table 1 shows a main effects design for the case of three attributes that each take two levels. Design tables are typically presented using {-1, 1} notation (or, 1,2,3,… when there are more than two attribute levels). The researcher is expected to substitute his or her pre-specified attribute levels for these values.

Because the main effects design is a reduced design compared to the full factorial, it is referred to as a "fractional factorial design." The limitations to using such a design are demonstrated in Table 1: each of the two-way interactive effects are confounded with a main effect (e.g. the occurrences of x_1x_2 are equivalent to the occurrences of x_3) and the three-way effect does not vary. Under the assumption that these effects are negligible though, the main effects model is identifiable. In general, it would be preferable to include the interactive effects, at least for testing purposes, although the sample size increases substantially to do so.[4] Despite the limitations, main effects designs are standard with choice experiments because they do not require inordinately large sample sizes.

Note that the design array in Table 1 only provides information about the placement of attribute levels for one alternative. With binary or multinomial choice experiments, one or several other alternatives must be generated. Louviere (1988) describes several possible approaches to designing these alternatives. One is to take the "foldover" of the first attribute, or the exact opposite on an attribute-by-attribute basis. This approach is, obviously, useful only for a binary choice. The primary methods for generating larger choice sets are randomized or cyclical procedures. A randomized procedure is just what it sounds like. Alternative attributes are generated randomly. Cyclical procedures seem to be used more often. Here, attribute levels are chosen for each alternative in turn by taking the next level available. For example, when there are three attributes, and the first alternative uses level one, the second alternative uses level two and the third uses level three. After level three, the cycle returns to level one.

Generally researchers try to maintain balance across choice sets, so that each attribute level appears an equal number of times. They also try to obtain minimal overlap, that is, as few repeats of the same attribute level within choice sets as possible; and prevent dominated alternatives, or alternatives that offer less utility on an attribute by attribute basis. As will be shown later in the paper, the first and third of these principles are not necessarily optimal. On the other hand, the optimal designs derived here guarantee that no alternative will be dominated, so that the second principle always holds.

[3] The statistical underpinnings for experimental design are thoroughly described by Winer et al. 1991.

[4] Although this paper does not address interactive effects, it will certainly be important to include them in future design research, as substitution effects are often significant in environmental valuation amenities (Cummings, Ganderton and McGuckin 1992, Hoehn and Loomis, 1993).

3. D-Optimality

D-optimality refers to the maximization of the determinant of the Fisher information matrix, which is equivalent to the minimization of the generalized variance of the parameters, or the minimization of the joint confidence sphere surrounding the parameter estimates. It is, in a sense, a criterion that seeks statistical efficiency for the overall model.

In this section, D-optimal designs are derived for the univariate and multivariate linear and binary choice models. The results for the linear and univariate binary models exist in the literature already. They are described here to illustrate to the reader where and how particular aspects of the later design solutions emerge. Results for the multi-attribute binary choice models are the work of the author.

D-Optimality for the Linear Model with One Independent Variable

The example of the linear model illustrates the importance of placing design points at the extremes of their domains. This result is tempered later in the paper when discrete choices are introduced.

Consider the one-variable linear model:

$$y_i = a + bx_i + e_i \tag{6}$$

with e_i independent and identically distributed N(0,♦^{2}) for $i = \{1, ..., N\}$.[5] The single independent variable is assumed to be continuous and, for convenience later, bounded: $x_i \in [-1, 1]$, $i = \{1, ..., N\}$. The Fisher information matrix is:

$$I(a,b) = \frac{1}{s^2}\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \tag{7}$$

The determinant can be written as:

$$\left|I(a,b)\right| = \frac{1}{s^2}\sum_{i=1}^{N}\sum_{j=1}^{N} x_i^2 - x_i x_j$$

which simplifies to:

$$\left|I(a,b)\right| = \frac{1}{s^2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(x_i - x_j)^2 \tag{8}$$

The principle behind the D-optimal solution is easily understood from equation 8: the determinant is a linear function of the squared differences of all pairs of the x variable. First, it can be seen that, because the right hand side of equation 8 is a sum, the optimal solution is an arbitrary pair (i, j) of

[5] The parameter definitions in the linear model are not analogous to the parameters in the choice model. This model is provided for illustrative purposes, rather than as a direct link to the choice situation.

the x variable. This means that the optimal design solution is a two-point design with $N/2$ (or, in the case of an odd N, $(N+1)/2$) observations at x_i^*, the optimal solution for x_i, and $N/2$ (or $(N-1)/2$) at x_j^*, the optimal solution for x_j. Second, the optimal pair $\{x_i^*, x_j^*\}$should be spread as far from each other as possible; in other words, one variable (let it be x_i) should be placed at the maximum possible value for x (+1, by assumption) and the other (x_j) should be placed at the minimum value (-1, by assumption).

The optimal solution is intuitive, in that it takes only two design points to draw a regression line, and, given that those two points will be observed with error, the regression line will most closely approximate the true relationship between the regressor and independent variable if the two design points are positioned as far apart as possible. No other point along the domain of x is necessary for model identification, or statistically more informative, from a D-optimal perspective.

It should be noted, immediately, that even this simple and straightforward design solution comes with caveats. Principally, for the optimality result to hold, the specified model must be the true one. If, for example, there are interactive or higher order terms in the true model, this solution is no longer optimal. A basic fact of life in the world of optimal design is that researchers must know a lot, up-front, about what they will ultimately be estimating. This caveat is usually mentioned in association with nonlinear models, when researchers must even know the parameter values beforehand, but it bears noting for the case of the linear model as well.

D-Optimality for the Linear Model with Multiple Independent Variables

For the general case of K experimental variables:

$$y_i = b_1 x_{1i} + b_2 x_{2i} + \ldots + b_K x_{Ki} + e_i \tag{9}$$

The K x K Fisher information matrix is:

$$I(\mathrm{B}) = \frac{1}{s^2}\begin{pmatrix} \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \cdots & \sum x_{1i}x_{Ki} \\ & \sum x_{2i}^2 & \cdots & \sum x_{2i}x_{Ki} \\ \bullet & & \ddots & \vdots \\ & & & \sum x_{Ki}^2 \end{pmatrix} \tag{10}$$

and the determinant is:

$$\left|I(\mathrm{B})\right| = (1/s^2)\,|\,X'X\,| \tag{11}$$

where X is a N x K matrix containing all vectors, $x_1 \ldots x_m$. Maximizing $|\,I\,|$ is equivalent to maximizing $|\,X'X\,|$.

To understand the properties of the design solution, it is useful here to consider the geometric properties of a determinant. In the case of a matrix consisting of two, two-element vectors, the determinant is equivalent to the area that results from completing the vectors into a parallelogram. In the case of a multi-dimensional matrix, the same act of completing the vectors results in a multi-dimensional "parallelogram." If the matrix were nonorthogonal, the dimension of

the parallelogram would be less than the dimension of the matrix and completion of the vectors would result in a partially collapsed cube. Further, the area of the cube is maximized by maximizing the length of each vector.

Two conclusions can be drawn from this: first, that, to the extent possible, the D-optimal solution will be *orthogonal* and second, that all design points will placed be at their boundary points, or endpoints of the domain of x. These two properties will maximize the diagonals of the information matrix and zero out the off-diagonal terms. Overall, the design solution will contain points that are as far apart from each other as possible. For a main effects model, any orthogonal main effects array that can be drawn from the full factorial is optimal.

Assuming the bounds of [-1,1] for all attributes, one optimal design solution for three quantitative variables is the design presented in Table 1. In general, when an orthogonal design exists for a particular number of attributes, the optimal design will be that orthogonal design, modified to reflect the assumed upper and lower bounds on the experimental variables. In a sense, the optimal solution reduces $x_2 \ldots x_K$ to a series of qualitative (two-level) variables with the two levels being the respective upper and lower bounds of each attribute.

D-Optimality for the One Variable Logit Model

The case of one independent variable with a constant term gives: $q_i = a + bx_i$. This is essentially the CV model. The Fisher information matrix for this model (dropping the q term for simplicity) is:

$$I = \begin{pmatrix} \sum P_i(1-P_i) & \sum P_i(1-P_i)x_i \\ \sum P_i(1-P_i)x_i & \sum P_i(1-P_i)x_i^2 \end{pmatrix} \tag{12}$$

and the determinant is:

$$|I| = \sum_{i=1}^{N} \sum_{j=i+1}^{N} P_i(1-P_i)P_j(1-P_j)(x_i - x_j)^2 \tag{13}$$

To derive the optimal design in terms that are independent of the specific parameters values for a and b, equation 13 can be converted to:

$$|I| = \left(\frac{1}{b}\right)^2 \sum_{i=1}^{nN} \sum_{j=i+1}^{N} P_i(1-P_i)P_j(1-P_j)(q_i - q_j)^2 \tag{14}$$

This determinant is a function of two design points and is therefore maximized with only two points: q_i and q_j. The expression has two components: a squared utility difference term, $(q_i - q_j)^2$ and a probability weighting term: $P_i(1-P_i) \bullet P_j(1-P_j)$. Taken alone, the probability weights would be maximized at $P_i = P_j = .50$. This illustrates the influence of "utility balance" (Huber and Zwerina, 1996) in optimal design for binary response models. With probabilities of .50, consumers are, on average, perfectly indifferent between the two alternatives offered. On the other hand, the squared difference term would be maximized by design points placed at their extreme limits: where P_i and P_j

are closer to 0 or 1. This influence is just the opposite of utility balance: with probabilities of 0 or 1, consumers prefer one choice over the other 100% of the time. The optimal solution can be derived numerically and is a compromise between these two influences: $\{q_i^*, q_j^*\} = \{-1.54, +1.54\}$, a symmetric design at the 18th and 82nd percentiles of the underlying response function.

To generate the price offers associated with this design solution, the researcher can solve for $x_i = (q_i^*, - a)/b$ and $x_j = (q_j^* - a)/b$, or, more directly, determine the levels of x_i and x_j that would give $P(q_i) = .18$ and $P(q_j) = .82$. Prices should be set so that, for half the cases, 18% of respondents accept the bid offer and 82% reject, and for the other half, 82% accept and 18% reject.

To implement the design solution exactly, the research must know, or be able to approximate, the underlying model. In practice, researchers generally have some knowledge of the underlying model, based on focus group or pretest information, before conducting their final version of the survey. Further, Kanninen (1993b) and Nyquist (1992) have shown that a sequential approach to conducting CV surveys can substantially improve the information available to the researcher and the efficiency of the ultimate estimates obtained. Note that, by sequential approach, these researchers meant that prices, or bids, would be updated over the course of the experiment, not during an interview with one experimental respondent. Rather, bids would be updated after as sets of observations have been collected.

Although these researchers both examined parametric approaches to bid updating, where each update would be based on the estimated model parameters at each point in time, it is also possible to update bids nonparametrically. With this approach, only the empirical acceptance rates for each choice set are used to update bids. Bids for subsequent observations would be raised when empirical acceptances fall below the optimal level and lowered when acceptances are too frequent. Such a procedure was implemented in practice for a multivarite binary choice experiment and is described in the next sub-section.

D-Optimal Design for the Multivariate Logit Model

The K x K Fisher information matrix for the case of multiple attributes is:

$$I = \begin{pmatrix} \sum w_i x_{1i}^2 & \sum w_i x_{1i} x_{2i} & \cdots & \sum w_i x_{Ki} \\ & \sum w_i x_{2i}^2 & \cdots & \sum w_i x_1 x_{Ki} \\ \bullet & & \ddots & \vdots \\ & & & \sum w_i x_{Ki}^2 \end{pmatrix} \tag{15}$$

where $w_i = P_i(1\text{-}P_i)$. In matrix notation, equation 15 becomes:

$$|I| = (PX)'[(I - P)X] \tag{16}$$

where P is a N x N diagonal matrix with diagonal elements p_i and X is the N x K matrix with rows equal to the vectors x_{ki} for $k = \{1, ..., K\}$ and $i = \{1, ..., N\}$.

Because of the complexity of this optimality problem, it is useful to begin the process by determining the optimal number of distinct design points that will comprise the optimal solution. Using the additive property of determinants, the sums within each row in the determinant of I can be deconstructed, one at a time, into individual components so that the determinant can be expressed as a sum over all combinations of K out of the N observations:

$$|I| = \sum_{i=1}^{N}\sum_{j=1}^{N}\cdots\sum_{z=1}^{N}\begin{vmatrix} w_i x_{1i}^2 & w_i x_{1i} x_{2i} & \cdots & w_i x_{Ki} \\ & w_j x_{2j}^2 & \cdots & w_j x_{1j} x_{Kj} \\ \bullet & & \ddots & \vdots \\ & & & w_z x_{Kz}^2 \end{vmatrix} \tag{17}$$

Equation 17 expresses | I | as a sum over $\binom{N}{K}K!$ functionally equivalent terms that each contain K observations. Its maximum can therefore be obtained by maximizing one particular determinant from equation 17 for an arbitrary set of K observations. This is not surprising, as K is the minimum number of distinct observations necessary to identify a model with K parameters. With the full sample, the optimal design will contain N/K sets of the K optimal design points.

Converting the determinant in equation 17 to matrix notation, and using the fact that for square determinants, A and B, |AB| = |A| |B|, we have:

(18)

$$|I| = \sum_{i=1}^{N}\sum_{j=1}^{N}\cdots\sum_{z=1}^{N} p_i p_j \ldots p_z (1-p\ )(1-p_j)\ldots(1-p_z)\left|X^*_{i,j,\ldots,z}\right|^2$$

where $X_{i,j,\ldots,z}{}^*$ represents the matrix with rows composed of the vectors x_i, x_j,...,x_z.

To maximize equation 18 it is useful to construct a reparameterization of the problem. Without loss of generality, let $q_{1i} = b_1 x_{1i} + b_2 x_{2i} + ... + b_K x_{Ki}$, $q_{ki} = x_{ki}$ for $k = \{2,...,K\}$ and $i = \{1, ..., N\}$ and $Q_{i,j,\ldots,z}{}^*$ represent the matrix with rows of vectors q_i, $q_{j,\ldots,}$ q_z. Equation 18 can then be expressed as:

$$|I| = \left(\frac{1}{b_1}\right)^2 w_i w_j \ldots w_z \left|\Theta^*_{i,j,\ldots z}\right|^2 \tag{19}$$

What is convenient about this formulation is that q_2 through q_K appear only in the determinant part of the right hand side in equation 19. The expressions w_i, w_j, ..., w_z are functions only of the q_{1i}'s. With this separation, the maximization problem can be solved in two stages: first, maximizing equation 19 with respect to q_2 through q_m for an arbitrary set of q_{1i}'s then plugging these solutions into equation 19 and maximizing with respect to the q_{1i}'s

The first stage of the problem, maximizing with respect to the vectors $q_2 \ldots q_K$ (and, therefore, $x_2 \ldots x_K$), is equivalent to maximizing the determinant of $Q^*_{i,j,\ldots,z}$. The optimal array should be orthogonal and contain values as large in absolute value terms as possible. The solution for the design of these *K*-1 attribute vectors is therefore to set them to their extreme limits according to *K*-1 arbitrarily chosen columns of the familiar, 2^m orthogonal main effects design, for example the columns x_1 , x_2 and x_3 from Table 1.

Recall that under a choice framework, $x_2 \ldots x_K$ refer to attribute level differences. To maximize these differences, not only is an attribute level placed at one of its extreme points, but the level of the same attribute in the base alternative is placed at its opposite extreme. So, when the design calls for the level of one attribute to be +1, the level of the same attribute in the base alternative is placed at –1, and vice versa.

Once the solutions for the *K*-1 attributes have been established, the second stage of the maximization problem, maximizing with respect to the vectors q_{1i}, is qualitatively similar to the problem of optimal design for a binary choice model with one variable. The determinant alone would be maximized by setting the design points at their extremes, where probabilities go to 0 or 1; and the $P_i(1-P_i)$ components are maximized in the middle range, where $P_i = .50$ for all *i*.

Taking the first order conditions for an arbitrary design point, q_{1j}, gives:

$$\frac{(1-e^{q_{1j}})}{(1+e^{q_{1j}})} = -2\frac{\left|\Theta^+_{1j}\right|}{\left|\Theta\right|} \tag{20}$$

where Q_{1j}^+ represents the signed (1,*j*) cofactor of *Q*. The optimal solutions for q_j are derived numerically using the *FindMinimum* (to minimize the negative of the determinant) command in Mathematica 3.0.

The optimal solutions for q_j for the cases of two, four and eight attributes are derived numerically and displayed in Tables 2 through 4. These particular cases are chosen because they are each associated with unique fractional factorial designs. For attributes between these numbers, the appropriate design arrays are simply reduced versions of the ones displayed here. For example, if a researcher has three attributes, the design array would be drawn from the array for *K* = 4: Table 3. For five, six or seven attributes, the design would be drawn from Table 4.

What is particularly pleasing about the design solutions is how closely they resemble standard 2^K fractional factorial designs. The optimal solutions for all attributes but one follow the 2^K main effects orthogonal design exactly, modified to accommodate the assumed upper and lower bounds of the attribute levels. The final attribute, x_1, is used as a manipulator, to balance choice sets to achieve certain response rate splits, depending on the number of attributes in the experiment.

The optimal designs in these tables have several interesting features. First, the optimal solutions for the q_{1i}'s are all equal within each design. This results in the predicted response probabilities (displayed in the final column in each table) being the same for each choice set. The q_{1i}'s move inward, toward zero, as *K* increases. As this happens, the response probabilities move toward (but does not get too close to) utility balance. For the case of two attributes, the response

split is 82/18, or 82 percent of the sample choosing alternative 1 in the first choice set and 18 percent choosing the base alternative, 0. Moving to four attributes, this response split goes to 74/26. With eight attributes, the split is 67/33, or a two-thirds / one-third split.

Note that, although the q_{1i}'s are equal, the levels for the optimal x_{1i}'s, which can be derived algebraically from the optimal q_{1i}'s, the levels of the other attributes, and the true parameter vector, differ across the choice sets.

Although the optimal levels for the x_{1i}'s appear complicated, and basically, impossible to derive before conducting the study, a sequential procedure can be used, as suggested in the previous subsection, to adjust these levels according to whether the empirical response rate splits are above or below the optimal splits. Steffens et al. (2000) conducted such a study in Michigan and achieved success in improving the efficiency (on average) of the parameter estimates and increasing the determinant of the information matrix substantially. The experiment used in-person interviews of birders, offering each of the sixty interviewees eight different binary choices of birdwatching sites with six attributes, including an entrance fee. The entrance fee was chosen to be the balancing variable. This experiment represents the first attempt to implement the D-optimal multivariate binary choice designs in practice. The approach worked well and was not too burdensome on the researcher. Of course, as with any first attempt, a number of lessons were learned that can guide future attempts to implement optimal design. In particular, the empirical response rates forced the researcher to move many of the fees to their highest reasonable values. Even at these high values, the optimal response rates were not always achieved. In such cases, perhaps it would be best to employ a second balancing variable.

C-Optimality

C-optimality refers to the minimization of a function of the model parameters. Using the delta method, the asymptotic variance (*avar*) of a function of the model parameters, $g(B)$, is:

$$a\,\mathrm{var}(g) = g'(B)'\,I(B)^{-1}\,g'(B) \tag{24}$$

where $g'(B)$ is the vector of derivatives of g with respect to the parameter vector, B. Preserving generality, let $g'(B) = \{ g_1, g_2, \ldots, g_k \}$. Using matrix differentiation, the first order conditions for minimization can then be expressed as:

$$\frac{\partial\,\mathrm{var}(g)}{\partial x_{ij}^{q}} = g'\,I^{-1}\,\frac{\partial I}{\partial x_{ij}^{q}}\,I^{-1}g = 0 \tag{25}$$

for all observations, *i*, attributes, *j*, and alternatives, *q*.

Because equation 25 is a quadratic form, and I is symmetric, the first order conditions can be re-expressed as:

$$\frac{\partial\,\mathrm{var}(g)}{\partial x_{ij}^{q}} = \sum_{l=1}^{k}\sum_{m=1}^{k}\left(g_1 I^{1l} + g_2 I^{2l} + \ldots + g_k I^{kl}\right)\left(g_1 I^{1m} + g_2 I^{2m} + \ldots + g_k I^{km}\right)\frac{\partial I_{lm}}{\partial x_{ij}^{q}} = 0 \tag{26}$$

where I_{lm} is the (l,m) element of I and I^{yz} is the (y,z) element of I^{-1}, which is equal to the signed (y,z) minor of I divided by the determinant of I.

A specific C-optimal criterion can be specified through the function, g. Two different functions will be considered here. The first is a marginal WTP between two attributes as shown in equation 4. Without loss of generality, letting attribute l be attribute 1 and attribute m be attribute 2, the derivative vector is $g' = \{-b_2 / b_1^2, 1 / b_1, 0,\ldots,0\}$. The asymptotic variance of $-b_2 / b_1$ is:

$$\text{var}\left(\frac{b_2}{b_1}\right) = \frac{1}{b_1^2}\left[\left(\frac{b_2}{b_1}\right)^2 {}^{11} + {}^{22} + \left(\frac{b_2}{b_1}\right) I^{12}\right] \tag{27}$$

The second function of interest is total WTP for a specific attribute bundle, x, as shown in equation 3. For this criterion, the derivative vector is $g' = \{-(b_2x_2+b_3x_3+\ldots+b_kx_k) / b_1^2,\ x_2 / b_1,\ x_3 / b_1, \ldots,\ x_k / b_1\}$.

C-Optimality for the Linear Model

As with the development of the D-optimal designs, the presentation of C-optimality will begin with the linear model. It will turn out that the results for this case will exactly match those for the binary model. Under the linear model, the first order conditions reduce to:

$$\frac{\partial \text{var}(g)}{\partial x_{ij}} = 2\left(g_1 I^{1j} + g_2 I^{2j} + \ldots + g_k I^{kj}\right)\sum_{m=1}^{k}\left(g_1 I^{1m} + g_2 I^{2m} + \ldots + g_k I^{km}\right)x_{ij} = 0 \tag{28}$$

for all observations, i, and attributes, j.

For each j, the summation within equation 28 is the relevant part of the first order conditions for C-optimality. Now the first order conditions can be simplified to:

$$\left(g_1(-1)^{1+j}\left|X_{1j}^{-\prime} X\right| + \ldots + g_k(-1)^{K+j}\left|X_{Kj}^{-\prime} X\right|\right) = 0 \tag{29}$$

where X_{kj}^{-} is equal to the X matrix with the j^{th} row replaced by a row of zeros and a one in the k^{th} column.

Looking, first, at the case where $g(\text{B}) = b_2 / b_1$, the first order conditions reduce further to:

$$\frac{1}{b_1^2}\left((-1)^{1+j} b_2 \left|X_{1j}^{-\prime} X\right| + (-1)^{2+j} b_1 \left|X_{2j}^{-\prime} X\right|\right) = 0 \tag{30}$$

or,

$$\frac{\left|X_{2j}^{-\prime} X\right|}{\left|X_{1j}^{-\prime} X\right|} = \frac{b_2}{b_1} \tag{31}$$

for all attributes, *j*. Recall that when both matrices inside the determinant are square matrices (as when we assume that the number of distinct design points is no greater than *K)* the determinant of the product is equal to the product of the determinants. The first order conditions, therefore, simplify to:

$$b_1 \left| X_{2j}^{-} \right| = b_2 \left| X_{1j}^{-} \right| \tag{32}$$

Turning to the second C-optimal criterion, the minimization of the asymptotic variance of total WTP, the first order conditions are:

$$\frac{1}{b_1^2}\left((-1)^{1+j}(b_2 x_2 + \ldots + b_K x_K)\left|X_{1j}^{-\prime} X\right| + (-1)^{2+j} b_1 x_2 \left|X_{2j}^{-\prime} X\right| + \ldots + (-1)^{K+j} b_1 x_K \left|X_{Kj}^{-\prime} X\right|\right) = 0 \tag{33}$$

for all *j*. The solution to this set of first order conditions requires:

$$\frac{(-1)^{i+j}\left|X_{ij}^{-}\right|}{(-1)^{1+j}\left|X_{1}^{-}\right|} = \frac{b_j}{b_1} \tag{34}$$

for all $i = 1,\ldots, K$ and $j = 1,\ldots,K$.

Linear Model with One Independent Variable

With this model, there is only one estimator to be considered: $-\alpha/b$. Looking only at two observations, the first order conditions are:

$$a \begin{vmatrix} 1 & 0 \\ 1 & x_2 \end{vmatrix} = b \begin{vmatrix} 0 & 1 \\ 1 & x_2 \end{vmatrix}$$

and

$$a \begin{vmatrix} 1 & x_1 \\ 1 & 0 \end{vmatrix} = b \begin{vmatrix} 1 & x_1 \\ 0 & 1 \end{vmatrix}$$

The C-optimal solution is:

$$x_i = -\frac{a}{b} \tag{35}$$

for all observations, *i.* The optimal location of the only attribute in linear model is exactly at the point of the estimator itself. Unfortunately, this solution disallows estimation of a regression line, as there is only one point for which data is collected. That point, however, is exactly the point that has

been identified as the point of most interest by the C-optimal criterion. The solution implies that the researcher would simply collect data at the exact point of interest and forgo estimating the model. Additional comments on this solution are provided at the conclusion of this section.

Multivariate Linear Model

Now consider the case of a two attribute model. The first order conditions are:

$$b_1 \begin{vmatrix} 1 & 0 \\ x_{21} & x_{22} \end{vmatrix} = b_2 \begin{vmatrix} 0 & 1 \\ x_{21} & x_{22} \end{vmatrix}$$

and

$$b_1 \begin{vmatrix} x_{11} & x_{12} \\ 1 & 0 \end{vmatrix} = b_2 \begin{vmatrix} x_{11} & x_{12} \\ 0 & 1 \end{vmatrix}$$

The C-optimal solution is:

$$x_{i1} = -\frac{b_2}{b_1} x_{i2} \tag{36}$$

for all observations, *i*. This solution is analogous to the solution for the univariate model. It implies perfect collinearity between the two columns of the *X* matrix and disallows estimation of a regression line.

For larger numbers of attributes, it can be shown that the optimal placement of all attributes, 3-*K*, is at zero, for all observations.[6] In other words, the attributes are dropped from estimation all together. Essentially, additional attributes cannot improve the information provided by the relevant attributes; they can only increase the overall variance by adding parameters to the model and reducing degrees of freedom.

To minimize the variance of b_2/b_1, the full design solution for the multivariate case is, therefore, to remove all attributes but the two relevant ones from estimation all together. The remaining two would be placed at their exact point of indifference, as in equation 36.

The conditions for the solution to the second C-optimal criterion are provided by equation 34. These conditions are a multivariate version of the conditions for the first C-optimal criterion and imply that every attribute be placed at its exact point of indifference between itself and Attribute 1. With this solution, every column is collinear and, again, model estimation is impossible.

[6] The demonstration is not provided here to save space but follows by solving the first order conditions.

C-Optimal Design for the Binary Logit Model

For the binary model, the first order conditions in equation 26 reduce to:

$$\frac{\partial\, a\ \mathrm{var}(\ \ g\ )}{\partial\, x_{ij}} = \ p_i(1-\ p_i)\sum_{m=1}^{k}\left(g_1 I^{1m} + g_2 I^{2m} + \ldots + g_k I^{km}\right)x_{im} \ *$$

$$\left[2(g_1 I^{1j} + g_2 I^{2j} + \ldots + g_K I^{Kj}) + (1-2p_i)\sum_{l=1}^{K} g_1 I^{1l} + g_2 I^{2l} + \ldots + g_K I^{Kl})x_{il}\right] \qquad (37)$$

Similar to the linear model, the numerator of the summation is the relevant section of the first order conditions. Converting the conditions from the linear model to the binary model gives:

$$\frac{\left|\bar{X}_{ij}'P(I-P)X\right|}{\left|\bar{X}_{1j}'P(I-P)X\right|} = \frac{b_j}{b_1} \qquad (38)$$

where i=1,…,k and j=2 for the first C-optimal criterion and i=1,…,k and j=2,…k for the second C-optimal criterion. Since P is a square matrix, the first order conditions in equation 38 simplify to exactly those in equation 34. The C-optimal design solutions are identical for the linear and binary models.

Unfortunately, as with the linear model, the design solutions produce a practical dilemma. By generating a dataset where every observation is perfectly balanced, one generates a dataset with a multicollinearity problem. Specifically, one attribute must be used to perfectly offset the utility contribution of the other attributes. It is therefore impossible to estimate a model that follows the C-optimal design solution.

Short of taking the exact design solution, one might be tempted to assume that approximating the design solution would be a recommended approach for obtaining efficient estimates for marginal or total WTP. The author is not so sure. Approximating the design solution would presumably mean designing choice sets that are close to utility balance. There are two potential dangers to using this approach.

First, one would obviously be generating a near-multicollinear dataset. Although estimable, such a dataset might produce very high variances for the parameter vector, as it would with the linear model (Greene 1993). The C-optimal solution seems to suggest that such a dataset has the potential to estimate a ratio of parameters efficiently, even though the actual parameters might be estimated quite inefficiently. More likely, though, what it suggests is that when we are interested in estimating WTP, we are best off finding a way to estimate it directly, rather than as a ratio. In a sense, then, the C-optimal design solution is alerting us to the fact that our indirect approach to estimating WTP is inefficient. From a statistical perspective, it would, of course, be preferable to estimate WTP directly, rather than as a ratio of two estimated parameters. Without such a model, WTP cannot be estimated efficiently.

Second, choice sets with utility balance are probably cognitively difficult for experimental respondents. How does a person choose between two choices of which he or she is perfectly indifferent? Dellaert, Brazell and Louviere (1999) found that when alternatives within a choice set offer similar utility levels but contain large attribute differences, respondents can have a hard time distinguishing among them and identifying their most preferred. This might lead to heteroskedasticity among responses.

Conclusions

This paper has extended the literature on optimal design for nonmarket valuation experiments by deriving D- and C-optimal designs for binary choice experiments with multiple attributes. Between the two design criteria, the author finds that the D-optimal design recommendations are the most useful for practical applications. These designs place all attributes but one at their extreme points, leaving one attribute to balance response rates to their optimal levels. The C-optimal design solutions, although they are intended to optimize estimation of WTP, turn out to be impractical as real design solutions. They require that all observations be perfectly balanced at the point of indifference. Choice sets based on this idea will not provide enough information to separately identify all of the model parameters. Essentially, this result seems to illustrate the inefficiency of our approach to estimating WTP. From a statistical perspective, it is best to estimate WTP directly. Unfortunately, this is impossible with choice experiments.

Given this dilemma, it seems the most reasonable approach is to focus our efforts on estimating the best choice models possible. This means employing the D-optimal designs. Yes, this means our estimates of WTP will remain less efficient than should be theoretically possible. But the individual parameter estimates will be estimated as efficiently as possible, and given our preference for choice experiments as a way to indirectly understand WTP, it seems to be the most appropriate way to go.

There are a number of caveats that must always be mentioned with optimal design. First, the design solutions are always specific to the assumed model, in this case, logit with a linear utility specification. Since logit is by far the most popular model for choice experiments, this assumption is not so bad. Future research by the author will look at how higher order terms and cross-terms will affect optimal designs.

One caveat that is always mentioned when nonlinear (such as discrete choice) models are examined is that the optimal designs are always functions of the unknown parameters. Obviously, we can only guess at these values, perhaps using pre-test information, before conducting the experiment. This has always seemed to be the most serious flaw to this area of research and the reason that the results are rarely used in practice. However, as this paper describes, the optimal designs can be applied over the course of a sequential data collection process using nonparametric information. Specifically, researchers can update one or more attributes to move response rates for each choice set toward the optimal response rates. This procedure was tested in practice recently with positive results. Clearly, there is room for much more experimentation with such a process.

Finally, optimal design is a statistical analysis only. Optimal designs assume responses will be accurate and truthful. Humans are not always capable or willing to be so. There is a great need to

combine the design results in this paper with controlled experimental situations to see how they perform in practice.

Table 1: Fractional Factorial (Main Effects) Design for Three Two-Level Attributes

Main Effects			Two-way Interactions			Three-way Interactions
x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$
-1	-1	+1	+1	-1	-1	+1
-1	+1	-1	-1	+1	-1	+1
+1	-1	-1	-1	-1	+1	+1
+1	+1	+1	+1	+1	+1	+1

Table 2: D-Optimal Design for 2 Attribute Binary Choice Experiment

Observation	Alternative	q_1^*	z_1^*	z_2^*	P
1		1.54	$(1.54 - b_2) / b_1$	+1	.82
1	0	0	b_2 / b_1	-1	.18
2	1	1.54	$(1.54 + b_2) / b_1$	-1	.82
2	0	0	$- b_2 / b_1$	+1	.18

Table 3: D-Optimal Design for 4 Attribute Binary Choice Experiment

Observation	Alternative	q_1^*	z_1^*	z_2^*	z_3^*	z_4^*	P
1	1	1.04	$(1.04 - b_2 - b_3 - b_4) / b_1$	+1	+1	+1	.74
1	0	0	$(b_2 + b_3 + b_4) / b_1$	-1	-1	-1	.26
2	1	1.04	$(1.04 + b_2 - b_3 + b_4) / b_1$	-1	+1	-1	.74
2	0	0	$(- b_2 + b_3 - b_4) / b_1$	+1	-1	+1	.26
3	1	1.04	$(1.04 + b_2 + b_3 - b_4) / b_1$	-1	-1	+1	.74
3	0	0	$(- b_2 - b_3 + b_4) / b_1$	+1	+1	-1	.26
4	1	1.04	$(1.04 - b_2 + b_3 + b_4) / b_1$	+1	-1	-1	.74
4	0	0	$(b_2 - b_3 - b_4) / b_1$	-1	+1	+1	.26

Table 4: D-Optimal Design for 8 Attribute Binary Choice Experiment

Observa-tion	Alterna-tive	q_1^*	z_2^*	z_3^*	z_4^*	z_5^*	z_6^*	z_7^*	z_8^*	P
1	1	.72	+1	+1	+1	+1	+1	+1	+1	.67
1	0	0	-1	-1	-1	-1	-1	-1	-1	.33
2	1	.72	+1	+1	+1	-1	-1	-1	-1	.67
2	0	0	-1	-1	-1	+1	+1	+1	+1	.33
3	1	.72	+1	-1	-1	+1	+1	-1	-1	.67
3	0	0	-1	+1	+1	-1	-1	+1	+1	.33
4	1	.72	+1	-1	-1	-1	-1	+1	+1	.67
4	0	0	-1	+1	+1	+1	+1	-1	-1	.33
5	1	.72	-1	+1	-1	+1	-1	+1	-1	.67
5	0	0	+1	-1	+1	-1	+1	-1	+1	.33
6	1	.72	-1	+1	-1	-1	+1	-1	+1	.67
6	0	0	+1	-1	+1	+1	-1	+1	-1	.33
7	1	.72	-1	-1	-1	+1	-1	-1	+1	.67
7	0	0	+1	+1	-1	-1	+1	+1	-1	.33
8	1	.72	-1	-1	+1	-1	+1	+1	-1	.67
8	0	0	+1	+1	-1	+1	-1	-1	+1	.33

References

Abdelbasit, K.M. and R.L. Plackett, 1980,"Experimental Design for Binary Data," *Journal of the American Statistical Association,* 87: 381, 90-98.

Adamowicz, W.L., J. Louviere and M. Williams, 1994, "Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities," *Journal of Environmental Economics and Management,* 26, 271-92.

Alberini, A., 1995, "Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-bound, Double-Bound and Bivariate Models," *Journal of Environmental Economics and Management,* 28(3), 287-306.

Alberini, A. and R.T. Carson, 1993, "Efficient Threshold Values for Binary Discrete Choice Contingent Valuation Surveys and Economic Experiments," Resources for the Future Discussion Paper, Quality of the Environment Division, Washington, D.C.

Bateman, I.J. and K.G. Willis, Eds., 1999, *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the U.S., E.U. and Developing Countries,* Oxford University Press, Oxford.

Ben-Akiva, M. and S. Lerman, 1985, *Discrete Choice Analysis,* MIT Press, Cambridge.

Carson, R., J. Louviere, D. Anderson, P. Arabie, D. Bunch, D. Hensher, R. Johnson, W. Huhfeld, D. Steinberg, J. Swait, H. Timmermans, J. Wiley, 1994, "Experimental Analysis of Choice," *Marketing Letters,* 5, 351-67.

Cummings, R.G., P. Ganderton and T. McGuckin, 1992, "Substitution Biases in CVM Values," Comment No. 61 submitted to the National Oceanic and Atmospheric Administration Blue Ribbon Panel on Contingent Valuation, December.

Dellaert, B.G.C., J.D. Brazell and J. J. Louviere, 1999, "The Effect of Attribute Variation on Consumer Choice Consistency," *Marketing Letters,* 10:2, 139-47.

Fedorov, V.V., 1972, *Theory of Optimal Designs,* New York: Academic Press.

Greene, W.H., 1993, *Econometric Analysis,* Second Edition, MacMillan Publishing Company, New York.

Hoehn, J.P. and J.B. Loomis, 1993, "Substitution Effects in the Valuation of Multiple Environmental Programs," *Journal of Environmental Economics and Management,* 25, 56-75.

Huber, J. and K. Zwerina, 1996, "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research,* 307-17.

Jones, C. And W.M. Hanemann, 1996, "Theoretical Foundations of Resource Compensation," Paper presented at the W-133 Meetings of the U.S.D.A.

Kanninen, B.J., 1993a, "Design of Sequential Experiments for Contingent Valuation Studies," *Journal of Environmental Economics and Management,* 25: S-1-11.

Kanninen, B.J., 1993b, "Optimal Experimental Design for Double-Bounded Dichotomous Choice Contingent Valuation," *Land Economics,* 69(2): 138-46.

Kanninen, B.J., 2000, "Optimal Design for Multinomial Choice Experiments," Currently under review.

Louviere, J.J, 1988, *Analyzing Decision Making: Metric Conjoint Analysis,* Sage Publications, Inc., Newbury Park, CA.

Louviere, J.J. and G. Woodworth, 1983, "Design and Analysis of Simulated Consumer Choice of Allocation Experiments: A Method Based on Aggregate Data," *Journal of Marketing Research,* 20, 350-67.

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

Mathematica 3.0, 1996, Wolfram Research, Inc., Champaign, IL.

McFadden, D., 1974, "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics,* Zarembka, P., ed., Academic Press, New York, 105-42.

Minkin, S., 1987, "Optimal Designs for Binary Data," *Journal of the American Statistical Association,* 82:400, 1098-1103.

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

Nyquist, H., 1992, "Optimal Designs of Discrete Response Experiments in Contingent Valuation Studies," *Review of Economics and Statistics,* LXXIV(3), 559-562.

Opaluch, J.J., S.K. Swallow, T. Weaver, C.W. Wessels and D. Wichelns, 1993, "Evaluating Impacts from Noxious Facilities: Including Public Preferences in Current Siting Mechanisms," *Journal of Environmental Economics and Management,* 24, 41-59.

Silvey, S.D., 1980, *Optimal Designs,* London: Chapman & Hall.

Steffens, K, F. Lupi, B. Kanninen and J.P. Hoehn, 2000, "Implementing an Optimal Experimental Design for a Binary Choice Experiment: An Application to Bird Watching in Michigan," mimeo.

Winer, B.J., D.R. Brown and K.M. Michels, 1991, *Statistical Principles in Experimental Design*, Third Edition, McGraw-Hill, Inc., New York.

Wu, C.F.J., 1988, "Optimal Design for Percentile Estimation of a Quantal Response Curve," in *Optimal Design and Analysis of Experiments*, Y. Dodge, V. Fedorov and H. Wynn, eds., Elsevier Science Publishers.

CONSTRUCTED PREFERENCES AND ENVIRONMENTAL VALUATION

Presented by John Payne, Duke University
Co-authored with David Schkade, University of Texas, Austin

Summarization

Dr. Payne began his presentation saying he would focus not on theory but on data. He and his colleagues' research is based on an idea in the NOAA report suggesting that when you think about willingness to pay (WTP) or any kind of valuation issue, it is important that respondents think about substitutes and budget constraints in generating WTP answers. There is a strong statement in the report that in contingent valuation studies researchers need to remind people explicitly about substitutes and budget constraints. One way to do that is to give people an opportunity to value not just one good but several in a bundle or set of goods.

Whenever you do this, though, you raise issues of context effects, he cautioned. These raise concerns about to what extent does the value of a good as part of a bundle differ from the value of the good by itself and to what extent does the value depend on where the good is valued in a sequence of goods. These are the kinds of issues that he, David Schkade, and Bill Desvousges have been working on. Dr. Payne said he would talk first about the study and the results and then about the implications for accessing values for purposes of cost-benefit analysis or other uses in deciding policy.

In their study, they presented people with a series of five environmental goods and asked them to evaluate all five. This was done across two sessions. In one session people were asked to give a WTP response and in the other to evaluate the goods by answering attitude questions. The researchers looked at how those valuations differed according to the order in which the goods were evaluated and found strong evidence of serial order effects. Serial order effects occur when a good evaluated first in a series receives a much higher value than a good evaluated later. This is not inconsistent with economic theory and the ideas of substitutes and budgets, Payne said. What was interesting to them, though, was that they found a big effect between the valuations of a good that was first in a sequence and all the other goods anywhere else in that sequence.

In a sequence there can be substitute effects and budget effects (the further out in a sequence you go, the more you have spent). The question they were interested in was, can you look at not only the sequence effects but at what happens to the total value of the bundle of goods? If you look at the sum across all of the goods, is that sum dependent on the order in which you do things? As we will show, said Payne, it does.

They hypothesize that what is creating this effect is, that in doing a contingent valuation, people are being put in a position where they need to construct a response. That construction often depends on the first answer and that answer drives everything else.

In their study, they used five goods, selected both because they seemed interesting and because they had been used in other contingent valuation studies. One was visibility improvements

in the Grand Canyon; another was something they had worked on before, providing protection to migratory birds in the Central Flyway. The other three were salmon protection in the Northwest, oil spill prevention, a major environmental issue, and the reintroduction of the red wolf into the Great Smokies. Two of the goods were chosen for their proximity to the survey groups. They chose the birds in the flyway as a good because half the people in the survey where in Texas; the red wolf, because half were in North Carolina. This helped when they looked at whether distance mattered in terms of use value. Information on each good was provided in both text form and through pictures. The researchers told people that they would be asked to express values on environmental programs, or goods. They also told them early on and explicitly that they would be valuing five goods in all. They next presented information on each good and checked to make sure people understood it.

They conducted the survey over two sessions separated by a two-week interval. Half the people did the WTP-related questions in the first session and the attitude questions in the second. The other half did the evaluations in the reverse order. They randomized the order of the five goods for each respondent. The same random order was used for both sessions. In the WTP session, they asked people how confident they were of the numbers they had supplied in their WTP responses, their views of the likelihood of success of the programs, and demographic questions. In the second session, they presented people with the same five goods and the same information about them, but instead of WTP questions, they asked attitude questions such as how important is the problem, how serious, what is the good's use value, and what is its importance for future generations. They then asked people to do a rank ordering of the goods in terms of importance. Payne stressed that at the beginning of the survey they told people that there would be five goods, so people who had done the first session by the second session were aware of this and also had a lot of information about the goods.

Dr. Payne showed some of their results, organized by whether respondents got the contingent valuation first or the rating task first. He pointed out the WTP amounts for goods when they were in the first position — for air when air was the first good in the sequence, for birds when birds was first, etc. This is a classic design, he said, where you give people a single good to evaluate and you get a response. A general effect, which has been found before, is that the WTP for a good when it is the first good in a sequence is much higher then the WTP for that same good if it comes later in the sequence. The means and the medians of WTP amounts for serial position show the first position to be valued much higher than the other positions. This effect holds whether people do the WTP responses first or the ratings first. The bottom line seems to be that reminding people that there are five goods, even letting them see the five goods, is not sufficient to get away from the sequence effect. What seems to matter is that people have to go through the process of assigning a value. Once they have done that, it becomes real to them that there are budget constraints and substitutes.

The idea of substitutes and budget constraints is consistent with the serial position effect. It is interesting, Payne noted, that from a design perspective you see the effect between just the first and second positions. It suggests that if you want people to consider substitutes and budget issues in valuing a good you might want to have them value another good before they value the one you are interested in.

Substitutes and budget constraints have an impact on valuation as a function of sequence position. But what happens at the end, after you have valued all five goods? By the time you have

gone through all five, substitute effects and budget effects should have combined and washed out. So one prediction you can get from economic theory is that, while the value of a good will vary with its position, by the time you have done all five goods the value of the sum of the five goods should be essentially the same. What they found, said Payne, is that they are not.

If, he said, you start the task by valuing a good that is higher in value, like oil spills, you end up with all other goods being given higher WTP values. They looked at the effect on oil spill values, their highest valued good, when it was in the first position followed by the red wolf, which was their lowest, in the second position. They also looked at how the wolves were valued if they were in the first position and oil spills in the second position. Holding those two goods constant, they then looked at how that order affected the sum of the WTP for the other three items. Their data showed that the values assigned to air, birds, and salmon were much higher if the first good valued in the sequence was oil spills than if it was wolves. Interestingly, if a relatively low valued good was second to a relatively high valued good in the first position, it received a higher WTP.

Their tobit analysis results show the same sort of effects. People did discriminate among the goods, particularly between oil spills and the other goods. The demographic effects seem to be consistent with the literature: people's WTP went up with income, females were willing to pay more, and there were marginal effects for age. The tobit analysis results also confirm the serial order effect — the first good was valued at a much higher WTP than the other goods.

Dr. Payne said that he, Dr. Schkade, and others have argued that when you get a response to a WTP question or a contingent valuation question, you are getting a constructed response, a number that, in some sense, is made up at the time the question is asked. One view is that this partly accounts for why you get procedural variance (how you ask the question matters), descriptive variance or framing effects (how you present information and describe problems matters), and what they call context effects (the order in which you do things matters). So, they argue, a lot of those effects are due to the fact that people are constructing responses. That raises the question, is there anything there at all to be measured? Is it all constructed or are there any stable core values?

One of the things they did in their study was compare WTP responses with a variety of other attitude measures. Looking at the mean responses, what struck them was that there indeed seemed to be something there. Whether they looked at WTP or importance measures, etc., there was evidence that oil spills, no matter how they asked the question, was consistently valued more highly than the other goods. Because they had five goods, they were able to look at the relationship among responses within an individual across the five goods to see if there were any stable core values. Looking at the mean correlations of responses, they found indications of stable core values but the WTP responses were actually the less good why of getting at those numbers. It is not that there is nothing there; for example, WTP does relate to the final ranking but not as well as some of the others. Another way to get at this is to look at comparisons across what proportion of the variance is explained by the goods across different ways of measuring value. There is some variance being explained by the goods, such as WTP, but the attitude measures are capturing more of it.

Their conclusions are that they found two strong sequence effects in terms of valuing across a set of goods. They argue that these sequence effects suggest caution when using dollar amounts as measures of the economic value, in any absolute sense, of a set of goods. They found a strong serial position effect that was concentrated on the difference between being first in a sequence and being

later. The sequence effect was similar for both response modes. This suggests that simply reminding people of substitutes and budget restraints may not be sufficient. You may need to have people go through a prior evaluation exercise to get them to internalize those issues. The total WTP amount for a bundle of goods is not invariant to the valuation sequence. And in fact the effect is consistent with other literature, in psychology and in other areas, of anchoring effects. The first response can be defined as what in psychometrics is called a modulus or is sometimes called an anchor value, where all valuations are related to that first number. These effects are not uncommon in a lot of work in psychometrics. The effects reflect the cognitively difficult task they were giving people.

While they found strong context effects, they believe that there are some regularities — stable values or attitudes that are better viewed, not as economic values, but as expressions of attitudes. Their view is that you must consider expressions of values or attitudes and the two sources of systematic variance, as well as random error or noise. The first are the stable values associated with the attributes of an object, the second, the systematic effects due to the nature of the task, (how you ask the question, describe the problem, etc.). Those task and context effects are predictable because they result from the interactions between the properties of human cognition and the nature of specific tasks. They are systematic biases and predictable. They argue that in tasks involving things like the contingent valuation of unfamiliar environmental goods, task and context effects are often as large or larger than those of stable core values or random error.

This does not mean that there is not value to doing good experimental design or to providing good incentive structures but, Dr. Payne argued, having done those things, there will still be situations where task and context effects are large. Therefore, researchers need to acknowledge this and, perhaps, change their approach to valuation. The approach needs to change in a way that recognizes the psychology of people's judgments and provides them with tools and techniques to better construct values. Reminding people about substitutes and budgets in a way that they internalize the information is a device for helping people construct better attitudes and preferences. Researchers know something about how to do that and should be using that knowledge in their valuation techniques, he concluded.

Dr. Payne added that the profession needs to recognize that there are limits to what people can give researchers and they need to develop systems that recognize those limits. Perhaps the approach should be, acknowledging that all that people can give researchers are attitudes, that those attitudes can provide relative importance across goods and can be mapped onto dollar values, using techniques such as damage schedules, for use in cost-benefit analysis. He cautioned that researchers should do so recognizing that people are neither totally dumb nor are they super people but instead recognizing that some people can provide some information that can be used to build valuations.

Discussion of Session I Papers

by Julie Hewitt, US EPA, National Center for Environmental Economics

I have four papers/three presentations to discuss, and in more ways than one, the authors have made my job easy. First, all of them are well written, and straightforward to follow. I'll discuss them in turn, and then offer some concluding comments.

First the Carson paper. Richard and his co-authors start with two oft-cited reasons why some economists dislike SP surveys: the hypothetical nature and the possibility for respondents to respond strategically. In their paper, they address both, though the second requires more of the paper, and this is important. That is, rather than merely providing a list of reasons why respondents would not act strategically, they go a step further and ask, under what conditions would we expect respondents to act strategically, and what effect does such strategy have on their responses? If we understood the answer to this last question, could we not simply build the strategic behavioral rules into a structural model, rather than be left with a reduced form model with the strategic behavior built into it? This trio of authors, with plenty of expertise in the areas of stated preference, mechanism design and utility theory, have done quite a service in addressing the strategic behavior questions in a true discussion paper, with no equations. I hope the next step is the empirical application, complete with econometric details.

A seemingly specific point about word choice: throughout, people are referred to as agents, short for economic agents; I think a better term would be actors, short for economic actors, and I suggest this as a way to be clear about who these respondents are: they are not agents in a principal/agent sense, for they are more than that; they are simultaneously principals and agents. They are principals in the sense that it is their tastes and preferences we are interested in understanding; they are agents in the sense that they may be flawed representatives of the principals and their tastes and preferences. This is a point that is raised in a 1992 volume edited by George Loewenstein and Jon Elster called *Choice over Time*. That volume discusses the variety of observed behaviors that appear to be prima facie evidence of irrationality, and offers a variety of explanations as to how such behavior could indeed be rational.

I want to highlight one point they make which is related to a thought that has been rolling around in the back of my brain in a not very articulate format. They also raise the issue of how well respondents deal with cost uncertainty. They raise this issue with the example of a respondent who does not believe that the cost offered to them (would you be willing to pay $X) is a realistic estimate of what the government would actually have to spend to provide the good in question. I have a small quibble with referring to this as a cost, since after all, a large portion of the environmental amenity that the government provides through the EPA is accomplished not through spending but through regulation. Nonetheless, their discussion reminded me that we should perhaps be thinking of survey respondents not as utility-maximizing actors, but as actors who are involved simultaneously in production and consumption of the same commodity. The precedent for this type of behavioral model is in the literature on agriculture in developing countries, where subsistence farmers are consuming the same good that they take to market. Their behavior can't be taken at face value as either consumption—in the traditional sense—alone or production alone, but is best modeled as utility maximization subject to a full income budget constraint, as suggested by Becker in his 1965 paper on the allocation of time. Perhaps we should think of survey respondents in a similar sort of fashion. For instance, we might expect that respondents would adjust their WTP responses

according to the source of the pollution, for a given level of pollution: if the source is comprised of a few firms with deep pockets, would we expect consumers to be WTP as much as they would if the source were many small firms that were the source of their neighbors jobs? This would lend a public choice flavor to the analysis of WTP, but this seems perfectly in keeping with the commonly used payment vehicles of SP surveys. Furthermore, it is not inconsistent with the notion for some of the SP formats covered in the Carson et al. paper that an individual's response depends on how they think others will respond.

In the section on continuous response formats, I find myself shocked, shocked, shocked to learn that there are decisions not made according to a true cost-benefit criterion.

They discuss sequence effects and I shall return to this point later.

From the standpoint of policy, this work does result in some clear guidance to EPA, in the sense that they have laid out SP formats that are incentive compatible versus those that are not. And the groundwork is laid for the next logical step from this research, which is the empirical application.

Now to the Kanninen paper. On average there are the right number of equations in these four papers, they just all happen to be in Barbara's paper! But in seriousness, Kanninen has extended earlier work on optimal design of experiments to the more recently employed SP valuation variants, those of multi-attribute binary choice models and of multinomial choice, or conjoint models. The idea of optimal design is a straightforward one: we have two choices to gain more confidence in our estimated models: one is to survey more respondents, and the other is to apply optimal experimental design, wherein the survey designer chooses the various thresholds to give respondents in a binary choice question, or how the attributes of the package vary in a conjoint survey.

I want to mention the number of equations in Kanninen's paper again, because I want to emphasize a point that anyone who is frightened away by the equations will miss. While optimal experimental design is of most use to researchers on a limited budget, the results here do not require more sophisticated techniques for estimation than what researchers are already using, nor do survey designers need to re-derive the equations here. That is, the results here are fairly precise prescriptions for optimal design that can be transferred to a broad range of SP experiments.

There is another point that Kanninen makes clearly regarding D-optimality, which focuses on gaining efficiency in the parameter estimates themselves, but I'll restate it in quiz format for emphasis:

If attribute A can take on any values between 0 and 10, what is the best set of values to present to respondents for the *linear* multi-attribute binary choice model?

A) 5

B) 0 and 10

C) 0, 5, and 10

D) 0, 3, 7, and 10

Without having been through this presentation, most of us would quickly rule out A (no variation at all); B seems OK, but if more is better, wouldn't C and D dominate B? And choosing between C and D is easy: with C, I can say I chose the low, medium, and high values, while D offers one more value but is much harder to describe without sounding a bit arbitrary. What this naïve approach ignores is that for the *linear* multi-attribute model, the effect of varying attribute A is linear, so there's only one coefficient to estimate, and there is nothing to be learned by using the midpoint: the effect of moving A from 0 to 5 is exactly the same as the effect from moving from 5 to 10, which is also double the effect of moving from 0 to 10. There's nothing to be gained from measurement at the midpoint, but every observation at the midpoint is one that's not at an endpoint, and that has a real cost in terms of precision in estimation. This is precisely the problem with ad hoc experimental design.

When considering C-optimality, which focuses on optimizing for WTP, things don't turn out quite so nicely. However, this is an extremely interesting finding. Correct me if I've misinterpreted this, but before Kanninen derived these results for the multi-attribute and multinomial models, there were both C-optimality and D-optimality results and a survey designer had to make a choice between the two, which may have seemed ad hoc and therefore been unsettling. But now, there's no choice to be made between C- and D-optimality, and a clear-cut argument for choosing model structure that estimates WTP directly.

I would suggest modifying the term response rate, because it has a different meaning here than in the usual survey context (how many respondents answer the entire survey, not specific questions).

And now to the Payne and Schkade papers. There are two, and so my comments may vary a bit from their presentation.

In their first paper (1999), they start by contrasting pre-existing and constructed preferences, noting that the latter typically apply in the case of SP surveys. The survey designer is essentially an architect guiding the construction process. And just as building codes protect householders, a building code for SP surveys would protect . . . maybe I'm pushing the analogy farther than is appropriate. Figure 1 in their paper gives a list of stages of construction, problems that occur in each stage and remedies. This list is extensive (8 stages, several problems per stage), and in the interest of time, I won't talk about the problems in which I largely agree with the solution.

Regarding myopic decision frames: the question here is why it is often observed that the willingness to pay for a package is less than the sum of WTP values for the components of the package, or WTP(A+B) < WTP(A) + WTP(B). Is this truly a problem? Having read the Carson, Groves and Machina paper, I'm now less concerned. It could well be reasonable for respondents to value A at 10, B at 15 and A+B at 22, where the package value incorporates an expectation of volume discounts. If I don't get B when you ask me about A, but then you ask me about both, I might think there should be a volume discount associated with the package, because after all there may be administrative economies of scale in providing these goods! Perhaps such responses are rational in this light.

Regarding manipulation checks, the checks would need to be carefully constructed, particularly in light of the Carson et al. paper.

I'd like to commend them on their words of caution section: they raise an idea attributable to Sunstein (1990) early in their paper: that not all preference expressions are created equally. This made me nervous, and as I read on, I thought, oh thank heavens they didn't go anywhere with this. I was poised to say, well this requires researcher judgement and of course, not all researcher judgements are created equally. Anyway, I just want to add my observation that I get very nervous when I hear survey designers say things such as, "we left that question out because their answers didn't make sense." For many of the items being valued, if we haven't got a good idea how individuals value something and need to ask them to state their preferences, then we likewise ought not to have a very strong prior about what their answers would be. Off the soapbox.

I want to end discussion of this particular paper on one of their final notes: constructed preferences are really future preferences, and we're not very good at prediction. I mentioned the *Choice over Time* book earlier, which gives a good example: we tend to say we will do more good deeds and reduce our bad habits, only to have a hard time holding ourselves to such resolutions. It seems to me that if there's any "strategic behavior" going on in SP surveys, it's this variety. Do people report WTP amounts in this fashion? Well, I can give a higher answer to WTP than my current budget would allow because I plan to work very hard this year and get a big raise next year, all motivated by wanting to be a better as evidenced by being able to contribute more to saving Mother Earth. It's not a particularly self-serving strategy, but does it lead to overstated WTP? And again, though I might wish to revisit one or two of their points in this paper, nonetheless they have given EPA some guidance regarding the proper conduct of SP surveys.

Now the fun part: the only empirical results I get to talk about, those of their forthcoming Payne et al. (2000) paper. I think they've offered a good bit of evidence in this paper that sequencing matters. This study seems to be motivated at least in part by two things: 1) the result that is frequently the case in empirical papers that $WTP(A+B) < WTP(A) + WTP(B)$, and 2) a theoretical result from Carson and Mitchell (1995) which I'll recast in similar notation as $WTP(A+B) = WTP(B+A)$, implying that order of presentation of programs does not matter when asking about WTP for the whole package. Now let a subscript on WTP denote the order in which the WTP question is asked; the two-program version of the Payne et al. (2000) result is that $WTP_1(A) + WTP_2(B) > WTP_1(B) + WTP_2(A)$ if $WTP_1(A) > WTP_1(B)$. What is clear when casting these results in similar notation is that the result of Payne et al. (2000) is a statement about a different notion of sequencing than that of Carson and Mitchell (1995). Furthermore, I am not convinced that these two results are inconsistent with each other.

In fact, I find it reassuring that WTP falls with the order. This also seems to be in concert with the ideas raised in the Carson et al. paper. Now that you're offering me more items, even though I get to bid on individual items, I now realize that of course I paid too much for the previous programs and will want to adjust downward my bids for later individual programs. But, does that justify a conclusion that it's the later WTP values that are closer to truth than the earlier WTP values? The question can be recast as, under what circumstances would later WTP values be closer to true WTP than earlier WTP values?

Once again, this research provides EPA with some pretty clear guidance: to get a lower WTP value, ask other WTP questions prior to the WTP question of interest. Of course, there is also the implicit advice on how to achieve higher WTP values!

I said earlier that these authors made my job easy in more ways than one, and I offer the final reason: I have the greatest confidence that they are the most expert to judge each other's papers, and so I ask:

Barb: is John and David's experimental design optimal?

Richard: should Barb consider incentive compatible formats in extending optimal design results?

John & David: do you think Richard probably didn't mean what he said in his 1995 paper with Robert Mitchell, that WTP(A+B) = WTP (B+A)?

Richard: can the mechanism design approach be applied to John and David's paper?

References

Carson, Richard T. and Robert Cameron Mitchell. 1995. "Sequencing and Nesting in Contingent Valuation Surveys," *Journal of Environmental Economics and Management*, 28(2): 155-173.

Loewenstein, George and Jon Elster, editors. 1992. *Choice Over Time*, New York: Russell Sage Foundation.

Discussion of Session I Papers

by John K. Horowitz, University of Maryland
Horowitz@arec.umd.edu

In her paper on optimal design of choice experiments, Kanninen notes that in order to make the most powerful inferences, researchers should choose explanatory variables that are spread as far from each other as possible. Such a design has been achieved in this session. These three papers contain 163 citations of which only one – Mitchell and Carson's contingent valuation book – is common to all three.[1] A truly optimal design has been achieved, and my purpose here is to see what it reveals.

I. The One-Shot Dichotomous Choice Question

Carson, Groves, and Machina (CGM) ask whether the results derived from a vast array of stated preference surveys are conceivably compatible with economic theory. They conclude that the answer is yes. Because of strategic properties of various survey questions and because of respondents' plausible beliefs about those surveys, many response patterns that are often puzzled over are shown not to shed much doubt on stated preferences (SP). CGM conclude that the body of survey work does not, as yet, provide a compelling reason to drop the underlying economic model. Much of their paper is thus focused on what that model implies for survey responses in a wide variety of circumstances.

In this paper, I will take it at face value that CGM have identified the essential issues and drawn the correct conclusions about the literature. My task here will be to tease out the implications of their arguments.

(Some readers will want me to take a different tack. I could investigate whether CGM have drawn correct conclusions about the literature; that is, whether their analysis is correct and the set of papers they have examined complete. Alternatively, I could question whether their economic model – based on a blend of agency theory and theory of the consumer, especially choice under uncertainty – is sufficiently specific that its hypotheses are conceivably falsifiable. I leave it to other reviewers to broach these arguments. Note that CGM do not pursue all of the possible implications of the "economic maximization framework" but focus on those based on incentives and information.)

As I read the paper, I see only one kind of elicitation method that might reasonably be said to elicit "true preference" in the kinds of situations that EPA and other environmental economists must address: the one-shot dichotomous choice question. Here is an example based on Hagen *et al.*:[2]

Q1. "If adopting the [spotted owl] conservation policy would cost your household $28.00 per year (for the foreseeable future), would you vote YES or NO?"

My impression is that most researchers believe this type of question is robust, perhaps even unassailable, for pure public goods. They believe this partly on the basis of CGM-type arguments

[1]The papers are Carson, Groves, and Machina; Kanninen; and Payne, Schkade, Desvousges, and Aultman.
[2]My version removes an ambiguous cost statement from the original Hagen *et al.* question. I further recommend removing "if."

and partly on experience and intuition. Researchers may also rely on this sort of question for a more fundamental reason: Since this question is essentially the choice we face as a society, how could it be wrong for us to ask it?

But the properties of the one-shot dichotomous choice question deserve their scrutiny as well. Here are some of the issues:

A. The question does not state how responses will be used.

CGM note that not describing how a subject's response will be used is a serious problem with most open-ended questions, but they do not explore the ways this omission could contaminate closed-ended questions.

Suppose that researchers intend to use the responses to estimate the median willingness-to-pay (WTP). In this case, the one-shot dichotomous choice question is incentive compatible.

Suppose on the other hand that researchers intend to use responses to estimate *mean* WTP, the more common approach. Estimating mean WTP requires the researcher to vary the policy's stated cost across respondents and then calculate the implied distribution of WTP. In this case, one of two problems must arise. Either the researchers must lie about the policy's costs to the respondents *or* the costs must be randomly distributed across the population. Both of these conditions present serious problems.

To see the first problem, suppose I, as a respondent, know that the average cost for the policy in Q1 is actually $20. If my valuation is above $20, then I will say yes to Q1 even when my true value is below the stated cost of $28, since a yes response increases the probability that the estimated mean will be above $20.[3] Thus, the question is not incentive compatible at the stated cost.

Note that this response strategy does *not* depend on my knowing the true cost exactly; I need only believe that there is some probability that the true cost is below my valuation for me to have an incentive to say yes when sometimes my "true" response is no, or vice versa. Furthermore, this belief seems legitimate given the cross-sectional variation in costs invoked by the mean-WTP approach. Note also that CGM's results on cost uncertainty apply to the case where the mean of the uncertain cost is equal to the stated cost, an assumption that my example does not invoke.

The result that Q1 is not incentive compatible relies on the subject knowing that the stated cost is not necessarily the cost the subject will actually face. Thus, as a solution we might ask whether subjects have to know that costs have been artificially randomized and that the stated cost is not necessarily the cost they will face. There are two reasons why the answer is yes.

First, in an open society it is important that citizens know what mechanism is being used to make public goods decisions. At one time, it was suggested that estimated WTP be divided by two for calculating "true" WTP. One prominent critic pointed out, "Are we going to tell subjects this before or after they answer the WTP question?" It does not seem desirable, and may not even be possible, to keep the survey mechanism secret.

[3] If a "yes" vote increases P(WTP > stated cost), then it also increases the estimate of mean-WTP. This condition is needed for CGM's first Result, so it seems reasonable to invoke it here.

The second reason is that making dichotomous choice questions "better" (that is, more like real world choices) will almost surely require allowing subjects to talk among themselves and discuss their responses. Differences in policy costs will then become apparent and need to be explained.

The second solution to the mean-WTP problem is to randomly assign *true* costs. In this way, the stated cost in the dichotomous choice question will be the true cost that will be faced by the subject. But for this to work, true costs must be assigned *independently of preferences*. This requirement rules out making use of any cross-sectional variation in costs that is due to differences in income. Indeed, it effectively rules out any of the mechanisms by which true costs might be expected to vary naturally in the cross-section such as (besides income) family size, place of residence (*e.g.*, which state or county the individual lives in), or consumption of particular goods such as gasoline or recreational equipment.

Thus, estimating mean-WTP requires "truly randomized" costs. Such randomization is probably politically unacceptable. It certainly seems like a high price to pay for estimating mean-WTP. Cost randomization is even less palatable given that the dichotomous question remains incentive compatible under the median-WTP rule.

The incentive compatibility of the median-WTP rule does appear rather robust. Suppose that EPA must estimate values before it knows the true cost of the policy but that it will know the true cost before it makes the final decision.[4] In this case, what is required is that if the true cost turns out to be $20, the EPA must base its policy recommendation based *solely* on responses to the $20 question. Such a decision rule is possible only if EPA uses the median ("Are half of the responses yes at $20?") or other percentile rule. The EPA essentially throws out all of the responses at costs other than $20.[5] Such a mechanism is incentive compatible.

As a respondent, I then know my response will only be used when the stated cost is the true cost and I will give a "true" yes or no.

In summary, the EPA *can* use a mean-WTP rule, but to make the choice question incentive compatible each person must be charged the stated price (just as under the median-WTP rule), which then must differ across the population. This latter condition is severe, since random assignment of costs will likely seem unfair to most citizens, even if economists think it would make a fine social choice mechanism. The median-WTP rule avoids these problems but runs afoul of other issues raised (and not raised) by CGM, which I take up next.

B. Dichotomous choice is *not* the choice we face as a society.

The bigger problem with Q1 is perhaps more obvious: The dichotomous choice question is *not* the choice we face as a society. It is only one among a vast array of choices. Payne *et al.* conduct a valuation survey for *five* programs – salmon preservation, oil spill prevention, Grand Canyon

[4] As CGM correctly point out, if the EPA itself will not know the true cost before making its decision, then the respondent should face a choice-under-uncertainty. A distribution of potential costs should then be included in the survey question.

[5] This is not a severe restriction. Under this mechanism, the EPA could ask each respondent several dichotomous choice questions at the different prices with the instruction that once the true cost is determined, all of the subject's responses at costs other than the true cost will be discarded. Incentive compatibility is preserved.

visibility, migratory waterfowl protection, and wolf reintroduction. All of these programs represent choices that environmental decision-makers face on behalf of society. Multiple-program survey questions may or may not yield "true preferences" but they do portray "true choices."

As CGM note, the incentive structure of survey questions breaks down when we consider N > 2 where N is the number of options to be considered, a result economics has been cognizant of, in varying forms, since Arrow. Since we as a society do indeed face "N > 2," there is no avoiding Arrow's or Gibbard-Satterthwaite's diagnosis. The conceptual bind is inescapable but there are two possible practical remedies.

First, it is possible that subjects' responses will not be particularly sensitive to the number of policy options or the order in which they are presented. If this were the case, the multiple-program problem would seem not to have any practical complications.

The evidence about multiple-program dichotomous choice valuation questions is scanty, so it may be premature to draw conclusions about their performance. Most of the evidence, including Payne *et al.*, is based on open-ended questions, but since we do not expect open-ended questions to work very well for single program situations, it is unrealistic to expect them to work well for multiple-program situations.

Still, it is not hard to imagine that Payne *et al.*'s results will also be observed under closed-ended questions. Their main result is that WTP is higher for the first program in a series of programs, so the remedy of simply ignoring the ordering effect likely would fail. Subjects' responses do appear to be sensitive to the number of policy options and the order in which they are presented

Such a result has straightforward and, to my mind devastating, implications for environmental policy. It means that the decision about what to value – that is, what problem to conduct a benefit analysis for – may have greater consequences than the actual valuation evidence.

It is important to note that Payne *et al.*'s result, or any of the similar results reported in the literature, is *not* a sequencing effect as laid out by Carson, Flores and Hanemann (CFH). The difference between CFH's sequencing results and most multiple-program results does not seem always to be recognized. CFH's model of diminishing WTP applies when a subject pays for and receives the environmental good. In Payne *et al.*, the subjects have simply been *asked* about paying for a good. Their valuation survey gives no suggestion that the program will indeed be carried out, the payment exacted, or the good provided. Therefore, there should be no diminution of WTP throughout the sequence of programs. If WTP diminishes by any substantial amount (as indeed Payne *et al.* found), it is not because of a neoclassical sequencing effect.

A second remedy is to separate the prioritization and valuation problems. The EPA could first develop an explicit and systematic method for setting priorities, that is, for setting up the sequence in which policies will be analyzed and considered. Then, a one-shot dichotomous choice framework could be used to assess the top-ranked policy, then the second-ranked policy, and so forth.

One method for setting these priorities is for the EPA to set up a panel of ecologists, economists, and other concerned scientists who would consider the full range of possible

environmental policy decisions. A variety of policies and their costs would be laid out. The scientific panel would answer the question:

Q2. "If society had $100 million to spend on an environmental problem, which of these policies should it spend it on?"[6]

Note that this panel's members must face a budget constraint in setting their priorities, as in Q2, otherwise they too would be failing to help us make the choices we actually have to make.

Survey respondents would then be asked to answer one-shot dichotomous choice questions for the individual programs, starting with the one that is top-ranked. It would make sense to follow Payne *et al*. by letting the respondents know that they face a series of programs and choices.

My recommendation for this procedure is based on the belief that prioritization is best done by scientists and also on the belief that survey respondents believe that attention to an environmental problem by the government already reflects a serious scientific consensus about the importance of particular environmental issues (see Horowitz).

C. Dichotomous choice differs from voting in substantial ways.

The similarity between opinion polls taken before a referendum and the referendum's outcomes is often cited as evidence in favor of the accuracy of SP methods in eliciting true preferences, as CGM do. The very framework of Q1 lends itself to this argument; who in a democracy could object to posing a Q1-type question?

But there is a substantial difference between Q1 and democratic voting: Voting takes place at a specific, anticipated time. This set-up has two important effects; it allows arguments to be aired and it allows subjects to get into a decision-making frame of mind. (Of course, it is much easier to think about the psychology of voting when facing the 2000 presidential election.)

The ability for arguments to be aired has rather obvious effects, so I leave it for readers to contemplate how its presence or absence might affect valuation outcomes. Tom Schelling has, for a long time, suggested that valuation experiments be conducted with subjects who are allowed to discuss their responses among themselves. Allowing interest groups to form, hire experts or advocates, and make interest-driven recommendations to their adherents would be an even more realistic step.

The ability for respondents to get into a decision-making frame of mind is a neglected element of this problem.[7] It is illustrated by the following exchange:

Student (musing about a valuation question): Some days I feel rich and some days I feel poor, so my answer [to Q1] would vary depending on what day you asked me.

[6] A $50 million and $5 million question might also be asked.
[7] V. Kerry Smith noted that mail surveys allow subjects this kind of opportunity.

Professor (slyly): Well, if we took a good random sample, we would get some people who were feeling flush and others who were feeling penurious. Then we would have the right mix of your two sentiments, wouldn't we?

Student: I only want to make these kinds of decisions when I'm feeling flush.

Allowing respondents to know ahead of time about the decision they will have to make, and to know that their friends are also being asked to make this decision, would be a change that would make valuation more like real choices *and* real democracy. While we might argue that life frequently forces us to make unusual or unexpected choices, it rarely forces us to make substantial policy decisions in an afternoon, by ourselves.

II. Preferences *vs*. Choice: Relationship to Kanninen and Payne, Schkade, Desvousges, and Aultman Papers

The papers in this session and the accompanying discussion painted a stark difference between *preferences* and *choice*. The difference may be a deep conceptual one, but it manifested itself in our session in terms of different practical recommendations for conducting stated preference surveys.

Payne, Bettman, and Schkade and, in a different way Payne *et al*., are interested in *preferences*; in "values," as the term as traditionally been used. They are interested in the attitudes, emotions, and deeply held beliefs of individuals about the environment and the economy. The multiple-program questionnaire of Payne *et al*. is thus seen by them as a fruitful way of getting people to formulate and express those values; a way that is potentially more fruitful than a single, isolated question about a single, isolated program.

Although it is not initially clear, so too is Kanninen. She adopts a "design approach" in which program characteristics and costs are survey variables that can be manipulated so as to provide an optimal survey design.[8]

My discussion instead focuses on *choice* by asking subjects questions that are closest to the questions that we actually face as a democratic society. In this context, program characteristics and costs are not survey design variables but policy design variables.[9] What is important in the framework I have adopted is for subjects to be asked to make serious and realistic choices. Those choices may tell us less about what subjects would have done in other choice situations, but they tell us more precisely about what subjects want to do in the choice situation we actually face. The two frameworks thus have sharply different implications for the design of SP surveys.

III. Conclusion

In summary, let me reiterate what I see as the most important implication of CGM: It is wrong to assume that benefit-cost analysis can set our priorities for us. It is not possible to use valuation tools to solve the prioritization problem. The reason is that when N > 2, valuation

[8] CGM treat preferences and choice as inseparable, as does most of economics.

[9] For example, under the "choice" set-up there will be little cross-sectional variation in the survey questions. Under Kanninen's framework, there will be a great deal of cross-sectional variation in the questions.

methods break down. Setting priorities is important because the order in which we ask survey questions greatly affects the answers we get, as Payne *et al*. have shown.

The one-shot dichotomous choice question is likely to remain our main SP tool for estimating benefits, especially non-use benefits, for benefit-cost analysis. I have recommended making the questions more like the real choices we face, whenever possible, rather than devising more elaborate preference-eliciting formats. I have also recommended setting up an explicit framework for setting environmental priorities. Such a priority-setting framework could mesh well with the one-shot dichotomous choice survey format.

References

Carson, Richard T., Theodore Groves, and Mark J. Machina, "Incentive and Informational Properties of Preference Questions," this volume.

Carson, Richard T., Nicholas E. Flores, and W. Michael Hanemann, "Sequencing and Valuing Public Goods," *Journal of Environmental Economics and Management* **36** (1998) 314-23.

Hagen, Daniel A., James W. Vincent, and Patrick G. Welle, "Benefits from Preserving Old-Growth Forests and the Spotted Owl," *Contemporary Policy Issues* 10 (1992) 13-26.

Horowitz, John K., "A New Model of Contingent Valuation," *American Journal of Agricultural Economics* 75 (1993) 1268-72.

Kanninen, Barbara, "Optimal Design of Choice Experiments for Nonmarket Valuation," this volume.

Payne, John W., James R. Bettman, and David A. Schkade, "Measuring Constructed Preferences: Toward a Building Code," *Journal of Risk and Uncertainty* 19 (1999) 343-70.

Payne, John W., David A. Schkade, William H. Desvousges, and C. Aultman, "Valuation of Multiple Environmental Programs," *Journal of Risk and Uncertainty* 21 (2000) 95-115.

Question and Answer Period for Session I

Richard Carson noted that John Payne raised two substantive issues. The first concerned sequence. Economic theory predicts sequence effects. Payne suggested perhaps the sequence effects are due to the unfamiliar nature of the choices, but other work has found sequence effects when the choices involve common market goods. The effect is certainly not limited to environmental choices.

At a deeper level, given that you will have sequence effects, you could actually write the entire agenda control problem in terms of willingness to pay (WTP) and willingness to accept (WTA) to control the sequence. Nothing in any contingent valuation (CV) or stated preference (SP) method can solve that problem. Further, it is not in politicians' interests to hand over to policy analysts the power to set the sequence of public debate.

Next is the issue of whether a response in a survey truly reflects preference. The literature shows that information offered in a survey can distort preference. Giving more information is not necessarily better: giving only part of the truth can distort people's responses. What you need is balance.

In California, the public gets little information about most of the referenda on the ballot. A survey that lays out the issues in detail to the respondents takes more time than people are likely to devote to the issues in ordinary life.

Carson's final point concerned the notion of multinomial choice questions. He observed that the marginal rates of substitution between attributes are often well identified. Multinomial choice questions can be useful in understanding how people trade off attributes, which can be more important to decisionmakers than total WTP.

James Hammitt, Harvard University, addressed the problem of sequence effects. He noted that a person's WTP depended on many factors, including possible substitutes, complements, and opportunity costs for money. Offering a new alternative may change those factors and so change WTP.

When researchers plan to offer respondents a series of options, what should they tell them about upcoming choices? Hammitt liked Payne's idea of telling people how many choices they were going to be given, but should respondents also get more specific information about their upcoming choices before they make their first choice?

When we ask respondents about a second choice, we often do not make it clear whether they are bound to stick by their decision on the first choice. Is that important?

John Payne replied that telling respondents the number of goods that they will be asked about is not effective in eliminating sequence effects. Neither is giving people information about all the choices before asking for valuations. If you really want people to think hard about substitutes and budgets, you need to precede the target question with an explicit valuation task.

Payne's co-author, David Schkade of the McCombs School of Business, University of Texas, concurred that you must focus people's attention to be sure they consider substitutes. There may be other ways to do it, but giving a valuation task seems to work.

Carson spoke about the case of people's WTP for two goods, *A* and *B*, offered as a package being different than their WTP for *A* plus their WTP for *B*. It is almost impossible to get people to think in terms of "what would you be willing to pay for *B* given that you already have *A*." The more goods you put in the sequence, the harder it becomes for people to think about the scenario.

Hammitt asked, can you convince people of the opposite, that they will not get the earlier goods?

Carson said that also is problematic, once you put multiple goods into play. Research has shown this using simple consumer goods, so we should expect no less when we ask about complex public goods.

V. Kerry Smith, North Carolina State University, raised three points. Regarding Kannenin's paper, he noted that an experiment that presented limited choices may yield clear results, but that the data might not be useful for other purposes, such as understanding marginal effects. When researchers design a survey, do they have any obligation to collect broadly useful data?

Second, Carson's paper emphasized that theory demands that people must believe their answers to be consequential if we want them to reveal their true preferences. What does it take to get people to take surveys seriously?

Third, regarding John Horowitz's comments, Smith wondered if anyone had studied whether giving people advance notice about a survey and its contents gives different results than simply asking people the survey questions.

Barbara Kanninen responded, agreeing with Smith's first point. The work she presented assumed a linear utility function. Where there are nonlinearities or uncertainty about linearity, her study's conclusions about survey design may not apply. You may need a design that will allow you to estimate the nonlinearities.

The survey her paper used as an illustration involved a small sample looking at six attributes. In that case, linearity was a reasonable and necessary assumption if you wanted to draw any useful conclusions from the data.

Julie Hewitt said she did not mean her presentation to suggest that Kanninen's results speak to all situations — they are only for situations where the linear model applies.

Kanninen agreed that you should add midpoints to your design if you suspected nonlinearity.

Carson, addressing Smith's first point, said a growing number of studies explored how big a sample you need if there are nonlinearities. You need a fairly big sample size to detect even moderate departures from linearity.

Regarding how to convince people their survey answers are consequential, Carson said you must construct good questions with realistic, credible choices. Also, you can tell people how the results of the survey will be used. He noted that researchers have room for improvement here. It is

easy to find studies where some people seemed to respond as if their answers were non-consequential.

Mike Christie, University of Wales, Aberystwyth, addressed a question to Kanninen. Most choice experiments use more than two levels. Can you actually derive information about in-betweens if you just offer the extremes as choices?

Kanninen replied that if utility is linear, offering in-between levels reduces attribute differences and actually yields less information from the respondents. Of course, if you do not have a linear model, this is not true. However, the linear model is a reasonably good fit for many situations.

Carol Mansfield, Research Triangle Institute, noted that if she were asked to value a private good like a cashmere sweater, she might be willing to pay $150. If she were next asked if she were willing to pay $15 for a pair of socks, she might be in a frame of mind to accept that high price. However, if she were offered the same socks alone on a separate occasion, she might only be willing to pay $2. She suggested that to get her actual value for socks, it would be better to ask her about socks alone, without other questions that might bias her response.

Payne noted that some choices involve both private goods and public goods. Studies have suggested that the magnitude of sequence effects may vary with the respondent's familiarity with the goods. The more familiar people are with the choices, the less important ordering seems to be.

Schkade said if you know that sequence effects will matter, you have to try different sequences in your surveys. He noted that marketers of private goods love sequence effects and try to take advantage of them to get the highest prices. Surveys looking for an accurate measure of public values have to try different sequences to at least get boundaries for the values.

John Horowitz offered a different perspective. He thought the best cost estimate is the amount you think people might pay and the best sequence to use is the sequence in which the choices might arise in real life.

Payne noted that there is a difference between assessing values for policymaking purposes and assessing values for marketing purposes or for predicting behavior. If you are concerned about predicting behavior, you should do "context matching" — matching the order of questions as closely as you can to the expected real context.

Smith commented further on assumptions about linearity, noting that the translation between goods and income may not be transparent, especially when valuing public goods.

Stephen Swallow, University of Rhode Island, asked about what costs to present in a survey. He distinguished between the benefit side and the supply side. One is the WTP of an individual to get the benefit, and the other is willingness to supply, which is the willingness to pay the government to get the government to supply the benefit.

Horowitz suggested that if what you want to know is, will at least fifty percent of the public be happier if all have to bear a particular cost, then that problem doesn't arise. But if we want to

estimate total WTP, then incentive compatibility kicks in and doesn't let us do it in a way that we envision.

Carson remarked that all consumer preference models are fundamentally unidentified. In real markets, the price variability is small. Stated preference models allow using cost numbers outside that narrow range. But numbers well outside the range may not be plausible. That means there are real problems ever identifying mean WTP measure. If you cannot offer plausible extreme alternatives, you have to settle for a truncated WTP. You cannot necessarily get mean WTP estimates from any kind of data.

Schkade drew a distinction between evaluating a particular program and evaluating a particular change in the state of the environment. Often we want to evaluate the latter. However, we often fall back on offering the former as a choice in a survey, since it is a much more specific, concrete question.

Hewitt noted that respondents' behavior in a survey may fit a household production model developed to explain behavior of subsistence farmers. Subsistence farmers provide and consume the same good. Their motivations are a mixture of desire to enjoy the good and desire to maximize profit. Similarly, responses in an environmental survey reflect both interests in enjoying the good and in contributing to its provision, leading to a more complicated model of response behavior than one has when respondents are simply consumers.

Addressing Smith comments, Hewitt stated that Kanninen has provided an analytic framework for survey design to improve upon the ad hoc nature of design to date. However, Kanninen's work does not completely turn survey design from art to science. It just lets us push back the ad hoc assumptions one level. If you are not comfortable assuming that the utility function is linear, Kanninen's results do not apply.

Kanninen emphasized that when she refers to extremes, she means what the researcher thinks are the limits of the domain. You cannot then extrapolate beyond those bounds.

Joseph Cooper, Economic Research Service, USDA, addressed a comment to John Payne. Three years ago Cooper did a survey with three questions on water contamination. He found sequencing problems. He asked about WTP to reduce nitrate contamination by fifty percent, to reduce nitrate contamination by one hundred percent, and to remove all contaminants from the water. Before asking those questions, the survey asked questions about substitutes and budget constraints, to get respondents to think about those kinds of things. It was clear from the responses that people were not considering budget constraints when they answered their first CVM question, but they were by the time they answered their second question.

He concluded that it is a good idea to have a "throw-away" first question to get people in the proper frame of mind.

In the case of nested options, such as the three in his survey, he believed it was best to ask people to value the comprehensive option first.

Patrick Welle, Bemidji State University, asked two questions. First, he asked Carson if he thought it was wise to follow binary choice questions with an open-ended question aimed at understanding the reason behind the binary choice.

Second, he asked Kanninen for practical guidance on how to use pretesting and focus group information in survey design.

Carson replied that open-ended "why" questions should not corrupt responses. One study, which allowed respondents to revise their answer to the binary choice after they tried to explain it, found some reconsideration.

Payne observed that in attribute value pretesting, they routinely ask questions aimed at identifying unacceptably low values.

Carson noted that in the marketing context, it may be tough to find clean, orthogonal choices. In environmental contexts, you may find choices that benefit one desirable indicator and harm another.

Kanninen replied to Carson that her results suggest you can alleviate the problem he described through use of a balancing attribute.

Replying to Welle, Kanninen said that you should update design as you go. Rather than do one small pretest followed by a large survey, you should divide the large survey into waves and adjust your survey design for each wave based upon the information you have gleaned.

John Hoehn, Michigan State University, noted that even a small sample can help refine design.

Walter Milon, University of Central Florida, asked Payne and Schkade about their work involving building codes. He wondered if there is information in existing studies to evaluate the costs and benefits of alternative building codes.

Schkade noted that the first building code, written by Hammurabi, punished the architects of fallen buildings with death. It worked, after a fashion. But the history of building codes is one of experiment and improvement. Analytically determining the optimal building code would be too much to hope for. Studies can help identify and improve key parameters in codes, but there is no tool yet that can identify the optimal code.

Payne noted that there are lots of examples of legal rules, such as the rules of evidence, that have been refined through the years by experiment and revision.

Over the last twenty to thirty years, investigators have gained insights on how people answer survey questions and have derived strategies to improve the way we ask questions. We are improving the quality of information we can get from preference studies.

Carson noted that the NOAA panel had a specific mandate, which concerned how the government could prove the cost of damage to natural resources. EPA and other environmental agencies face other problems that preference studies can help solve. The question is how to use

limited survey budgets most efficiently. What do we need to know to affect decisions? Given that, how can we extract as much useful information as possible through affordable surveys?

STATED PREFERENCE: WHAT DO WE KNOW? WHERE DO WE GO?

PROCEEDINGS
SESSION TWO

VALIDITY OF STATED PREFERENCE METHODS

A WORKSHOP SPONSORED BY THE US ENVIRONMENTAL PROTECTION AGENCY'S NATIONAL CENTER FOR ENVIRONMENTAL ECONOMICS AND NATIONAL CENTER FOR ENVIRONMENTAL RESEARCH

October 12-13, 2000
Doubletree Hotel, Park Terrace
Washington, DC

Edited by Sylvan Environmental Consultants for the
Environmental Law Institute
1616 P Street NW, Washington, D.C. 20036

ACKNOWLEDGEMENTS

Sections of this report, indicated as "summarizations," were prepared by Sylvan Environmental Consultants for the Environmental Law Institute with funding from the National Center for Environmental Economics. ELI wishes to thank Matthew Clark of EPA's Office of Research and Development and Kelly Brown, Julie Hewitt, Nicole Owens and project officer Alan Carlin of National Center for Environmental Economics.

DISCLAIMER

These proceedings are being distributed in the interest of increasing public understanding and knowledge of the issues discussed at the Workshop and have been prepared independently of the Workshop. Although the proceedings have been funded in part by the United States Environmental Protection Agency under Cooperative Agreement CR-826755-01 to the Environmental Law Institute, it may not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Session II Proceedings
Table of Contents

THE VALUE OF VISIBILITY: A COMPARISON OF STATED PREFERENCE METHODS

Thomas H. Stevens
University of Massachusetts

John M. Halstead
University of New Hampshire

Wendy Harper
Wellesley College

Ina Porras
University of New Hampshire

L. Bruce Hill
Clean Air Task Force

Theresa L. Walker
University of New Hampshire

Cleve Willis
University of Massachusetts[1]

Prepared for presentation at the US EPA National Center for Environmental Economics and National Center for Environmental Research Conference, "Stated Preference: What do we know? Where do we go? Washington, D.C. October 12-13, 2000. Funding from EPA/NSF is gratefully acknowledged. All opinions are those of the authors.

The 1977 Clean Air Act requires the U.S. EPA, the states, and federal land managers to protect and restore visibility in wilderness areas (Harper, 2000). However, despite national

[1] Professor, Department of Resource Economics, University of Massachusetts; Professor, Department of Resource Economics and Development, University of New Hampshire; Visiting Assistant Professor, Wellesley College; Former Research Assistant, University of Massachusetts; Senior Scientist, Clean Air Task Force; Research Assistant, 1 Department of Resource Economics and Development, University of New Hampshire; Professor and Dean, College of Food and Natural Resources, University of Massachusetts.

reductions in sulfur dioxide emissions, visibility in most of the northeastern wilderness has declined substantially since the 1970's. As noted by Hill, et al., (2000), human induced smog conditions have become increasingly worse and average visibility in Class 1 airsheds, such as the Great Gulf Wilderness in New Hampshire's White Mountains is now about one-third of natural conditions. Deregulation of electricity production is likely to result in further degradation as consumers switch to low cost fossil fueled generation, and although EPA regional haze rules attempt to address this problem, many policy makers question whether the value of improved visibility is worth the cost.[2]

This paper focuses on several of the problems associated with the valuation of atmospheric visibility in wilderness areas. One problem is that different forms of the stated preference valuation approach, such as contingent valuation and conjoint or choice analysis, may produce very different results (Stevens, et al., 2000). Results may also differ depending on whether valuation is conducted onsite or offsite, or by mail or in person. Also, little is known about the geographical extent of the "market" for visibility; is it a local, regional, or global public good? Finally, do visibility value estimates adequately exclude the value of joint products like health and ecosystem effects associated with atmospheric pollution?

We begin with a brief review of previous studies. A case study of visibility in the Great Gulf Wilderness of New Hampshire is then presented and discussed.

Background and Previous Studies

Most previous studies of the value of visibility have used the contingent valuation method (CVM). One of the first studies was conducted by Rowe, et al. (1980) who found that non-residents were willing to pay about $4 per day to preserve visual range in southwestern Colorado. Schulze et. al. (1983) reported that residents of Los Angeles, Denver, Albuquerque and Chicago were willing to pay $3.75 to $5.14 per month to preserve visibility in the Grand Canyon. Crocker and Shogren (1991) estimated that residents were willing to pay about $3.00 per day to preserve visibility in the Cascades of Washington State. And, Chestnut and Rowe (1990) found that respondents were willing to pay $4.35 per month to avoid a change in average levels of visibility in the Grand Canyon, Yosemite and Shenandoah National Parks.[3]

With respect to wilderness areas in the northeast, the Appalachian Mountain Club (AMC) administered a survey in the summer of 1996 to ascertain visitor's perceptions of visibility in the White Mountain National Forest. This survey was administered to individuals at three sites: The Pinkham Notch visitors' center at the base of Mt. Washington, the Cardigan lodge at the base of Mt. Cardigan and the Mt Washington Observatory (at the top of Mt. Washington). This survey asked respondents to rate photographs of Mt. Jefferson, a mountain in the Class 1 Presidential Dry River airshed, at various visibility conditions. Each photograph was correlated with a measurement of optical extinction measured by a nephelometer at the site where the photograph was taken. Results of this survey show that individuals were able to consistently perceive different levels of visibility. That is, respondents were clearly able to differentiate between improvements and degradations to visibility (Hill, 2000).

[2] The EPA regional haze rules were recently overturned in Federal Court. However, the EPA has appealed and the current administration plans to take the case to the Supreme Court (Harper, 2000).

[3] Many of these studies were modeled after research and ideas developed or presented at a 1982 conference on visual values (Rowe and Chestnut, 1983).

Although much has been learned, results of previous research suggest that several important questions remain unanswered. The first issue refers to the valuation technique used. As noted by Brookshire, et al.(1982), results should be tested by using valuation techniques other than the traditional CVM. Second, relatively little is known about the relationship between the onsite and offsite value of visibility or about the effects of location (distance) on the value of visibility. The results of location analysis might help to resolve two of the major problems in the valuation of environmental assets: the extent of the market area associated with damage assessments and whether benefit estimates derived from one region can be transferred to other areas. And, very few studies have included analysis of the potential problem of joint products which may be a very important factor in estimating the value of visibility itself.

Theoretical Foundations

As noted, most direct valuation studies of nonmarket goods have used the CVM. The theoretical foundations of generating compensating and equivalent variations have been extensively documented (see, e.g. Mitchell and Carson, 1989; Cummings, Brookshire, and Schulze, 1986; Freeman, 1993) and will not be repeated here. There are several venues for eliciting respondents' stated value for the good(s) in question in the CVM, including open ended valuation questions, iterative bidding, and referendum-style. However, given the primary goal of this study--that is, to directly compare conjoint analysis and contingent valuation--it was necessary to use the dichotomous choice format. Adapting from Freeman (1993), if the individual's indirect utility function can be represented as u(M, v, C) where M is income, v is the level of visibility, and C is a vector of individual characteristics, the individual responds yes if

$$(1)\ u(M - E, v^1, C) - U(M, v^0, c) \geq 0$$

If $w(\beta)$ is the observable component of utility, then

$$(2)\ Pr(yes) = Pr[w(M - E, v^1, C) + \varepsilon_1 > w(M, v^0, C) + \varepsilon_0]$$

Where the ε_i $(i = 0, 1)$ are the random, unobservable components of utility. Willingness to pay for the visibility improvement to v^1 is thus defined implicitly by

$$(3)\ u(M - WTP, v^1) = u(M, v^0)$$

and

$$(4)\ w(M - WTP, v^1) + \varepsilon_1 = w(M, v^0) + \varepsilon_0$$

Conjoint analysis, which has its foundations in hedonic theory, has principally been employed in marketing studies (e.g. Green and Srinivasan, 1978 and 1990; Gineo, 1990; Manalo, 1990), and more recently in recreation studies (Mackenzie, 1990) and community development research (Gruidl and Pulver, 1991). Conjoint analysis is designed to decompose a set of overall responses to "stimulus" factors (such as a list of product attributes) so that the utility of each attribute can be inferred from the respondents' overall evaluations of the stimuli (Green and Wind, 1975; Green and Srinivasan, 1978, 1990). The technique allows disaggregation of the relative importance of each facet of a multi-attribute product. Conjoint measurement has the attractive feature of decomposing the respondent's original evaluations into separate and compatible utility

scales by which the original bundle choices can be reconstituted; a relative value of the importance of each attribute can then be derived. By deriving marginal utility estimates for each attribute, a measure of willingness to pay for changes in the level of these attributes can be estimated, which may then be compared to estimates generated using other valuation methods.

Conjoint analysis, which is essentially a modification of the referendum CVM, asks survey respondents to rate or rank alternative programs. Following MacKenzie (1990), assume an environmental protection program, Q, with M attributes:

(5) $Q = q(Q_i,.., Q_m)$

Where Q_i is the quantity of the i^{th} attribute. In conjoint analysis, product price or cost, P_Q, is treated as just a another attribute so that:

(6) $Q = q(Q_i,.., Q_m, P_Q)$

If the utility function is separable, then:

(7) $U = U(Q_i, - P_Q + M)$,

Where M is all other goods (income).

Assume that respondents are asked to rate two alternative programs, B and C, which differ only in terms of price and attribute Q_i. The utility difference between them is given by:

(8) $U^b - U^c = [U^b(Q^b_i + M - P^b)] - [U^c(Q^i + M - P^c)]$.

If $U^b > U^c$, then program B is rated above C. Rating can, therefore, be expressed empirically as:

(9) $Rating = a(P_Q) + B(Q_i) + e$,

Where (9) represents the empirical utility function (MacKenzie, 1993). Since the two attributes in (9) can be varied while leaving the rating (utility) constant, $-B/a$ is the marginal willingness-to-pay for attribute Q_i (MacKenzie, 1990; Magat et al. 1988). If (9) is expressed in terms of rating differences, Roe et al. (1996) and Johnson et al. (1995) show that CV and conjoint results can then be directly compared, as subsequently argued here.

However, recent studies have explored some of the simplifying assumptions of these earlier approaches and have sought to reconcile CA with the general assumptions of utility theory (e.g. Adamowicz, Louviere, and Williams, 1994; Boxall et al. 1996). Roe et al. (1996) points out that most CA studies utilize an ad hoc functional form, noting that "...respondents' ratings of a single commodity do not provide the information necessary to estimate the welfare gains or losses of moving from one commodity to another" (p. 148). Therefore, this study will use the model developed by Roe et al. wherein it is assumed that the individual utility associated with environmental quality programs is expressed as

(10) $U^i (P_i, q_i, M, z)$

Where P_i is the cost or price of program i borne by this individual, q_i is a vector of program attributes, m is income, and z is a vector of individual characteristics such as age and education. Assuming that utility is related to individual ratings via a transformation function $\Phi(\bullet)$:

$$(11)\ r^i (P_i, q_i, M, z) = \Phi ([U_i (P_i, q_i, M, z)]$$

where $r^i (\bullet)$ is the conjoint rating. A change from the status quo visibility (level 0) to changed visibility (level i) is given by the ratings difference, Δr:

$$(12)\ \Delta r = r^i (P_i, q_i, M, z) - r^0 (P_0, q_o, M, z)$$

Assuming a linear, constant marginal utility of income:

$$(13)\ \Delta r = r^i (q_i, z) + a(M - P_i) - [r^0 (q_0, z) + a (M - P_0)]$$

where a is a constant. From (13):

$$(14)\ \Delta r = r^i (q_i, z) - r^0 (q_0, z) + a (P_i - P_0)$$

A binary response model can also be derived from the conjoint formulation presented in (14). Suppose, for example, that respondents are asked to rate alternatives visibility levels, including the status quo, on a scale of 0 - 10 (with 0 being totally unacceptable and 10 being definitely acceptable) indicating the alternative which the individual would definitely choose. As Roe et al. argue, this formulation follows the standard random utility model:

$$(15)\ \Pr(\text{program I chosen}) = \Pr (U^i (P_i, q_i, M, z) + \varepsilon_i > U^j (P_j, q_j, M, z) + \varepsilon_j)$$

Where ε_i and ε_j are random errors. This binary format should yield the same results as dichotomous choice contingent valuation; thus, equations (15) and (2) should theoretically yield the same results.

To account for possible embedding effects, subsamples of the survey group will be confronted with different levels of the environmental commodity (visibility) where possible. Standard statistical analysis will then be used to test for embedding. It is hypothesized that if embedding effects are not present, value estimates will differ significantly as further degradations in visibility are presented.

Methods

A case study of visibility in the Great Gulf Wilderness in New Hampshire was undertaken during the winter, spring and summer, 1999, and during summer of 2000. Visibility at the study area, which is about one quarter mile northeast of the Mt. Washington summit, is commonly impaired by regional haze that is largely a product of fossil fuel energy production (Hill et al., 2000).

Four surveys were used to measure the value of visibility in the Great Gulf Wilderness region. The first survey was administered *onsite* by a trained interviewer who used a personal

computer (laptop) to present respondents with computer modeled images derived from the WinHaze Visual Air Quality Program. This program allowed us to hold weather conditions constant (cloud cover) while changing visibility only.

The second survey was identical in all respects except that it was administered *offsite* to individuals residing in the Northampton/Amherst area in Western Massachusetts (about a 3 to 4 hour drive from the study site). The third survey which was conducted by mail involved a random sample of 1,000 New England residents. The fourth survey was conducted by mail of a random sample of residents of New Hampshire, Vermont, and Maine.

A split sampling approach was employed throughout. In each of the intercept surveys one half of the respondents received a contingent valuation question that asked for their willingness to accept reduced visibility in exchange for lower electricity bills. The other respondents were asked to rate, on a scale of 1 to 10, the status quo and a scenario with less visibility and lower monthly electricity bills. The first mail survey was modeled after the intercept surveys except that it was possible to confront respondents with multiple scenarios of visibility degradation in eliciting WTA measures via the electric bill vehicle. The second mail survey switched elicitation procedures and sought to estimate respondents' willingness to pay (again using a split sample of conjoint and contingent valuation methods) to *avoid* degradation of visibility.

This sampling strategy allows us to test for differences in economic value estimates due to respondent's place of residence, survey type (mail or personal), type of valuation question (contingent valuation or conjoint/choice), and whether the respondent was contacted onsite or offsite.

The first section of the surveys asked respondents to rate several pictures according to the amount of haze in each. Each picture was a view taken from Camp Dodge, directly across from the Great Gulf Wilderness that had been altered by WinHaze to simulate different levels of atmospheric pollution, all else held constant (cloud cover, etc). Respondents to the personal survey were asked to rate 15 pictures while mail survey respondents rated 4 pictures.

The CVM or ratings (conjoint/choice) question was then presented. Following an introductory statement about electricity deregulation and air quality in the White Mountains (see appendix), each respondent viewed two pictures in this section: picture A represented the status quo visibility and electric bill while picture B represented reduced visibility and a lower electric bill. The CVM and conjoint (choice) questions were asked as follows:

1. WTA

Conjoint Analysis:

- How would you rate the situation in photograph A on a scale of 0 to 10 with 0 being totally unacceptable and 10 indicating that you would definitely be willing to accept this level of visibility along with no change in your monthly electric bill?

- How would you rate the situation in photograph B on a scale of 0 to 10 with 0 being totally unacceptable and 10 indicating that you would definitely be willing to accept this level of visibility along a $x decrease in your monthly electric bill?

Contingent Valuation: Would you be willing to accept this new level of visibility (indicated by picture GB) in the White Mountain National Forest if your monthly electric bill were reduced by $x?

2. WTP

Conjoint Analysis: How would you rate the situation in photograph B on a scale of 1 to 10, with 1 being totally unacceptable and 10 indicating that you would definitely be willing to pay $x per month more for electricity to avoid this new level of visibility?

Contingent Valuation: Would you be willing to pay $x per month more for electricity to avoid this new level of visibility (indicated by picture B) in the White Mountain National Forest?

Picture A, which represented the base scenario, or status quo, describes the average visibility level at the site during the summer months. Picture B represented one of four visual range reductions. The electric bill reduction was 20 percent of the respondent=s total monthly bill in the personal survey and one of 1/4th, 1/3rd, or 2 of the monthly bill for the first mail survey respondents,[4] while respondents to the second survey were confronted with bids ranging from $10 to $50 per month (these values were chosen based on the initial year surveys).

A series of follow up questions were asked to obtain information about each respondent's socio-economic characteristics, motives involved in answering the valuation question, and plans, if any, to visit the wilderness area in the future.

Double wave mailings with postcard follow ups were used in each mail survey. Response rates were approximately 36 percent for the WTA survey and 39 percent for the WTP survey. These response rates are disappointingly low, and raise the issue of non-response bias.

Although previous efforts to obtain willingness to accept (WTA) estimates for environmental commodities have generally been unsuccessful (Hanley, et al., 1997), there are three reasons why a WTA format was initially employed in this study. First, from a theoretical perspective, property rights to a clean environment are often assumed to belong to the public, and consequently environmental losses should be evaluated using a WTA measure (Harper, 2000). If as suggested by Kahneman et al. (1990), individuals value losses more highly than gains, willingness to

[4] Twenty percent is the average savings expected from deregulation.

pay estimates could severely understate value. Second, given deregulation of electricity generation, acceptance of an increase in air pollution in exchange for cheaper electricity is, in our view, a very realistic scenario. Third, few, if any, comparisons of WTA derived from the CVM and conjoint or choice techniques have been conducted.[5] The current year's mail survey was motivated when the results of the original surveys were somewhat ambiguous, leading to use of a willingness to pay format to see if responses differed substantially.

Results

Characteristics of individuals responding to each survey are summarized in Table 1. Respondents to the mail survey tended to be older and have more income as compared to personal survey respondents. One reason for this difference is that personal interviews were conducted on randomly selected individuals who were contacted onsite or offsite at libraries and cafes in the college towns of Amherst and Northampton, MA. On the other hand, the mail survey was sent to a randomly selected list of households in the entire New England region. It is important at this juncture to note that none of the samples are representative of the population as a whole, and therefore the results should not be extrapolated beyond the sample itself.

WTA

Results from each survey in terms of the percentage of respondents accepting reduced visibility in exchange for lower monthly electricity bills is shown in Table 2. For the conjoint responses, three alternative criteria were used to define acceptance; scenario B ranked equal to or greater than A, scenario B ranked greater than A, and scenario B rated a 10 (definitely accept), but not equal to A. Table 2 also shows average electricity bill compensation. It is important to note that relatively few respondents were willing to make a tradeoff between electricity bills and reduced visibility and that willingness to accept was quite sensitive to the criteria of acceptance assumed in the conjoint format.

That relatively few respondents were willing to accept a tradeoff between visibility and electricity cost is not surprising. In this study average electricity bill reductions ranged from only $7.41 to $ 29.14 per month. However, it is important to stress that the scenarios presented are thought to be very realistic given projected conditions for electricity deregulation in New England (Harper, 2000).

To test for the effects of valuation technique, respondent's location, and type of survey (mail or personal), the two logit model set forth in Table 3 were specified. All data derived from the surveys were pooled and dummy variables were included to test for the effect of respondent's residence, whether the survey was a choice or CVM format, whether it was conducted on or offsite, and whether by mail or in person. The dependent variable in the first model is defined as those rating scenario B>A in the conjoint format and yes in the CVM. The dependent variable in the second model takes on a value of one if respondents rated B $\geq$A or answered yes to the CVM question.[6] These two definitions of "yes" responses were used because previous research suggests that value estimates often vary widely depending on how respondent uncertainty is incorporated in

[5] Since the conjoint method avoids pricing the environmental commodity directly, we hypothesize that conjoint or choice analysis might be more reliable in WTA applications.

[6] There were not enough observations to model B = 10 respondents.

the analyses (Elkstrand and Loomis, 1997; Alberini, et al., 1997; and Wang, 1997). One potential advantage of the choice format employed in this study is that as compared to the traditional CVM, respondents can express uncertainty directly. However, the criterion that should be used to define "yes" responses in the choice format has not been determined. One line of argument suggests that from a purely conceptual perspective, responses rating $B \geq A$ are consistent with "yes" CVM responses. On the other hand, the experimental literature shows that the CVM is often subject to the so called "hypothetical bias" problem and that this bias is reduced or eliminated in choice formats that only count $B > A$ or even $B = 10$ as yes responses (See Champ, et al., 1997; Cummings and Taylor, 1999).

In any case, the specifications presented in Table 3 are not rigorously grounded in economic theory. Rather, we view these specifications as similar to Meta Analyses in that we are primarily attempting to examine the influence of location and survey method (choice, personal, mail, onsite, offsite, etc.) on whether or not respondents would accept a reduction in visibility in exchange for cheaper electricity.

As shown in Table 3, WTA reduced visibility is expected to increase with compensation and visibility. We also expect that the probability of accepting a visibility reduction will be less for those who plan future visits to the site and for those interviewed onsite.[7] It is also important to note that about one-half of all respondents were interviewed personally, forty eight percent received a choice survey, 31 percent were interviewed onsite, about 8 percent lived in New Hampshire, and more than two-thirds had plans for future visits. And, only about 15 to 20 percent were willing to trade reduced visibility for cheaper power, depending on definition of a "yes" response in the choice format.

Results obtained from the models are presented in Table 4. WTA reduced visibility increases, as expected, with compensation and visibility.[8] However, residents of New Hampshire were more likely to accept reduced visibility, all else held constant. One possible explanation for this is that individuals who are most familiar with the resource being valued (live relatively nearby in New Hampshire) are simply less concerned about visibility. However, respondents planning future visits to the wilderness area were less likely to accept reduced visibility. It is important to note that whether the survey was conducted in person or onsite was not a statistically significant factor. However, conjoint (choice) respondents were less likely to accept reduced visibility in model 1, but not in model 2.

That the CVM and conjoint models can produce different results should not be too surprising. Although few comparisons of these techniques have been published, most previous empirical comparisons suggest substantial differences (see Stevens, et al., 2000). There are several reasons for this. First, when compared with the CVM, many conjoint questions provide more information about substitutes. Second, from a psychological viewpoint, respondents may react differently when choosing among options then they do when making dollar valuations (Irwin, et al., 1993; Brown, 1984). And, as noted above, Alberini, et al., (1997), Wang (1997), Elkstrand and Loomis (1997) ,Champ, et al. (1997) and others have shown that value estimates can vary widely depending on how respondent uncertainty is included in the analysis.

[7] Those onsite presumably have more at stake.

[8] This suggests that the models pass the so called scope test.

In the case study considered here, the CVM and conjoint questions presented respondents with the same set of substitutes, but conjoint responses were counted as "yes" in two different ways; if B>A or if B≥A. And, this difference seems to be responsible for whether the conjoint results are or are not different from CVM results. In other words, the way in which respondent uncertainty is handled appears to be responsible for the disparity between the CVM and conjoint results in this study.

Estimates of the median economic value of visibility derived from the logit models are presented in Table 5. All median values were calculated by:

$$(16) \quad \text{Pr Accept} = \frac{1}{1+e} - (a + b\,\text{Ln Compensation})$$

where 'a' and b are estimated parameters (see Table 4). The estimated visibility values suggest that the average respondent is not willing to make a tradeoff between energy cost and visibility. The average respondent's monthly electricity bill was approximately $70, substantially less than the median WTA estimates of $924 and $1006 per month derived from models 1 and 2, respectively. And, the median value estimates are very sensitive to whether model 1 or model 2 is used, whether the respondent lives in New Hampshire, or does not plan to visit the site.

Another issue concerns what it is that respondents were valuing. Responses to the follow up questions indicate that many individuals were not just valuing visibility; rather, air quality as a whole was valued. Many respondents linked their WTA response to health problems, now or in the future. Visibility per se did not seem to be the main concern in many cases, regardless of the respondent's geographical location. For example, consider the following quotes from the follow up questions:

> "This 'haze' would in fact be potentially dangerous pollution in the form of air born particulates accompanied by large amounts of invisible sulfur dioxide and some heavy metals. This pollution would be spread and/or funneled by the prevailing winds over a large area. It is the long term effect of these pollutants that is unacceptable. The technology exists to significantly reduce this emission".

> "It will increase sickness and allergies"... "With the increase of haze in the air, more health problems will result. Since I live in Vermont, this will affect my personal health."

> "To me visibility *per se* is cosmetic; what truly concerns me is the contents of that air and its long term effect on human existence..."

Other respondents were more concerned about the effects of the increase in pollution on the ecosystem and wilderness. Context is important here, and respondents felt affected by their environmental "responsibility". For example:

> "This condition is unhealthy for the living things. I am willing to pay a little more to protect the environment"... "Only a small amount of haze can have an enormous impact on the forest ecosystem."... "Endangered species/wild animals that depend on the wild will be likely to migrate or disappear".

> "Clean air and clean water are priceless. I do not think that money is the issue at stake. The health and well-being of humans as well as most other animals and plants is dependent upon the quality of the environment in which we live. To put a price on environmental quality and destroy the resources on which we depend is absurd."

> "Preserve these treasured landmarks"... "Preserving the pristine conditions of National parklands should be a national priority. One that does justify cost to consumers"... "Too much haze for a non-city vacation spot."

Some respondents were also concerned about the effect that visibility might have on the tourism, recreation activities and property values in the White Mountains:

> "As a landowner in the White Mountains I wouldn't accept any increase in air pollution."... "If visibility is poor the usual number of tourist do not come to Maine, New Hampshire or Vermont, there the ripple effect will be seen in less revenues for the states, hotels/motels, restaurants, etc."

Finally, some respondents were totally against energy providers using coal, and advocated the use of alternative forms of electricity that provide the same benefits (reduced costs) without increasing pollution. Some respondents did not believe the assumption that the reduction in visibility would only occur at the White Mountains of New Hampshire and were concerned about the effects of the increased pollution in their own area.

> "Why should the level of visibility be less than it is in the picture A?. There shouldn't be any pollution. Alternative renewable energy sources are available now, which would eliminate pollution and be cheaper than fossil fuels to produce. The use of solar energy and its applications to solar thermal electric and solar photovoltaic electricity, wind energy and hydropower could easily replace fossil fuels and nuclear energy. This would result over a period of just a few years in the elimination of pollution globally and actually reduce the cost of electricity."

Since many respondents valued air pollution in general as opposed to visibility only, the valuation results presented in this study are likely to be biased upward. However, in future studies it may be possible to employ a conjoint format that separates the effects of visibility from the problems of air pollution in general.

WTP

The summer 2000 mail WTP survey is still has only recently been completed, and so much analysis remains. However, Tables 6 and 7 present the results of analysis of the data set in isolation. As can be seen, a number of variables had statistically significant coefficients of the expected sign: gender, income, electric bill, natural log of bid, and natural log of visibility were all significant at the five percent level or above. Unlike the previous mail survey, there were no "state" effects. The model had reasonably good predictive power, as Table 7 illustrates that it predicted about 78 percent of responses correctly. Using equation (16) to calculate median WTP values yields an estimate of $511 monthly. This figure does appear quite high at first blush; however, it appears similar to the

WTA estimates. Clearly, further research aggregating this data set into the overall data pool is necessary and will hopefully shed more light on the issue.

Summary and Conclusions

The findings that emerge from this study can be summarized as follows. First, most respondents were not willing to accept cheaper electricity in exchange for reduced visibility over the range examined in this study. In fact, the estimated economic value of visibility suggests that compensation for improved visibility via lower priced electricity is simply not feasible; the necessary compensation is likely to be greater than the average respondent's actual electricity bill. If respondents are well informed, we might therefore infer that deregulation will not result in a substantial increase in pollution as a result of greater household demand for the cheapest source of electricity.

Second, the effects of location appear to be more complex than previously imagined. Respondents living nearby seem to value visibility less than those living further away, all else held constant. Perhaps absence does make the 'heart grow fonder'. On the other hand, valuation did not differ among those interviewed on or off site, yet those planning future visits were much less likely to accept reduced visibility.

The "market area" for visibility at popular unique sites, such as the Grand Canyon and Yosemite is known to be very large. Although much less is known about the extent of the market for less unique wilderness areas, like the Great Gulf in New Hampshire, this study suggests that its market area may also be quite extensive. On the other hand, conclusions about the effects of location are clouded by the finding that many respondents did not believe that air pollution would be limited to the study site.

Third, the CVM and conjoint models can produce very different results. In this study the difference seems to be a result of the criterion used to define a "yes" response in the conjoint format. Twenty percent of conjoint respondents were WTA the tradeoffs presented in this study when a yes response was defined as $B \geq A$; 9 percent of conjoint respondents were WTA if the criteria is $B > A$; and only about three percent indicated that they would definitely accept ($B = 10$ and $B \neq A$). We therefore believe that future studies should include tests for sensitivity to the valuation question format and to respondent uncertainty.

As noted above, we believe that conjoint (choice) responses rating $B \geq A$ are conceptually consistent with "yes" responses in the traditional CVM. And, our empirical estimates suggest no difference between conjoint and CVM in this case (see Tables 4 and 5). However, the problem of hypothetical bias suggests that "yes" responses should be defined by $B > A$, and when this was done, the resulting value estimates derived from the conjoint format were much different from those derived from the CVM (see Table 5).

Fourth, we did not find differences associated with whether the valuation question was conducted by mail or in person. Perhaps the NOAA guidelines requiring personal interviews should continue to be reevaluated.

Finally, despite survey pre tests and careful wording of the valuation question, many respondents valued air pollution in general. Consequently, the value of visibility is likely to be overestimated. A conjoint analysis that includes several attributes of pollution, including visibility, might clarify this issue, but the problem of sensitivity of this method to the definition of "yes" responses is likely to remain an issue.

Table 1. Socioeconomic Characteristics of Respondents; Sample Means[a]

Survey	Planned Future Visits (%)	Age (years)	Income (thousands)	Residence (%) MA	CT	NH
CJ, Personal Onsite (WTA)	98 (13)	38.8 (13.3)	52.2 (34.7)	20 (40)	07 (25)	12 (32)
CJ, Personal, Offsite (WTA)	60 (49)	36.7 (12.8)	43.8 (36.2)	100 (-)	-	-
CJ, Mail, Offsite (WTA)	71 (45)	48.9 (15.7)	67.2 (39.2)	84 (37)	05 (21)	06 (23)
CVM, Personal, Onsite (WTA)	95 (21)	38.3 (14.6)	38.9 (34.2)	27 (45)	07 (26)	16 (37)
CVM, Mail, Offsite (WTA)	62 (49)	48.0 (13.9)	52.3 (33.9)	16 (37)	39 (49)	11 (31)
CVM, Personal, Offsite (WTA)	54 (50)	31.7 (10.3)	22.4 (20.9)	100 (-)	-	-
CVM, Mail, Offsite (WTP)	67 (47)	51.2 (15.4)	55.1 (29.6)	33	15	52

[a] Standard deviations in parentheses.

Table 2. Summary of Survey Results[a]

Survey	Sample Size	Average Monthly Electric Bill Reduction ($)	Percent Accepting Visibility Reduction		
			B≥A or yes	B>A or yes	B=10[b]
Conjoint, Personal Onsite	60	$13.75 (7.24)	20 (40)	13 (34)	0 (-)
Conjoint, Personal Offsite	60	$12.23 (10.27)	25 (44)	12 (32)	5 (-)
Conjoint, Mail Offsite	105	$25.73 (15.95)	18 (39)	6 (23)	1 (-)
CVM, Personal Onsite	87	$7.41 (2.51)	17 (38)	17 (38)	-
CVM, Mail Offsite	102	$29.14 (25.86)	23 (42)	23 (42)	-
CVM, Personal Offsite	59	$11.35 (6.38)	20 (41)	20 (41)	-

[a] Standard deviation in parentheses.
[b] And B≠A.

Table 3. Logit Model Specification

Variable	Definition	Mean	Standard Deviation	Expected Sign
Model 1 Dependent	Rating B>A or yes to CVM	.15	.36	
Model 2 Dependent	Rating B≥A or yes to CVM	.20	.40	
Ln Compensation	Ln $ monthly electric bill reduction	2.6	.74	+
Ln Visibility	Ln miles	2.78	.56	+
Age	years	41.8	15.1	+/-
Income	thousands	48.3	36.8	+/-
MA	Dummy; Massachusetts resident = 1	.55	.50	+/-
CT	Dummy; Connecticut resident = 1	.12	.32	+/-
NH	Dummy; New Hampshire resident = 1	.08	.27	+/-
PER	Dummy; Personal Interview = 1	.56	.50	+/-
Onsite	Dummy; Onsite Interview = 1	.31	.46	-
FVisit	Dummy; Plans for future visit = 1	.73	.44	-
CJ	Dummy; Conjoint (Choice Survey) = 1	.48	.50	+/-

Table 4. Logit Model Results

Variable	Model 1 (B>A) Parameter Estimates	Standard Error	Model 2 (B≥A) Parameter Estimates	Standard Error
Intercept	-5.01***	1.33	-3.81***	1.09
Ln Compensation	.47*	.25	.35*	.21
Ln Visibility	.67**	.30	.44*	.24
Age	.008	.01	-.0004	.009
Income	.003	.004	.002	.004
MA	.48	.43	.57	.39
CT	.53	.50	.76	.47
NH	1 .10**	.49	1 .30***	.46
PER	.34	.47	.24	.39
Onsite	.54	.51	.23	.43
FVisit	-1.15***	.32	-.87***	.28
CJ	-.99***	.32	.01	.26
N	472		472	
F	38.79***		26.78***	
Percent correct predictions	72.6		65.5	

*** Significant at .01 percent level; ** Significant at .05 percent level; * Significant at .10 percent level.

Table 5. Visibility Value Estimates: Median WTA Per Month[a]

	Model 1 (B>A)	Model 2 (B≥A)
I. Average Respondent	$924	$1006
II. New Hampshire Resident; no visits planned	$36	$17
III. Average resident; No visits planned	$154	$162
IV. Average Respondent Conjoint model	$2790	C[b]
V. Average Respondent CVM Model	$331	C[b]
VI. Average Respondent CVM Model (WTP)	$511	

[a] Values rounded to nearest dollar.
[b] Conjoint dummy variable not different from zero (see Table 4).

Table 6. Coefficient Estimates, WTP to Avoid Degraded Visibility		
Variable	**Coefficient (Standard Error)**	**|P[|Z|>z]**
Constant	-0.9760 (1.6800)	0.5636
Gender	0.9326 (0.4886)	0.0563
Education	0.1408 (0.3118)	0.6516
Age	-0.0044 (0.0172)	0.7965
Income	0.0001 (0.0000)	0.0197
Electric Bill	0.0104 (0.5243)	0.0469
Ln Bid	-1.1742 (0.3882)	0.0025
Ln Visibility	1.3738 (0.4685)	0.0034
Maine	0.2581 (0.5043)	0.6087
Vermont	0.9035 (0.6055)	0.1357
Future Use	0.3035 (0.5337)	0.5696

n = 139

Log likelihood function: -69.87915

Chi-squared: 41.84588

Table 7. Actual vs. Predicted Values, WTP to Avoid Degraded Visibility				
		Predicted		
		0	**1**	**Total**
Actual	0	81	8	89
	1	23	27	50
	Total	104	35	139

References

Alberini, A., Boyle, K., Welsh, M. 1997. "Using Multiple-Bounded Questions to Incorporate Preference Uncertainty in Non-Market Valuation." In: Englin, J. (Compiler), Tenth Interim Report, W-133 Benefits and Costs Transfer in Natural Resource Planning, University of Reno, Reno, NV.

Brookshire, Thayer, Schulze and D=Arge. 1982. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches," *American Economic Review* Vol 72:165-177.

Brown, T. 1984. "The Concept of Value in Resource Allocation," *Land Economics* 60, 231-246.

Champ, P., Bishop, R., Brown, T., McCollum, D. 1997. "Using Donation Mechanisms to Value Nonuse Benefits From Public Goods". *J. Environ. Econom. Manag.* 32(2), 151-162.

Chestnut, L. and R. Rowe. 1990. "Economic Valuation of Changes in Visibility: A State of the Science Assessment for NAPAP." In Acidic Deposition: State of Science and Technology, Report 27, Washington, D.C. NAPAP.

Crocker, Thomas and Jason Shogren. 1991. "Ex-Ante Valuation of Atmospheric Visibility." *Applied Economics*. Vol. 23, 143-151.

Cummings, R. G. and L. O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review*. Vol. 89(3), 649-65.

Elkstrand, E., Loomis, J. 1997. "Estimated Willingness to Pay For Protecting Critical Habitat For Threatened and Endangered Fish With Respondent Uncertainty." In: Englin, J. (Compiler), Tenth Interim Report, W-133 Benefits and Costs Transfer in Natural Resource Planning, University of Reno, Reno, NV.

Harper, W. 2000. "A Comparison of Direct Methods for Valuing Environmental Amenities: A Case Study of the White Mountain National Forest." (Ph.D. dissertation, University of New Hampshire).

Hill, Bruce, W. Harper, J. Halstead, T. H. Stevens, Ina Porras, and K. D. Kimball. 2000. "Visitor Perceptions and Valuation of Visibility in the Great Gulf Wilderness, New Hampshire." in Cole, et al. Proceedings: Wilderness Science in a Time of Change. Proc., RMRS-P-000. Ogden, VT: U.S.D.A. Forest Service, Rocky Mountain Research Station.

Irwin, J. R., Slovic, P., Licktenstein, S., McClelland, G. 1993. "Preference Reversals and the Measurement of Environmental Values." *J. Risk Uncertain*. 6, 5-18.

Kahneman, D., J. Knetsch and R. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*. Vol. 98, 1325-1347.

Porras, Ina T. 1999. "Off-Site Value of Visibility: New Hampshire's White Mountain Case Study." (M.S. thesis, University of Massachusetts, Amherst).

Roe, B.; K. J.; Teisl, M. F. 1996. "Using Conjoint Analysis to Derive Estimates of Compensating Variation." *Journal of Environmental Economics and Management*. 31, 145-159.

Rowe, Robert, Ralph D'Arge and David Brookshire. 1980. "An Experiment on the Economic Value of Visibility." *Journal of Environmental Economics and Management*. Vol 7, 1-19.

Rowe, Robert and Lauraine G. Chestnut, editors. 1983. Managing Air quality and Scenic Resources at National Parks and Wilderness Areas. Westview Press: Boulder, CO.

Schulze, William D., David Brookshire, et. al. 1983. "The Economic Benefits of Preserving Visibility in the National Parklands of the Southwest." *Natural Resource Journal*. Vol. 23, 149-165.

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

Wang, H. 1997. "Treatment of Don't Know? Responses in Contingent Valuation Survey: A Random Valuation Model." *J. Environ. Econom. Manag.* 32, 219-232.

APPENDIX

Introduction to Valuation Questions

For the next question, consider the following: Currently, many states are debating the issue of deregulation in the electric utility industry. If deregulation occurs in your state, you may be able to choose your own power provider. Assume for the purposes of this question that cheaper power (that is, less than what you currently pay) is available through a mid-western power company. Further, this power company produces electricity by burning coal. Increased demand for this company's cheaper power will contribute to air pollution and poor visibility in the White Mountains.

Now suppose picture A represents the level of visibility most often experienced in this region during the summer months. Further suppose that you were faced with a situation where the visibility level would change to that in picture B. The purposes of this question assume that visibility would change ONLY in the White Mountain National Forest.

COMPENSATING FOR PUBLIC HARMS: WHY PUBLIC GOODS ARE PREFERRED TO MONEY

Presented by Carol Mansfield, Research Triangle Institute
Co-authored with George Van Houtven, Research Triangle Institute and
Joel Huber, Duke University

Summarization

Dr. Mansfield presented work based on two studies of willingness to accept (WTA) and difficulties in measuring WTA.[1]

Studies attempting to measure WTA often carry indications casting doubt on the results' reliability, such as high protest rates or very high stated WTA amounts. However, being able to measure WTA is important from a practical policy perspective.

In earlier work, Prof. Huber and others had found that surveys involving risk/risk trade-offs produced more stable results than surveys involving risk/dollar trade-offs.[2] That led Dr. Mansfield and her fellow researchers to wonder if surveys offering public goods to offset public bads would yield more valid results than surveys offering dollars to offset public bads.

Some studies in the literature suggest that people may have a negative reaction to cash compensation. Two articles by Frey and co-authors looked at the siting of a low-level nuclear waste dump in Switzerland.[3] The authors found that when residents were asked simply, would you accept the disposal site, 50 percent said yes. When residents were asked, would you accept the site if you received a large cash payment in compensation, only a quarter said yes. The authors concluded that offering compensation crowds out altruistic motivations to support public projects.

Other studies have looked at people who qualified for welfare but refused to accept it.[4] They theorized that there is a social stigma associated with accepting cash compensation.

---

[1] Two papers provide more detail on the results discussed in the presentation: Mansfield, C., G. Van Houtven and J. Huber (2000), "Compensating for Public Harms: Why Public Goods are Preferred to Money," RTI working paper, 2000; and Mansfield, C., G. Van Houtven and J. Huber (forthcoming) "The Efficiency of Political Mechanisms for Siting Nuisance Facilities: Are Opponents More Likely to Participate than Opponents?" Journal of Real Estate Finance and Economics.

[2] Viscusi, W.K., W.A. Magat, and J. Huber (1991), "Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-Offs for Chronic Bronchitis," Journal of Environmental Economics and Management, vol. 21, pp. 32-51.

[3] Frey, B.S. and F. Oberholzer-Gee (1997), "The Cost of Price Incentives: An Empirical Analysis of Crowding Out," American Economic Review, vol. 87 (4), pp. 746-755.
Frey, B.S., F. Oberholzer-Gee, and R. Eichenberger (1996), "The Old Lady Visits Your Backyard: A Tale of Morals and Markets," Journal of Political Economy, vol. 104 (6), pp. 1297-1313.

[4] Moffitt, R. (1993), "An Economic Model of the Welfare Stigma," The American Economic Review, December 1993, pp. 1023-1035. Keane, M.P. (1995), "A New Idea for Welfare Reform," Federal Reserve Bank of Minneapolis Quarterly Review, Spring 1995, pp. 2-28.

Possibly too, people find it easier to balance like against like – that is, people find it easier to balance public bad against public good than public bad against private compensation.

Still another study theorized that by accepting compensation, people were also mentally accepting moral responsibility for any future harm from the public bad.[5]

But why would people prefer public goods? Isn't a public good a bribe in the same way as direct cash compensation?

In general, people do not seem to perceive public goods as bribes in the same way as cash. Accepting public goods allows people to feel that they are doing something good for the community. Cognitively, public goods are easier for people to balance against public bads, and public goods may in some way mitigate the harm done by public bads better than cash alone could. For example, a public good could provide a long-term offset to the harm to a neighborhood's reputation caused by a public bad in a way that cash payments never could.

Mansfield and her fellow researchers explored two hypotheses. First, they hypothesized that the existence of a public bad in a neighborhood increases the utility from public goods relative to cash. For example, say you offered respondents the choice between two houses, one with a park nearby and the other with lower taxes. Then you offered a similar choice, except that both houses offered were near a public bad such as a landfill or airport. They hypothesized that with the public bad in the mix, more people would choose the public good (the park) as a counterbalance.

Second, they hypothesized that when you move from a market framework to a WTA or compensation-for-a-public-bad framework, the value of public goods as compensation will increase relative to cash.

In the study, the researchers first asked respondents some neutral market choice questions. They asked respondents to choose between two houses — one enjoying a nearby public good, such as a park, and the other with lower taxes. In some cases researchers also told respondents about a public bad that affected both houses, and in other cases they did not mention the bads.

They also asked some traditional WTA questions. In some they offered cash compensation and in others they offered a public good as compensation.

The researchers tested a variety of public bads and gave respondents a significant amount of detail about the unpleasant aspects of each scenario. In general, the scenarios posed noise, odor, or aesthetic problems, not health or safety risks. Some of the public bads had local effects, involving only the closest houses, and some had broader effects. Similarly, the public goods in the study varied from having local to more general effects.

They collected three data sets. The first, a pre-test, was from a mail-back paper survey offered to church members. The second and third were from self-administered computer surveys, one conducted in a mall in Greensboro, North Carolina, and the other at five different malls in North Carolina, Florida, New Jersey, and New York.

---

[5] Boyce, R.R., T.C. Brown, G.H. McClelland, G.L. Peterson, and W.D. Schulze (1992), "An Experimental Examination of Intrinsic Values as a Cause of the WTA-WTP Disparity," The American Economic Review, vol. 82, pp. 1366-1373.

In terms of demographics, the church respondents were older, wealthier, more likely to be white, and more likely to be retired than the other two samples. The Greensboro mall in the second data set had low income demographics, and for the third data set the researchers deliberately sought out malls with higher income demographics.

The researchers first compared the percentage of respondents who chose the house with the nearby public good over the house with lower taxes in the neutral market choice questions. In all scenarios except one, telling people about the presence of a public bad made them less likely to choose the house with the lower taxes over the house near the public good. Taken individually, not every scenario showed significant differences at the 10 percent level between the public-bad and no-public-bad responses. However, when the researchers pooled the data from all scenarios and ran a random effects logit, the results showed an significant, increased preference for the public good over lower taxes (cash) when the public bad was described in the neutral market choices between two houses. The basic result seemed to be that if you lived in a neighborhood with a public bad, you wanted to live near a public good.

In the same surveys, the researchers also framed questions in terms of compensation, asking what people were willing to accept in return for allowing a public bad to be sited in their neighborhood. In the church-member survey and the multiple-mall survey, more people accepted the public good offered as compensation than accepted cash. In the Greensboro mall survey, more people accepted cash. The sample sizes were small. Pooled together in a fixed-effects logit model, more people accepted the public good.

This alone proves nothing, except perhaps that the researchers picked public goods that people preferred, or that they offered too little cash. The analysis did not speak to what the researchers wanted to know: whether the relative attractiveness of cash dropped when they moved from the neutral market choice to the compensation framework.

To investigate the possibility that the relative value of cash and the public good changed as respondents moved from a neutral market choice to a compensation or WTA format, the researchers constructed log odds ratios calculated from the two WTA responses (WTA cash and WTA the public good). These odds ratios were used to derive predictions of how people would respond to the neutral market choice question offering a public good or cash in the presence of a public bad. In every case but one, the estimated preference for cash derived from the WTA questions underpredicted the number who actually chose cash in the neutral market choice questions. This suggests that the value of cash declines relative to the marginal value of extra public goods in a compensation framework.

Summing up these findings, Dr. Mansfield observed that the presence of a public bad in a neighborhood increases the value of public goods over monetary incentives, even in the market choice scenarios designed to minimize guilt or other intrinsic motivations for preferring the public good to cash. A compensation framework further increases the value of public goods relative to cash.

The researchers held a focus group between the first and second surveys, and some of the comments of the participants shed light on the thinking involved in these kinds of choices. Regarding cash as compensation, one participant said, “It won’t help if everyone just gets money.” Another said, “The government shouldn’t be able to just pay people to do whatever they want.”

Regarding a public good as compensation, one person said, "I think it would be good for the whole neighborhood." Another said that the public goods would "help make the neighborhood nicer, help address the problem."

Regarding the difference between the compensation framework and the neutral market framework, one participant said, "This [market] choice is easier to make, it's not an ethical dilemma."

Addressing the policy implications, Dr. Mansfield suggested that it might be easier to site nuisance facilities if you offer public goods as compensation. In case studies of actual sitings, public goods are often used as compensation, and Dr. Mansfield could find no examples involving only cash as compensation. Compensation is important to the siting process, and economic insights can help make it more likely to be accepted. People who supply the public bad and bear the burden of compensation should be interested to know that they may be able to save money by supplying a mitigating public good rather than cash.

On a related note, Dr. Mansfield made some observations on the characteristics of people who tend to say "no" to WTA questions. She discussed a study by Hamilton on the expansion of hazardous waste facilities.[6] Hamilton found that in counties with higher voter turnout, existing facilities were less likely to expand. He argued that voter turnout was a proxy for community activism.

In Mansfield's study, the researchers noticed that the demographic characteristics of the people who refused any compensation seemed to match the characteristics of people who tend to vote or engage in collective action. A formal analysis of the data tended to confirm this insight. This fits well with the general observation that it is hard to move sitings of nuisance facilities through the political process in politically engaged communities, even if only a minority of residents oppose the facility.

Dr. Mansfield observed that the percentage of people willing to accept compensation in her studies was rather high compared to other studies. She attributed this to the nature of the public bads presented. People are probably more likely to accept compensation for declines in odor, noise, or aesthetics than in health or safety.

The study highlights the importance of context. Values obtained in one setting may not translate to another. In particular, if a policy choice involves compensation, researchers should frame the issue in a compensation context when determining WTA values. Values of public goods should be seen as a lower bound on WTA.

Future research needs to shed more light on why the presence of the public bad matters. Also, research could further explore how the nature of public goods and public bads affects people's choices. Finally, further research could explore this work's implications in the environmental justice field.

---

[6] Hamilton, James T. (1993). "Politics and Social Costs: Estimating the Impact of Collective Action on Hazardous Waste Facilities," Rand Journal of Economics, vol. 24 (1), pp. 101-125.

PROGRESS TOWARD COMPARING STATED PREFERENCE AND REAL MONEY CONTINGENT VALUES

Stephen K. Swallow
Michael A. Spencer
Laurienne Whinstanley

Department of Environmental
and
Natural Resource Economics

University of Rhode Island
Kingston, RI 02881

Discussion paper prepared for the U.S. EPA Workshop on stated preferences research, Washington, DC., October 12-13, 2000.

This paper discusses work in progress. Any results presented are preliminary, and may change on further analysis. Research is supported by NSF/EPA Science to Achieve Results, Partnership for Environmental Research grant #R825307-01-0 and the University of Rhode Island Agricultural Experiment Station. All opinions and errors are the responsibility of the authors, and do not reflect the opinions of any of the funding agencies. We are deeply indebted to Donald A. Anderson, President of StatDesign, Inc., Evergreen, CO, who provided a fractional factorial design for wetland survey questions discussed in this paper.

Environmental goods affect the general welfare as public goods, by which many individuals can benefit from their conservation without necessarily paying for the costs of conservation. As is well accepted by economic principles, private market transactions, or the free market, will generally under-provide such public goods because the potential suppliers often lack any free-market means to earn compensation for their provision. Thus, wetlands and wildlife habitats have frequently been targets for development or conversion to other uses because markets reward landowners directly for investments made in agricultural, residential, or commercial land uses. For several decades now, economists have been working on methods to identify the value of such public goods in a manner that is comparable to the value of private goods sold in markets.

For both market and non-market (public) goods, the concept of economic value depends upon what an individual is willing to give up, or sacrifice, in order to obtain the good in question. Often this willingness to sacrifice is measured in terms of money, a convention on which economists often focus because individuals, the public, corporations, and politicians are believed able to translate, subjectively, money-measures of value into terms of personal relevance, such as an equivalent value of groceries, entertainment, or commercial value. For market goods, market prices represent the value that the marginal buyer or seller places on a good, so that the value of market goods can be estimated based on observations of behavior by individuals in actual market transactions. For some public goods, individual behavior in markets can also be used to estimate at least some portion of the economic value of the good, such as when home buyers are willing to pay a premium for residential properties located in aesthetically pleasing environments, or such as when outdoor recreationists purchase travel services in order to access environmental resources of wildlife refuges or parks.

However, individuals do benefit from public environmental goods in ways that are independent of their decisions in existing markets. In order to estimate the economic value, in monetary terms, of these services from public (environmental) goods, economists have proposed the contingent valuation method (CVM). CVM is thoroughly described in Mitchell and Carson's (1989) book. For the present purpose, it is a valuation methodology by which economists use a survey of the public to ask individual respondents whether or not they would be willing to pay a specified dollar amount to obtain, maintain, or conserve a particular environmental good, such as an undeveloped wetland acre or a particular level of water quality in a local river.

While many environmental economists have accepted the CV methodology, others remain skeptical concerning the validity of money measures of value derived from hypothetical statements, or from the stated preferences of survey respondents. This skepticism has bred controversy as the CVM has been applied to successively larger policy questions or to determine liability for large-scale damages to public trust resources caused by commercial accidents. In response, economists are providing research to identify whether estimates of public good values under CVM match estimates that would be obtained if survey respondents had faced the actual, rather than simply the hypothetical, consequences of their stated preferences, stated choices, or stated willingness to pay. In particular, researchers are concerned with identifying whether CVM-estimated willingness to pay departs from an individual's actual willingness to pay for a public good under conditions that provide her with incentives to be truthful (and realistic) about her true values (incentive-compatible valuation).

This paper provides an overview of some economic experiments or experimental surveys designed to look for and to examine any departures between willingness to pay values based on stated preference (CVM) and values based on surveys in which respondents were required to provide cash or a check if they stated they were actually willing to pay a particular amount for a public good. Two main groups of experiments are discussed, and partial results from some of these are provided.

One group of experiments follows from Spencer et al. (1998) who focused on estimating the value of a monitoring water quality at different ponds in Rhode Island. The new experiments discussed here examine features of the payment mechanism in relation to willingness to pay for adding a pond-site to the statewide water quality monitoring program. These experiments were intended to help identify the most appropriate way to present the payment mechanism in surveys involving real money.

The second group of experiments involves the valuation of wetlands in northwestern Rhode Island. In particular, these experiments focus on estimating the value of purchasing the development rights for wetlands with various attributes, such as the wetland's location relative to developed and undeveloped lands, an indicator of the potential role of the wetland in wildlife or biodiversity conservation, the availability of public access, and acreage. Of course, in Rhode Island, as in many states, wetland conservation laws are fairly strict, often severely restricting development. However, state law does provide a process by which landowners may apply for permits to allow the modification of wetlands, most often as a part of accessing or developing adjacent upland parcels considered suitable for development. Survey respondents were asked about their willingness to pay for a binding "conservation contract" or "easement" on different parcels of wetlands, where this agreement would be placed as a deed restriction on a real wetland parcel, but where the deed restriction would expire after 10 years. This group of experiments is intended to mimic actual land conservation programs wherein public or private conservation agencies purchase the development rights from landowners who continue to hold the title to the land. In these experiments, the survey question was formatted as a choice experiment or contingent choice (Opaluch et al. 1993; Adamowicz et al. 1998), wherein respondents viewed descriptions of two wetland parcels and stated their preference to pay and preserve one parcel or to forgo conservation on both parcels. The real money surveys also requested actual payment.

I. Background Literature and Implications

Previous research illustrates the potential that contingent valuation may produce willingness to pay (WTP) estimates that differ from WTP estimates produced when respondents must actually make a payment. In order to mitigate concerns about incentive-compatibility, a series of experiments applied contingent valuation to estimating the willingness to pay for traditional market (non-public) goods, such as chocolate bars and coffee mugs (e.g., Cummings et al. 1995; Cummings et al. 1997; Fox et al. 1995; Neill et al. 1994). The results generally indicate that hypothetical valuation exceeds valuation based on questions requiring real payments. These results have led to work in designing various methods to adjust or calibrate hypothetical values to real values, using, for example, experimental economics or statistical functions as complements to traditional CVM (Blackburn et al. 1994; Champ et al. 1997; Shogren 1993; Swallow 1994). Other research involves

developing hypothetical surveys that produce value estimates closer to estimates based on real payments (e.g., Cummings and Taylor 1997).[1]

However, there remain questions involving the comparability of results based on market (private or non-public) goods to those based on public goods, because, for example, with market goods an individual's stated value may reflect her prior expectation of market price but she may not actually be in the market for the particular good. Findings based on public (environmental) goods have also found departures between stated-preference values and values based on actual payments for public goods (Brown et al. 1996; Seip and Strand 1992).[2] Yet these results are not entirely conclusive because of the difficulty in eliminating incentive-incompatibility in the respondents' answer to real-money questions about public goods. Indeed, it was the concern about incentive-compatibility that led Brown et al. (1996) to focus on donations in their evaluation of question formats, explicitly conditioning their contribution to contingent valuation research on the assumption that, if the donation vehicle for payment induces a bias in real payments, any real-money bias would be similar to the bias with hypothetical payments.

Unfortunately, it appears that characteristics of the payment vehicle can induce differential effects on valuation of a particular public good or environmental attribute. Johnston et al. (1999) use a hypothetical survey to demonstrate that the relative value of environmental attributes can be conditional on attributes of the payment vehicle, or the means of collecting money, even though respondents may treat a marginal dollar equally under different payment vehicles. Their result is consistent with the substitutability of environmental policy attributes, as elucidated by Hoehn and Randall (1989) and Hoehn (1991), so that there may be a neoclassical explanation for differences between stated-preference (hypothetical) and real-money WTP estimates since, by definition, these payment vehicles (hypothetical and real) have different attributes. Moreover, Johnston et al. (1999) show that their result may not obtain for all environmental attributes, even within the same policy context (watershed management). Therefore, we expect the implied challenge to studies of the validity of CVM to be unpredictable.

Nonetheless, research into the validity of CVM is an important topic, due to the dependence of large-scale public policies on CVM-estimated WTP. Experimental economists, such as Shogren (1993), suggest using laboratory markets to develop improved contingent valuation approaches (see also, Blackburn et al. 1994; Shogren et al. 1994).

For example, experimental economists evaluate alternative mechanisms to reduce incentives for individuals to "free ride" on others' financial contributions to the provision of public goods (Marks and Croson 1998). These mechanisms include establishing a threshold of aggregate payments that are necessary to provide the particular public good at all, establishing a money back guarantee for contributors in the event the public good is not provided, and defining an means to dispose of contributions in excess of the threshold required. Excess contributions may be disposed of through rebates to the original donors or through provision of additional units of a public good, which creates "extended benefits" to contributors or to all individuals. Within these mechanisms, Marks and Croson (1998) hypothesize that the threshold provides an incentive to contribute to the

[1] Greg Poe and colleagues have evaluated a survey method wherein respondents indicate how certain they are that they would pay a specific amount, and using a multiple- bounded method of valuation.

[2] Of course, Bishop and Heberlein's (1979) classic study in the *American Journal of Agricultural Economics* lies in this literature as well.

public good, the money-back guarantee eliminates the risk of incurring costs when others fail to contribute to the threshold, and the disposition of excess contributions reduces the cost to an individual who over contributes. Rondeau et al. (1999) assess these mechanisms in a laboratory experiment to estimate real-WTP when researchers control the value of the public good to each participant, finding that the estimated WTP of the average experimental participant is a satisfactory predictor of the average value of the public good. Spencer et al. (1998) used a money back guarantee and a threshold for aggregate contributions (a provision point) in estimating the value of adding a site to a statewide set of water quality monitoring sites in Rhode Island, finding that stated-preference estimates of WTP exceeded real-money estimates by a ratio of 4.7 to one, although the difference was not statistically significant.

This literature contributed several considerations to the experiments described below. First, following the laboratory results, we designed a set of experiments to evaluate the implications of alternative mechanisms to reduce the tendency of individuals to free-ride on the contributions of others. In this case, we anticipate that a money-back guarantee coupled with a threshold provision point for an environmental good and with a rebate of excess contributions will reduce the divergence between hypothetical-money and real-money contingent valuation estimates.

Second, we attempt to control for the effects of changes in the attributes of the payment mechanism. In most cases, this implies that the stated preference survey and the real-money survey instruments are quite similar, except for the presence or absence of a requirement to pay actual dollars. That is, in most cases the presentation of a stated-preference (hypothetical) survey includes a description of the money-back guarantee, the provision point, and the rebate or other distribution of excess contributions.

Third, the wetlands experiments are designed to allow an evaluation of effects on the estimated WTP for several attributes of the wetland parcels. This design attempts to anticipate that attributes of the payment vehicle, possibly including the mere definition of hypothetical and real payment mechanisms, may induce a different degree of divergence between hypothetical and real-money WTP estimates for different attributes of the wetlands parcel, the public good of interest.

II. Experiments for Mechanisms to Reduce Incentives to Free Ride: Valuing Water Quality Monitoring

A set of experiments was designed around an "adopt-a-pond," water quality monitoring program similar to that reported in Spencer et al. (1998). This application, or environmental good, was defined in collaboration with the University of Rhode Island Watershed Watch program. As described below, the experimental surveys asked student college students about their willingness to pay to have a water quality monitoring schedule established at a pond that was not previously part of the URI Watershed Watch program. This experiment was designed to examine the effects of the method of payment on a respondent's willingness to pay in a discrete choice (yes-no) format, for either hypothetical or real payments to support water quality monitoring at a pond described in the survey.

Payment Formats

Three payment formats were used, as motivated by the experimental economics literature. In all three formats, respondents were informed that the pond would be monitored if the group of

respondents they were part of (50 respondents per group) reached an aggregate level of contributions of at least $350. This aggregate contributions level represents a "provision point" that is constant across each sub-sample of respondents. Each sub-sample, or group, of respondents then faced one format for payments supporting water quality monitoring.

The base format stated to respondents that if their group failed to reach the provision point, their personal contribution would be refunded to them. This format establishes a money-back-guarantee (MBG) but fails to address the issue of how excess contributions above the provision point would be distributed. This base format is called the No Rebate (NR) format.

The second format included the money-back-guarantee along with a statement that excess contributions would be rebated to contributors on a proportionate basis. In this case, if a group of respondents contributed 10% more than the provision point (e.g., contributing a total of $385 rather than just $350), then each contributor would receive a rebate of 10% of their personal contribution. This payment format is called the Uniform Proportional Rebate (UPR) format.

The third format again included the money-back-guarantee, but rather than a rebate, the format included a statement that contributions in excess of the provision point would be retained by URI Watershed Watch and used for the purpose of covering administrative costs. This format is called the Extended Benefits (EB) format.

The Experiment

The experiment was conducted at the University of Rhode Island with 435 students. The students-participants were obtained by prior arrangement with the instructors of one introductory class in sociology and one introductory class in psychology, and by direct solicitation in the university's cafeteria. The experiment took place during regular class hours in the sociology and psychology classes, and none of the students were forced or pressured to participate in the experiment. Classroom students were given prior notice that the experiment would not be a class assignment, but that they would be paid to make real decisions regarding a real environmental resource. The students recruited in the university's cafeteria were recruited on the spot during lunchtime to complete a survey involving environmental decisions. All survey respondents were paid to participate—respondents in the sociology class were paid $10, while respondents recruited from the psychology class and the cafeteria were paid $2.

The experiment required respondents to fill-out a survey, which elicited their attitudes and values for an environmental resource. Standard socio-economic questions were also elicited. In a split-sample style design, respondents were given either a hypothetical-money or a real-money survey treatment. Parallel language was used in both survey formats, except for slight modifications where respondents were asked to donate money to support a program involving the environmental resource. In the hypothetical-money survey treatment respondents faced a decision involving a hypothetical-money cost, while in the real-money survey treatment respondents faced decisions with real-money consequences. Of the 435 respondents 150 participated in the real-money treatment; 50 each under the three payment rules. The rest participated in some version of the hypothetical-money treatment.

The environmental good used in the experiment was an "adopt-a-pond" water-quality monitoring program. Over 90% of Rhode Island's water quality data on ponds and lakes comes

from volunteer monitors coordinated through a non-profit program called URI Watershed Watch. Watershed Watch relies on voluntary donations to operate, and each pond costs $500 per year to monitor [see Spencer, Swallow, and Miller (1998) for an overview of pond monitoring in Rhode Island]. If Watershed Watch collects enough funds to monitor a pond then the pond will be monitored; otherwise, it will not be monitored. Through Watershed Watch we were able to offer, for sale, a water-quality monitoring program for a specific pond that was not currently being monitored. Thus, those respondents who participated in the real-money survey treatments had a real opportunity to fund water-quality monitoring on the pond. Making a contribution involved a one-time payment to support water-quality monitoring on the pond for the upcoming monitoring season.

In both the hypothetical and real-money survey treatments, the WTP question was preceded by information regarding water-quality monitoring and the conditions under which the respondent was asked to donate money. Each respondent's survey contained information pertaining to either a no rebates, uniform proportional rebates, or extended benefits rule. The WTP question involved a "yes" or "no" response of the form depicted in Figure 1. Across all survey treatments, individuals faced a cost of $6, $8, $12, or $15.

In the real-money survey treatments, each respondent was assigned an arbitrary identification number. Real-money respondents were required to present this number in order to receive any money-back or rebates at the conclusion of the experiment. Experimental monitors collected all surveys and summed the contributions collected under each rebate rule. All cash transactions were handled on the same day before each respondent left the classroom.

Data Analysis

Data were analyzed using Hanemann's (1984) standard, utility-theoretic framework for estimating willingness to pay based on yes-no answers to discrete choice questions (such as in Figure 1). Within this random utility framework, the probability that individual respondent *i* will choose "yes," to monetarily support water quality monitoring on the pond, is modeled as follows:

$$\Pr(\text{Yes}) = \Pr(U_{i,Yes} > U_{i,N0}) ,$$

where

$$U_{i,Yes} = V(Q_{Yes}, \mathbf{S_i}, \mathbf{R}, M_i - C_i) + e_{i,Yes},$$

$$U_{i,N0} = V(Q_{N0}, \mathbf{S_i}, \mathbf{R}, M_i) + e_{i,N0},$$

Q_{Yes} represents the state of the world with monitoring on the pond, Q_{N0} represents the state of the world without monitoring on the pond, $\mathbf{S_i}$ is a vector of characteristics which describe individual *i*, $\mathbf{R}$ is a vector of variables indicating which payment rule a respondent faced, $V(\cdot)$ represents the deterministic component of utility that is econometrically measurable by the researcher, and e is the random or unobservable (to the researcher) component of utility. After rearranging terms, one has

$$\Pr(\text{Yes}) = \Pr(e_{i,N0} - e_{i,Yes} < V_{i,Yes} - V_{i,N0}).$$

If one assumes the *e*'s are independent and identically Type I Extreme Value distributed in standard form, then Pr(Yes) can be modeled as the following logit model (McFadden 1974):

$$\Pr(\text{Yes}) = \frac{1}{1+e^{-\Delta V}} ,$$

where $\Delta V = V_{i,Yes} - V_{i,No}$.

In estimating Pr(Yes), we hypothesized that the probability that a respondent would answer yes will be higher for both a uniform proportional rebate (UPR) and extended benefits (EB) payment rule than for a no rebate (NR) rule; will be lower for respondents facing a higher cost of monitoring the pond; will be higher for respondents who have declared an environmental major in college. Since environmental majors have expressed a preference to address environmental issues as professionals, it is reasonable to expect this factor to increase their willingness to pay (see also Spencer et al. 1998).

Finally, if one assumes V is linear in its parameters,

$$V_{i,Yes} = \beta'(\boldsymbol{Z_{i,Yes}}) + \beta_C C_{i,Yes},$$

then one can derive a utility-theoretic measure of WTP as follows (Hanemann 1984)

$$WTP = -(\beta'/\beta_C)\cdot(\boldsymbol{Z_{i,Yes}} - \boldsymbol{Z_{i,No}}) , \qquad \text{[equation (1)]}$$

where $\boldsymbol{Z_{i,Yes}} = z(Q_{Yes}, \mathbf{S_i}, \mathbf{R})$, β' is a vector of parameters, and β_C is the coefficient on individual cost for pond monitoring.[3] Estimates of β' and β_C can be obtained through maximum likelihood estimation.

Results and Discussion for Water Quality Monitoring

Table 1 describes the explanatory variables used in the behavioral model of pond monitoring, and Table 2 presents the logit estimation results for three specifications of the model. Specification 1 allows for varying intercepts and slopes across rebate rules (i.e., treatments), while specifications 2 and 3 represent reduced forms of specification 1 that only allow for varying slopes and intercepts, respectively, across rebate rules. For each specification, the χ^2-statistic for a likelihood ratio test of model significance is highly significant ($P< 0.001$). Across all specifications, the signs of the parameter estimates are as expected. Both *Hypothetical Cost* and *Real Cost* have negative signs that are highly significant. This indicates that the survey respondents are less likely to respond "yes" as the individual cost of pond monitoring increases. Dummy variables indicating the payment rule faced by a respondent, *UPR* and *EB*, both have positive signs, which indicate respondents are more likely to support pond monitoring under a uniform proportional rebates (*UPR*) or extended benefits (*EB*) rule than under a no rebates (*NR*) rule.[4] This result follows expectations. However, the effect of the *UPR* rule unexpectedly appears stronger than the *EB* rule.

[3] If cost is interacted with elements of R, then we mean hear that β_C includes the adjustment for this interaction.
[4] The dummy variable for NR is omitted to prevent colinearity with the intercept.

In specification 3, *UPR* is highly significant (*one*-tailed $P < 0.01$) while *EB* is only weakly significant (*one*-tailed $P < 0.09$).

Neither specification 2 ($\chi^2_{1vs.2}$-statistic = 1.0398, $P > 0.59$) nor specification 3 ($\chi^2_{1vs.3}$-statistic = 3.2758, $P > 0.51$) are statistically different from specification 1, suggesting that both specifications 2 and 3 are preferred to specification 1 based on a likelihood ratio test. However, since specification 2, unlike specification 3, allows one to obtain different hypothetical WTP to real WTP ratios across rebate rules, this discussion only reports WTP estimates for specification 2.

Table 3 reports both hypothetical and real WTP estimates for the average respondent based on the parameter estimates for specification 2 in Table 2. [5] For hypothetical WTP, the observed pattern suggests that $WTP_{NR} > WTP_{UPR} \approx WTP_{EB}$, where WTP_r is defined by equation one for the NR, UPR, or EB rule as indicated by subscript r. As expected, the UPR and EB rules yield higher hypothetical WTP estimates than the NR rule—in fact, both UPR and EB yield hypothetical WTP estimates that are nearly double those of the WTP estimates under the NR rule. The observation that $WTP_{UPR} \approx WTP_{EB}$, for the hypothetical (stated preference) respondents, is, however, unexpected. This result suggests that stated-preference willingness to pay may be robust with respect to the means by which excess contributions would be distributed if money were actually collected. Given these point estimates of hypothetical willingness to pay, it appears that leaving the issue of how excess contributions would be distributed may cause respondents to express a lower WTP, as illustrated by comparing the NR rule to the UPR and EB rules in table 3. Of course, the experimental design does not allow a comparison between these estimates of hypothetical WTP and the estimated WTP that might arise if a stated-preference survey omitted most of the details about how money could be collected and used. We leave this issue for future research.

For real-money WTP, the observed pattern suggests that $WTP_{NR} \approx WTP_{EB} < WTP_{UPR}$. As expected, $WTP_{UPR} > WTP_{NR}$, but unexpectedly WTP_{EB} is not greater than WTP with either the NR or UPR rules. Although the latter result seems counter-intuitive, based on experimental economics literature (e.g., Marks and Croson 1998) it may be that respondents did not value the extended benefits (EB) offered. In this experiment, any excess contributions were designated to help cover the general operating costs of the pond monitoring organization. Perhaps other forms of extended benefits, such as support for monitoring additional ponds, would have stimulated higher WTP estimates under the extended benefits rule. In contrast to the hypothetical WTP estimates, it appears that real WTP is not robust with respect to the disposition of excess contributions.

The last column of Table 3 reports the ratio of hypothetical WTP to real-money WTP for each rebate rule. All the ratios fall well within the range of ratios reported across the few other discrete-choice studies that compare hypothetical and real-money WTP for a public good (Foster, Bateman, and Harley, 1997). The NR and UPR rules have approximately the same hypothetical-to-real ratio, about 2, while the EB rule produces a ratio that is nearly twice that of the NR and UPR rules. Compared to Spencer, Swallow, and Miller (1998), who report a ratio of 4.67 for a water quality monitoring program, a ratio of 2 is a substantial improvement in the observed gap between hypothetical and real-money WTP.

[5] An interesting note is that specification 3 yields negative WTP estimates for the average respondent under the no rebate rule. This result could indicate that most non-environmental majors were not in the market for pond monitoring. However, willingness to pay differences from specification 3 remain consistent with those discussed for specification 2.

A tentative conclusion from this experiment is that the UPR rule appears preferable for use in real-money, contingent surveys. Experimental economics literature suggests that such rebate rules can result in valuation estimates that approximate well aggregate demand for an environmental good. Moreover, UPR produces a favorable ratio of hypothetical WTP versus real-money WTP, while the NR rule may be suspect in the case of a hypothetical survey that is worded in a fashion parallel to the corresponding real money survey. For these reasons, the UPR rule was adopted in experimental surveys concerning the valuation of wetlands development rights.

III. Experiments on the Valuation of Wetlands

The final set of experiments concerns the value of protecting wooded wetlands from development. These experiments constitute an attempt to compare stated preference (hypothetical-mone) values and real-money willingness to pay for wetlands in an actual "field experiment" involving randomly drawn citizens and a mail-survey format similar to formats commonly used in CVM. The wetland application is sufficient to raise challenges that may be associated with valuation of many environmental goods.

Wetland Focus Groups

Prior to and during survey development, we conducted extensive focus groups with citizens. These focus groups elicited citizens' views on what features of wooded wetlands would affect their willingness to pay for wetland protection. Focus group observations led to the following list of attributes considered most relevant to participants: size or acreage of the parcel, character of the surrounding land, availability of public access, role in expanding or connecting existing conservation areas or in establishing a new conservation area, relative diversity of wildlife, relative sustainability of wildlife habitat qualities, location of the parcel relative to roads, and the cost of protection.

In several focus groups, and in survey pretesting, a primary subject was the method by which to present a real-money survey question to potential respondents. Commonly, focus group or pretest participants expressed surprise after completing several preliminary parts of the survey and then finding a survey question that required that they contribute real money if their answer indicated a willingness to pay for protecting a wetland. These observations are difficult to quantify, but a general impression is worth noting for future researchers.

First, we attempted to convey, through cover letters and through instructions within the survey, that the survey was <u>not</u> a simple fundraising device, despite its request for money. This approach was soundly rejected by focus group participants who accepted that the survey was about valuing environmental goods, but failed to accept the idea that it was <u>not</u> a fundraising gimmick. In various revisions of the survey materials, we attempted to identify a timing (or location within the materials) that would mitigate these concerns, but we judged that the fundamental approach was ineffective.

We then altered the approach by dropping all claims that the survey was not about fundraising. Our cover letter included a statement that wetlands may be important to some citizens, as are the services of development, and that in some instances it is important for towns or state agencies to compare the value of wetlands to the value of development "in economic terms." The cover letter did not explicitly state that the survey would request real payments of money, but it did explicitly state that the questions involved money. Within the survey, we developed an introduction

to the real-money questions that reminded respondents that our purpose was to ask about the value of wooded wetlands to the respondent's household. At this point, we stated that we would be asking about their willingness to pay a particular amount to protect one of two wetland parcels. This information, within the survey, emphasized that whatever choice the respondent made, it was important to provide us with an answer.

The survey then provided additional details. These details described the method of protection for the wetland and indicated that protection for a particular wetland would occur ". . . *only if* enough respondents decide to financially support the wetland parcel you chose." This latter phrase is indicative of a provision point for the public good, but it does not specify a specific dollar amount for the provision of the wetland protection program for a parcel.[6] Prior to mailing the survey, we decided that the threshold for provision of the wetland protection contract would be equal to one year's payment on a 10 year conservation agreement with the respective landowners (amounting to $300 and $400 for the parcels located in the towns of Scituate and North Smithfield, respectively).

Next the survey details indicated that any excess money collected would be returned (rebated) to respondents who contributed, while any funds retained by the university would be used only for establishing a conservation contract for the wetland owners (while a research grant would cover all salaries and other costs at the university). The survey details also stated that, upon completion of the survey, contributors would either receive a letter indicating that a wetland conservation contract was executed and returning "your share of any left over funds" or a letter indicating that the contract was not executed and a refund of their entire contribution. While we did not explicitly state that we would give rebates to all contributors in proportion to the excess collected, focus group participants discussed their understanding of this refund mechanism in a manner consistent with the Uniform Proportional Rebate mechanism examined in the water-quality monitoring experiments above.

The Wetland Survey Experiment

The wetland survey experiment involves two landowners who have contracted with the university in a Land Conservation Contract that can be converted to a deed restriction that will legally prohibit development on a specified wetland parcel. One parcel is located in the Town of Scituate, Rhode Island, and is 29 acres, with medium wildlife habitat quality (as judged by a university biologist), located on local roads, an representing an opportunity to establish a new conservation area isolated from other legally protected lands. The second parcel is located in the Town of North Smithfield, Rhode Island, containing about 50 acres of wetlands on a 70 acre wooded tract. Given the locations of these parcels, the survey experiments focused on samples of residents in these towns.

Six survey-versions were designed and administered to a randomly selected sample of residents obtained from a commercial mail-marketing company. Table 4 summarizes the survey versions and samples, as well as response rates and collections of real money. All stated-preference

[6] We did not quantify the provision point because we had contracted with two different landowners for potential protection for different parcels of land at different costs and we decided not to examine whether differences in a quantitatively defined provision point would affect willingness to pay results in this field experiment. This decision leaves an open question for future research.

questions in all versions were presented in a contingent choice or choice experiment format (Opaluch et al. 1993; Adamowicz et al. 1998), wherein the respondent evaluates descriptions of two wetland parcels and then decides to state a preference to protect one of these parcels or to choose "neither" parcel and thereby forego protection on both parcels while saving any (hypothetical or real) monetary costs to their household.

Survey version 1 presented respondents with a single, real-money contingent choice question. Version 2 involved a hypothetical (stated-preference) contingent choice question, followed by a different, real-money question; prior to answering the hypothetical question, respondents were informed that a real-money question would follow. These two survey versions allow an evaluation of whether respondents altered their willingness to pay in switching between hypothetical and real-money questions within the same survey (version 2) and whether the opportunity to view a set of hypothetical substitutes would alter real-money willingness to pay in a split-sample design (version 1 versus version 2).

Versions 3, 4, and 5 all involved two hypothetical contingent choice questions. Version 3 was designed from comparability to real-money version 1. The main design feature here is that a single contingent choice question was presented in version 3, much as in version 1. However, after completion of the same demographic questions as presented in version 1 (and all versions), version 3 asked a second, hypothetical contingent choice question. This second question was included in version 3 in order to augment the sample of hypothetical responses.

Version 4 was motivated by concerns in the contingent valuation literature that respondents to hypothetical surveys may be motivated by a desire to show support for the environment. This motivation could be described as "yea-saying." In version 4, respondents were given three ways in which they could choose not to pay for conservation on either parcel (i.e., three versions of the "neither" choice): (a) choice to save the respondent's money rather than pay for conservation (a statement that was identical to the "neither" choice in other survey versions); (b) choice stating a general unwillingness to pay for wetland protection; and (c) a choice stating that protection of wetlands is valuable to the respondent's household, but that the respondent is unwilling to pay the specified costs at the time of response.

Version 5 is a straightforward format presenting two hypothetical contingent choice questions followed by the demographic questions. This version was intended to allow a comparison to the two questions (one hypothetical and one involving real-money) from version 2.

Finally, version 6 presented respondents with twelve hypothetical questions, primarily in an effort to gain additional data from hypothetical respondents. This motivation arises from results in Spencer et al. (1998) suggesting that the precision of willingness to pay estimation is poorer with stated-preference surveys than it is with real money surveys.

All versions of the survey were mailed following principles of the Dillman Total Design method, including an initial mailing of the survey, a reminder post card, a second mailing of the survey to non-respondents, a second reminder post card, and a follow-up letter to non-respondents with a brief, post-card survey collecting demographic information and a statement concerning the individual's choice to not respond. Inspection of table 4 shows that response rates (based on deliverable surveys) were generally in the range of 35% for real-money surveys (versions 1 and 2),

while response rates for surveys containing only hypothetical contingent choice questions (versions 3 through 6) were generally in the range of 45%.

Analysis and Results

In each contingent choice questions, wetland parcels were described according to the variables listed in Table 5. For hypothetical contingent choice questions, contingent choice questions were designed to cover the range of variables and combinations of their levels as described in the table. This design led to 36 different parcel-choice pairs in the hypothetical surveys. For real-money contingent choice questions, parcel descriptions were constrained to apply to all or at least a portion of the real wetland parcels that were contracted for this project. This constraint led to 9 parcel-choice pairs in the real-money surveys for Scituate, and 18 parcel-choice pairs in the real-money surveys for North Smithfield. The parcel-choice pairs were designed by a fractional factorial procedure provided by StatDesign, Inc.

At the time of this writing, the available data is limited to survey versions 2, 4, and 5 for both towns, and responses to version 1 from Scituate. These data are considered reliable, but are as yet unverified for coding accuracy. For the purpose of this preliminary analysis, the data were pooled for analysis using a multinomial logit version of the random utility model described above (see also, Adamowicz et al. 1998). This analysis assumes respondents stated a choice for a parcel, or for the neither-option, only if the chosen alternative would maximize their utility if implemented.

Three specifications of the choice model are presented in Table 6. All three models are statistically significant predictors of respondents' choices at a high level ($P < 0.01$). Specifications 1 and 3 include all of the attributes describing a parcel. Specification 2 is a model nested within Specification 1, omitting those parcel attributes that were found to be statistically insignificant in Specification 1 (omitting public access variables, variables for location on different types of roads, variables identifying whether the parcel connects to existing undeveloped land, the dummy variable for medium sustainability of wildlife habitat, and the dummy variable indicating whether the surrounding landscape is farmland).

The likelihood ratio test of Specification 2 versus the unrestricted Specification 1 fails to reject the null hypothesis that the coefficients on the variables dropped in specification 2 are zero (chi-square = 3.268, 8 df, $P>0.90$). Based on Specification 2, it appears that wetland parcels in residential landscapes were less preferred for conservation than were parcels in farmland or undeveloped-wooded landscapes, and it appears that parcels with low or medium diversity of wildlife or low sustainability of wildlife habitat were significantly less preferred for conservation as compared to parcels with high diversity.

Specification 1 allows the coefficient on cost to differ between real and hypothetical surveys, and for the version 4 format for the neither choice to lead to a willingness to pay estimate that is different than other versions of the hypothetical survey. Specification 3 restricts these cost coefficients such that there is a single coefficient on cost across all forms of the choice questions. A log-likelihood ratio test of this restriction, compared to unrestricted model 1, rejects the null hypotheses (chi-square = 20.539, 2 df, $P<0.005$). This preliminary analysis indicates that respondents treated money costs differently between the real-money surveys and the hypothetical-money surveys.

A comparison of the cost coefficient from real-money surveys to the cost coefficient from hypothetical surveys (other than version 4) indicates that respondents were less likely to state a willingness to pay to protect a parcel if the survey involved real money.[7] Based on these results, however, it is pre-mature to discuss whether this difference is significant. The coefficient on hypothetical costs is only significantly different from zero at a level of only about 11% for a one-tail test. Availability of the additional hypothetical responses (unavailable at this writing) may permit a more definitive comparison of these survey versions.

Finally, it is worth pointing out that respondents who contributed real money provided an average of $18 from North Smithfield, and an average of $16 from Scituate. Respondents from both towns contributed about $1000 to protect the wetland associated with their town. These contributions exceeded the pre-identified provision point of a one-year payment on the ten-year contract with landowners. Excess contributions are being rebated to these respondents.

[7] This conclusion is consistent with a few other preliminary tests, even though the variance on the coefficient of H_COST is rather high.

References

Adamowicz, W., P. Boxall, M. Williams, and J. Louviere. 1998. Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *American Journal of Agricultural Economics* 80(February):64-75.

Blackburn, M., G.W. Harrison, and E.E. Rutström. 1994. Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics* 76(December):1084-88.

Brown, T.C., P.A. Champ, R.C. Bishop, and D.W. McCollum. 1996. Which response format reveals the truth about donations to a public good? *Land Economics* 72(2):152-66.

Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum. 1997. Using donation mechanisms to value nonuse benefits from public goods. *Journal of Environmental Economics and Management* 33(2):151-62.

Cummings, R.G., S. Elliott, G.W. Harrison, and J. Murphy. 1997. Are hypothetical referenda incentive compatible? *Journal of Political Economy* 105(3):609-21.

Cummings, R.G., G.W. Harrison, and E.E. Rustrom. 1995. Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review* 85(1):260-66.

Cummings, R.G. and L.O. Taylor. 1997. Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. Typescript, Department of Economics, School of Policy Studies, Georgia State University. August.

Foster, V., I.J. Bateman, and D. Harley. 1997. "Real and Hypothetical Willingness to Pay for Environmental Preservation: A Non-Experimental Comparison." *Journal of Agricultural Economics* 48:123-138.

Fox, J.A., J.F. Shogren, D.J. Hayes, and J.B. Kliebenstein. 1995. CVM-X: Calibrating contingent values with experimental auction markets. Kansas Agricultural Experiment Station Contribution no. 95-470-J. Manhattan: Kansas State University. November.

Hanemann, W.M. 1984. Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics* 66(August):332-341.

Hoehn, J.P. 1991. Valuing the multidimensional impacts of environmental policy: Theory and methods. *American Journal of Agricultural Economics* 73:289-299.

Hoehn, J.P., and A. Randall. 1989. Too many proposals pass the benefit cost test. *American Economic Review* 79:544-551.

Johnston, R.J., S.K. Swallow, and T.F. Weaver. 1999. Estimating willingness to pay and resource tradeoffs with different payment mechanisms: An evaluation of a funding guarantee for watershed management. *Journal of Environmental Economics and Management* 38:97-120.

Marks, M., and R. Croson. 1998. Alternative rebate rules in the provision of a threshold public good: An experimental investigation. *Journal of Public Economics* 67:195-220

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, ed. P. Zarembka, pp. 105-142. New York: Academic Press.

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

Neill, H.R., R.G. Cummings, P.T. Ganderton, G.W. Harrison, and T. McGuckin. 1994. Hypothetical surveys and real economic commitments. *Land Economics* 70(2):145-54.

Opaluch, J.J., S.K. Swallow, T.F. Weaver, C.R> Wessells, and D. Wichelns. 1993. Evaluating impacts from noxious facilities: Including public preferences in current siting mechanisms. *Journal of Environmental Economics and Management* 24(January):41-59.

Rondeau, D., W.D. Schulze, and G.L. Poe. 1999. Voluntary revelation of the demand for public goods using a provision point mechanism. *Journal of Public Economics* 72:455-470.

Seip, K., and J. Strand. 1992. Willingness to pay for environmental goods in Norway: A contingent valuation study with real payment. *Environmental and Resource Economics* 2(1):91-106.

Shogren, , J.F. 1993. Experimental markets and environmental policy. *Agricultural and Resource Economics Review* 22(October):117-29.

Shogren, J.F., J.A. Fox, D.J. Hayes, and J.B. Kliebenstein. 1994. Bid sensitivity and the structure of the Vickrey auction. *American Journal of Agricultural Economics* 76(December).

Spencer, M.A., S.K. Swallow, and C.J. Miller. 1998. Valuing water quality monitoring: A contingent valuation experiment involving hypothetical and real payments. *Agricultural and Resource Economics Review* 27 (April):28-42.

Swallow, S.K. 1994. Value elicitation in laboratory markets: Discussion and applicability to contingent valuation. *American Journal of Agricultural Economics* 76(December):1096-1100.

<u>Review the pond described below and then check</u> the appropriate box (in the middle of the page) indicating whether you do or do not prefer to have U.R.I. Watershed Watch monitor the pond. <u>Note</u>: we ask that you please make an <u>actual</u> monetary payment, here and now, if you prefer to have the pond monitored.

DESCRIPTION OF POND

R 46 acres,
Suspected to have average to poor water quality,
Located inland,
2 Surrounded by wooded area,
\+ Is accessible to public,
Has low to moderate fishing usage, and
Monitoring will help find source(s) of current problem.

<u>Cost to you</u>:
We ask that you pay **$8**, here and now, to help fund monitoring for next season.

⇓

Please check only ***<u>one</u>*** box below:

❑ **<u>Yes</u>**, I prefer to pay the $8 and have the pond monitored.		❑ **<u>No</u>**, I prefer to pay nothing and <u>not</u> have the pond monitored.
⇓		⇓
If you choose **"Yes"** to having the pond monitored, ***please turn-in an envelope <u>filled with</u> $8 cash or a check for $8 (payable to U.R.I. Watershed Watch)*** with your completed survey. If you do not include such funds, we will assume that, at this time, you prefer not to monitor the pond.		If you choose **"No"** to having the pond monitored, ***please turn-in an <u>empty</u> envelope with your completed survey***.

Figure 1. Example of WTP Question

Table 1. Description of Explanatory Variables

Variable	Description
Hypothetical Cost	The required contribution (or cost) for an individual in the *hypothetical-money* survey treatment to support water quality monitoring on the pond. [a]
Real Cost	The required contribution (or cost) for an individual in the *real-money* survey treatment to support water quality monitoring on the pond.
UPR [b]	Indicator variable = 1 if *uniform proportional rebate* rule; = 0 otherwise.
EB	Indicator variable = 1 if *extended benefits* rule; = 0 otherwise.
UPR × *Hyp. Cost*	Interaction term between the *uniform proportional rebate* rule and the required *hypothetical-money cost* to support pond monitoring.
EB × *Hyp. Cost*	Interaction term between the *extended benefits* rule and the required *hypothetical-money cost* to support pond monitoring.
UPR × *Real Cost*	Interaction term between the *uniform proportional rebate* rule and the required *real-money cost* to support pond monitoring.
EB × *Real Cost*	Interaction term between the *extended benefits* rule and the required *real-money cost* to support pond monitoring.
Environ. Major	Indicator variable = 1 if environmental major; = 0 otherwise. [c]

[a] Notice that the no rebate (*NR*) rule is used as the base level.

[b] For both the hypothetical-money and real-money survey treatments, individuals faced a cost of $6, $8, $12, or $15.

[c] Environmental majors included animal science and technology, aquaculture and fishery technology, biological sciences, biology, clinical laboratory science, environmental science, marine biology, microbiology, plant science, water and soil science, wildlife biology, and zoology.

Table 2. Estimation Results for Various Specifications of the Logit Model

Variable	Specification 1		Specification 2		Specification 3	
	Parameter Estimate	Pr > |*Z*| (*P*-value)[a]	Parameter Estimate	Pr > |*Z*| (*P*-value)	Parameter Estimate	Pr > |*Z*| (*P*-value)
CONSTANT	–0.4796	0.4381	0.0463	0.8902	–0.3511	0.3506
	(–0.775)		(0.138)		(–0.933)	
Hypothetical Cost	–0.0571	0.3362	–0.1043	0.0061	–0.0700	0.0306
	(–0.962)		(–2.741)		(–2.162)	
REAL COST	–0.1595	0.0336	–0.2123	0.0002	–0.1754	< 0.001
	(–2.124)		(–3.715)		(–4.345)	
UPR	0.7553	0.3568			0.6549	0.0171
	(0.921)				(2.385)	
EB	0.7416	0.3898			0.3913	0.1654
	(0.860)				(1.387)	
UPR × *Hyp. Cost*	–0.0214	0.7862	0.0462	0.1256		
	(–0.271)		(1.532)			
EB × *Hyp. Cost*	–0.0211	0.7985	0.0453	0.1384		
	(–0.255)		(1.482)			
UPR × *Real Cost*	0.0193	0.8399	0.0933	0.0926		
	(0.202)		(1.682)			
EB × *Real Cost*	–0.0977	0.3831	–0.0218	0.7494		
	(–0.872)		(–0.319)			
Environ. Major	0.9304	0.0036	0.9386	0.0033	0.9528	0.0028
	(2.910)		(2.934)		(2.990)	
Total Observations	427		427		427	
Log-likelihood	–238.476		–238.996		–240.114	
χ^2-statistic	42.849	< 0.001[b]	41.809	< 0.001	39.573	< 0.001
Degrees of Freedom	9		7		5	

Note: numbers in parentheses represent *Z*-statistics.

[a] The *P*-values reported in this table correspond to a *two*-tailed test of H_0: $\beta = 0$ versus H_A: $\beta \neq 0$.

[b] The level of significance (i.e., *P*-value) for the χ^2-statistic.

Table 3. Willingness-to-Pay Estimates for the Average Respondent

Rebate Rule	Hypothetical WTP	Real-Money WTP	Ratio ($WTP_{Hyp.}/WTP_{Real}$)
No Rebates	\$1.58	\$0.78	2.03
Uniform Prop. Rebates	\$2.84	\$1.39	2.04
Extended Benefits	\$2.80	\$0.71	3.94

Note: the WTP estimates reported here are based on the parameter estimates to specification 2 in Table 2 and the average respondent, identified by the sample mean of the variable *Environ. Major*, which is 0.1265.

Table 4. Summary of Survey Treatments in Wetlands Survey

Town of North Smithfield	**Town of Scituate**
1NS: Real-Money Survey • Sample size = 700 • Deliverables = 629 • Response rate = 34% • 40 respondents were willing to pay for a parcel. Total amount sent in to preserve parcel in **North Smithfield** was $725.00	1S: Real-Money Survey • Sample size = 500 • Deliverables = 468 • Response rate = 39% • 40 respondents were willing to pay for a parcel. Total amount sent in to preserve parcel in **Scituate** was $640.00
2NS: Hypothetical-Real Money Survey • Sample size = 200 • Deliverables = 171 • Response rate = 30% 13 respondents were willing to pay for a parcel. Total amount sent in to preserve parcel in **North Smithfield** was $255.00	2S: Hypothetical-Real Money Survey • Sample size = 200 • Deliverables = 185 • Response rate = 35% 16 respondents were willing to pay for a parcel. Total amount sent in to preserve parcel in **Scituate** was $260.00
3NS: Two question Hypothetical Survey with Demographic Questions • Sample size = 129 • Deliverables = 115 • Response rate = 46%	3S: Two question Hypothetical Survey with Demographic Questions • Sample size = 271 • Deliverables = 256 • Response rate = 39%
4NS: Two question Hypothetical Survey with no responses allowing "support for environment" • Sample size = 150 • Deliverables = 136 • Response rate = 45%	4S: Two question Hypothetical Survey with no responses allowing "support for environment" • Sample size = 250 • Deliverables = 240 • Response rate = 50%
5NS. Two question Hypothetical Survey • Sample size = 100 • Deliverables = 93 • Response rate = 43%	5S. Two question Hypothetical Survey • Sample size = 100 • Deliverables = 94 • Response rate = 54%
6NS: Twelve-question Hypothetical Survey • Sample size = 100 • Deliverables = 95 • Response rate = 42%	6S: Twelve-Question Hypothetical Survey • Sample size = 100 • Deliverables = 95 • Response rate = 44%

Table 5. Description of Explanatory Variable for Wetlands Survey

Variable	Description
Wildlife diversity	Wildlife diversity of the parcel. Can be listed as one of the three: WILD_DVL-- Low wildlife diversity WILD_DVM-- Medium wildlife diversity WILD_DVH --High wildlife diversity
Access	The level of access the respondent would have to the parcel No_ACC—No access LIM_ACC—Limited access, permission required from owner FULL_ACC—Open access to users
Road	Type of road bordering the parcel LCL_RD—Locally traveled road CMN_RD—Commonly traveled road HVY_RD—Heavily traveled road
Sustainability	Indicates how much it was able to sustain habitat quality LOW_SUS—Low sustainability MED_SUS—Medium sustainability HI_SUS—High sustainability
Conservation	Indicates the role of the parcel as a conservation area CNSRVTNI—Parcel is isolated from other protected areas CNSRVTNE— Parcel is connected to an unprotected, undeveloped area CNSRVTNC—Parcel connects two protected areas
Surrounding land	Indicates the type of land surrounding the parcel SUR_LNDR—Parcel is surrounded by rural residential land SUR_LNDF—Parcel is surrounded by farmland SUR_LNDW—Parcel is surrounded by woodland
Size	Acreage of parcel (29, 45, 60)
Cost	The required contribution for a respondent to support protection of the parcel ($5, $10, $15, $20, $25, $30) R_COST—Dummy variable for real money interacted with the cost variable H_COST—Dummy variable for hypothetical money interacted with the cost variable LEGITCST—Dummy variable for legitimizing no's interacted with the cost variable

Table 6. Estimation Results for Various Specifications for Wetland Survey Data

Variable	Specification 1		Specification 2		Specification 3	
	Parameter Estimate	Pr>|Z| (P-value)	Parameter Estimate	Pr>|Z| (P-value)	Parameter Estimate	Pr>|Z| (P-value)
DMY_N (neither dummy)	-.1261 (-.393)	.6940	.1069 (.423)	.6725	-.0885 (-.272)	.7853
WILD_DVL (wildlife low)	-.5549 (-3.969)	.0001	-.5246 (-3.878)	.0001	-.5651 (-3.996)	.0001
WILD_DVM (wildlife medium)	-.4186 (-3.154)	.0016	-4137 (-3.184)	.0015	-.5189 (-3.875)	.0001
NO_ACC (no public access)	-.02988 (-.212)	.8319			.02094 (.148)	.8827
LIM_ACC (limited public access)	-.1288 (-1.015)	.3101			-.1251 (-1.011)	.3120
LCL_RD (local road)	.01205 (.078)	.9376			-.1106 (-.719)	.4720
CMN_RD (road commonly traveled)	.01897 (.120)	.9042			.0281 (.177)	.8595
LOW_SUS (low sustainable habitat)	-.5297 (-3.773)	.0002	-.4669 (-3.797)	.0001	-.5289 (-3.727)	.0002
MED_SUS (medium sustainable habitat)	-.1020 (-.784)	.4333			-.2215 (-1.704)	.0884
CNSRVTNI (isolated area)	-.1054 (-.726)	.4678			-.2474 (-1.714)	.0865
CNSRVTNE (expands conservation area)	-.04735 (-.309)	.7570			-.0354 (-.230)	.8184
SUR_LNDR (residential lands)	-.5328 (-3.895)	.0001	-.4564 (-3.734)	.0002	-.5820 (-4.273)	.0000
SUR_LNDF (farmland)	-.1364 (-1.020)	.3075			-.2175 (-1.644)	.1002
SIZE (acres)	.01762 (3.538)	.0004	.01741 (3.556)	.0004	.0236 (4.816)	.0000
COST					-.01357 (-2.368)	.0179
R_COST (COST x real dummy)	-.05226 (-4.717)	.0000	-.05714 (-5.560)	.0000		
H_COST (COST x hypothet. Dummy)	-.009651 (-1.295)	.1954	-.009015 (-1.239)	.2152		
LEGITCST (COST x version 4 dummy)	.008194 (1.049)	.2943	.008066 (1.039)	.2990		
Total Observations	847		847		847	
Log-Likelihood	-858.5940		-860.2280		-868.8636	
χ^2- statistic	124.633	<.005	121.365	<.005	104.094	<.005
Degrees of freedom	17		9		15	

Note: numbers in parentheses represent Z-statistics

ASK A HYPOTHETICAL QUESTION, GET A VALUABLE ANSWER?

by

Christopher D. Azevedo
Joseph A. Herriges
Catherine L. Kling*

Department of Economics

Iowa State University

June 12, 2000

This paper models the recreation demand for Iowa wetlands, combining survey data on both actual usage patterns (i.e., revealed preferences) and anticipated changes to those patterns under hypothetical increases in trip costs (i.e., stated preferences). We formulate and test specific hypotheses concerning potential sources of bias in each data type. We consistently reject consistency between the two data sources, both in terms of implied wetland values and underlying preference parameters. Careful attention is paid to the interpretations of the test results, noting particularly how the interpretation of the same results can vary with the "school of thought" of the reader.

* The authors would like to thank Dan Phaneuf, Kerry Smith, and Wally Thurman for helpful comments on earlier drafts of this paper. We would also like to thank seminar participants at North Carolina State University, the University of California-Davis, and the University of Kansas. This research was supported in part by the U.S. Environmental Protection Agency and by the Western Regional Research Project W-133. Although the research described in this article has been funded in part by the United States Environmental Protection Agency through R825310-01-0 to the authors, it has not been subject to the Agency's required peer review policy and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

"Economists have inherited from the physical sciences the myth that scientific inference is objective, and free of personal prejudice. This is utter nonsense." ¾ Leamer [1983, p. 36]

I. Introduction

Economists began investigating the use of surveys to elicit consumers' willingness-to-pay for public goods over three decades ago. The vast majority of the early applications concerned environmental goods, although numerous other public goods have been the subjects of valuation surveys. In the last decade, the use of surveys to elicit welfare values has come under debate by the profession at large. The now famous NOAA blue ribbon panel (Arrow, *et al.* [1993]) was assembled to consider the validity of these survey methods (also called contingent valuation or stated preference methods) and the *Journal of Economic Perspectives* ran a symposium with competing views on the topic (Diamond and Hausman [1994]; Hanemann [1994]; and Portney [1994]).

The fundamental question in this debate might be stated: can carefully designed survey methods provide informative data on consumers' willingness to pay for public goods, or does the hypothetical nature of these instruments render them irrelevant, regardless of how much attention is given to truth-revealing mechanisms in their construction? The initial reaction of most economists to this question is that hypothetical questions will generate hypothetical answers that likely bear little relation to the "true" answers. Critics of stated preference (SP) methods point to numerous potential sources of bias. For example, it has been argued that survey respondents may ignore or downplay their budget constraint in answering hypothetical questions (e.g., Arrow, *et al.* [1983]; Loomis, Gonzales-Caban, and Gregory [1994], and Kemp and Maxwell [1993]). Additional criticisms include concerns that SP-based willingness-to-pay estimates fail to vary sufficiently with the scope of the resource being valued, the so-called "embedding effect" (e.g., Desvousges, *et al.* [1993] and Kahneman and Knetsch [1992]), and that they are inordinately sensitive to the elicitation format used (e.g., McFadden [1994] and Diamond and Hausman [1994]).

Yet, there is also substantial evidence that answers to carefully designed surveys contain at least some information content. At the most basic level, Mitchell and Carson [1989, pp. 206-207] note that valuation estimates based upon SP methods are typically correlated in the expected direction with those independent variables that theory predicts should influence consumer preferences. Thus, the demand for an environmental improvement is usually found to decrease with the cost of its provision and increase with the income level of the respondent. Of course, internal validity checks are not sufficient to justify a reliance on SP-based valuations. To further investigate their validity, researchers have turned to comparisons to valuation estimates based upon revealed preference (RP) data, including actual market transactions for related goods (e.g., Cameron [1992a], and Adamowicz, Louviere, and Williams [1994]) or simulated market transactions obtained in a laboratory setting (e.g., Cummings *et al.* [1997], and Cummings and Taylor [1999]). The shear number of such comparisons has grown dramatically over the past two decades. In summarizing 83 studies with 616 such comparisons, Carson *et al.* [1996] find that the ratio of SP to RP valuations for essentially to same good ranges from 0.005 to 10.269. However, the ratio generally lies near 1.00 and the RP and SP estimates are highly correlated (with a rank correlation between 0.78 and 0.92). For many, including the now infamous NOAA panel, these results hold out the hope that useful information can be gleaned from properly designed SP surveys. The question remains as to how best

to use that information: in isolation, calibrated to similar revealed preference results, or combined with revealed preference data.

In this paper, we address this issue in the context of valuing recreational wetland usage in Iowa. The empirical analysis draws on a 1997 survey of 6000 Iowa residents, which elicited both revealed and stated preferences regarding the use of wetland regions in the state. Like previous efforts to combine RP and SP data (e.g., Cameron *et al.* [1999]; McConnell, Weninger, and Strand [1999]; and Huang, Haab, and Whitehead [1997]), we develop a joint model of individual responses to both the RP and SP questions and use the resulting framework to test for consistency between the two sources of information. In doing so, however, we also emphasize testing specific hypotheses concerning sources of bias in *both* revealed and stated preference data. In carrying out these sets of hypothesis tests, we carefully investigate alternative interpretations of the results, noting particularly how the interpretation of the same results can vary with the "school of thought" of the reader. These arguments are in the vein of Leamer's [1983] famous discussion of alternative interpretations of the same data by those with different initial views of capital punishment.

The remainder of the paper is divided into six sections. Section II begins by reviewing the various schools of thought that have emerged in the literature in interpreting RP and SP data. The underlying behavioral and econometric models are detailed in Section III, along with hypothesis tests designed to test for specific sources of bias in each data source. The Wetlands database is then described in Section IV, with the empirical specification of the model detailed in Section V. Section VI provides the results of the estimation and hypothesis testing and Section VI concludes the paper.

II. Schools of Thought

Nearly two decades ago, Leamer warned of the potential fragility of econometric analysis to hidden "whimsical assumptions" and the delusion of the "…goal of objective inference" [1983, p. 37], arguing that the initial beliefs of a researcher too often drive the conclusions gleaned from a database. These concerns are no less relevant today and nowhere more apparent than in the debate over the relative merits of stated and revealed preference methods in valuing nonmarket amenities. With literally hundreds of comparisons between RP and SP valuations appearing in the literature, the prior beliefs of the analyst often color the hypotheses tested and inferences made from conflicts between the two sources of data. In this section, we describe three basic schools of thought that seem to have emerged.

The first school of thought, which we will refer to as the "RP-lovers" sect, views SP preference data with skepticism. In its extreme form, the group denounces stated preference methods, characterizing its use in the policy arena as misguided. Proponents of this view argue that "…contingent valuation surveys do not measure the preferences they attempt to measure…" and that "…changes in survey methods are not likely to change this conclusion" (Diamond and Hausman [1994, p. 46]). The less extreme portion of this sect holds out some hope for stated preferences to inform policy makers, while holding onto the sanctity of revealed preference data. McConnell, Weninger and Strand [1999, p. 201] echo this sentiment when they consider a joint model of RP and SP data.

"In other words, the preference structure that governs the RP decision may be completely independent from the preference structure that governs the SP decision. If so, the information

gleaned from SP methods is arbitrary and can lead to unreliable estimates of trip value. On the other hand, when the underlying preference structure is identical, both the RP and SP decisions provide useful information for valuing environmental amenities."

Thus, SP methods are only of value if they are consistent with their RP counterparts. Indeed, the notion of "calibrating" welfare estimates obtained from SP data is essentially a tenant derived from this school of thought, as it implies a calibration to revealed preference estimates. A smaller, though no less devoted, sect is the "SP-Lovers" school of thought. Much like its counterpart, "SP-Lovers" view with skepticism results based on commonly used revealed preference techniques, such as hedonic pricing and travel cost methods, noting their many sources of bias and limited applicability to the valuation of environmental amenities. Travel cost methods suffer, among other things, from "…value-allocation assumptions related to multi-purpose 'visits'; dependence of costs on assumptions concerning fixed/variable direct travel costs, costs (benefit?) of time in travel and on-site; and problems in obtaining values which are appropriately 'marginal' vis-à-vis the site/activity in questions." (Cummings, Brookshire, and Schulze [1986, p. 95]). As Randall [1994, p. 88] argues, the trip quantities employed in a travel cost model may be "revealed," but the prices necessary to derive trip demand functions, and hence welfare estimates, are not. Consequently, the travel cost method "…yields only ordinally measurable welfare estimates…" and "…cannot serve as a stand-alone technique for estimating recreation benefits." Similarly, the hedonic pricing methods is plagued by "…persistent collinearity between 'important' variables and extraordinarily low explanatory power in regression equations" (Cummings, Brookshire, and Schulze [1986, p. 96]). At the very least, this school of thought would argue that SP methods deserve equal footing in the pantheon of valuation techniques. At its extreme, it would hold that RP preference methods are not only fraught with limitations, but are simply not capable of capturing the range of use and nonuse values associated with environmental changes, leaving SP methods as the only game in town.

In recent years, a third school of thought has emerged, arguing for a reliance on both revealed and stated preferences (e.g., Cameron [1992a], Adamowicz, Louviere, and Williams [1994], and Kling [1997]). These "agnostics" suggest a shift in focus away from viewing RP and SP as competing sources of values and towards seeing them as complementary sources of information. In this view, both data sources illuminate, though imperfectly, consumer preferences for environmental amenities of interest. Revealed preference methods bring the "discipline of the market" to stated preference valuations, whereas stated preference methods can shed light on consumer preferences for price and quality attribute levels that are not currently observed in the market (Cameron, [1992b]). Discrepancies between the individual parameter estimates obtained using RP and SP estimates are not necessarily indicative of a failure of either method, but instead suggestive that the two sources are working in correcting the limitations inherent in each method. This ecumenical view would not fall back on either data source as "correct", but rely instead on their combination as providing the best source of welfare estimates.

III. Behavioral and Econometric Models of Preferences

In developing behavioral models underlying revealed and stated preference data sources, we focus our attention on a single recreation good (q), though obvious extensions exist to the multiple good situations. Individual i's "true" preferences are assumed to result in an ordinary (Marshallian) demand equation given by:

$$q_i = f\left(p_i, y_i; b\right) + e_i \qquad (1)$$

where q_i denotes the quantity consumed by individual *i*, p_i denotes the associated price, y_i is the individual's income, and b is a vector of unknown parameters. The additive stochastic term is used to capture heterogeneity in individual preferences within the population. Ideally, the analyst would have available data on the actual price and quantity for environmental good in question, along with quality attributes for that good and individual socio-demographic characteristics. Unfortunately, for both revealed and stated preference models, environmental valuation efforts often rely upon imperfect and indirect measures for these variables.

Revealed Preferences

Revealed preference models, such as travel cost models, rely upon surveys to elicit information on the number of trips taken to the site in question (q_i^R) and to construct proxies for the price of traveling to the site (p_i^R), including the cost of expended time in traveling to the site. The demand relationship modeled using this data is assumed to take the form:

$$q_i^R = f^R\left(p_i^R, y_i^R; b^R\right) + e_i^R. \qquad (2)$$

As is often the case in the literature, the error term is assumed to be drawn from a normal distribution, with $e_i^R \sim N\left(0, s_R^2\right)$. Standard econometric procedures can then be employed to obtain consistent estimates of the parameters of the function accounting for the censoring from the left at zero. Specifically, maximum likelihood estimation would rely upon the log-likelihood function

$$LL^R = \sum_{i=1}^{n}\left(D_i^R \ln\left\{ s_R^{-1} f\left[\frac{q_i^R - f^R\left(p_i^R, y_i^R; b^R\right)}{s_R} \right] \right\} + \left(1 - D_i^R\right) \ln\left\{ \Phi\left[\frac{-f^R\left(p_i^R, y_i^R; b^R\right)}{s_R} \right] \right\} \right) \qquad (3)$$

where $\Phi(\cdot)$ and $f(\cdot)$ are the standard normal cdf and pdf, respectively, and $D_i^R = 1$ if $q_i^R > 0; = 0$ otherwise.

There are, of course, numerous reasons for discrepancies between the demand relationships in equations (1) and (2). First of all, trip data are typically quantities based upon survey questions asking the individual to recall their number of trips over the past year. Recall errors and rounding by survey respondent contribute to differences between both the quantities and the error distributions in the two equations. Second, and perhaps of greater concern, are the discrepancies between the actual visitation costs (p_i) and those constructed by the analyst (p_i^R). It is standard practice to compute travel costs as the sum of the out-of-pocket cost of visiting the site in question (typically using a fixed cost per mile times the round-trip distance to the site) and a proxy for the value of the time used in getting to the site. Time is most often valued as a fixed fraction of the individual's wage rate (Cesario [1976]).[1] The reliance on this, or similar, conventions, drives a wedge between the individual's true demand equation in (1) and the estimated relationship in (2). In turn, this has led a number of critics to argue that, because "…visitation costs are inherently subjective…any particular welfare estimate is in part an artifact of the particular conventions selected for imposition" and hence suspect (Randall [1994, pp. 88-89]).

[1] There have been a number of efforts to better capture the value of travel time by estimating the appropriate fraction of the individual's wage rate (McConnell and Strand [1981] and Bockstael, Strand, and Hanemann [1987]) or the monetary value of leisure time (Feather and Shaw [1999]). However, the use a fixed wage fraction (between twenty-five and fifty percent) continues to dominate the applied literature.

Stated preference data can take a variety of forms, including direct value elicitations (i.e., contingent valuations) or contingent behavior data (e.g., providing information on how individual trip quantities might change as the price per trip or site attributes changed). We will focus on the latter type of SP data. In particular, suppose that in the process of gathering RP data, the survey respondents are asked: "How many recreation trips would you have taken to the site if the cost per trip increased by $B?" In this case, the resulting database looks identical to the RP data set, providing both price (p_i^S) and quantity (q_i^S) data for each individual. The preferences revealed through these questions can be represented by the demand equation

$$q_i^S = f^S\left(p_i^S, y_i^S; b^S\right) + e_i^S . \tag{4}$$

where $p_i^S = p_i^R + B$ and $e_i^S \sim N\left(0, s_S^2\right)$. Estimation of the demand relationship in (4) alone would proceed much like in the RP setting, with a log-likelihood function identical to that in equation (3), except that *R* would be replaced with *S* in all of the superscripts and subscripts. Efficiency gains can be achieved by recognizing the likely correlation between RP and SP responses (i.e., between e_i^R and e_i^S). The appropriate log-likelihood function for the joint estimation is given by:[2,3]

$$LL = \sum_{i=1}^{n}\left\{ D_i^R\left[\ln f\left(\frac{q_i^R - f_i^R}{s_R}\right) - \ln\left(s_R\right)\right] + D_i^R D_i^S\left[\ln f\left(\frac{\left(q_i^S - f_i^S\right) - q^S\left(q_i^R - f_i^R\right)}{s_S\sqrt{1-r^2}}\right) - \ln\left(s_S\sqrt{1-r^2}\right)\right]\right.$$
$$\left. + D_i^R\left(1 - D_i^S\right)\left[\ln\Phi\left(\frac{-f_i^S - q^S\left(q_i^R - f_i^R\right)}{s_S\sqrt{1-r^2}}\right)\right] + \left(1 - D_i^R\right)\left(1 - D_i^S\right)\ln\int_{-\infty}^{\frac{-f_i^R}{s_R}}\int_{-\infty}^{\frac{-f_i^S}{s_S}} f_2\left(h_1, h_2; r\right)dh_1 dh_2\right\} \tag{5}$$

where $Corr\left(e_i^R, e_i^S\right) = r$, $q \equiv r s_S / s_R$, $f_i^k = f^k\left(p_i^k, y_i^k; b^k\right)$ (k = *R*, *S*), and $f_2\left(\cdot,\cdot; r\right)$ denotes the standard bivariate normal pdf. This model can be used to test a variety of hypotheses concerning the consistency of the RP and SP data.

As with the RP data, there are numerous reasons one might expect preferences revealed by SP questioning, and modeled in equation (4), to differ from the individual's true preferences. As noted above, one of the most common concerns is that individuals fail to adequately account for their budget constraint and the availability of substitute commodities in responding to hypothetical scenarios presented to them in a survey. Thus, the income they perceive as being available for allocation towards the commodity in question may be inflated (i.e., $y_i^S > y_i$). In addition, individuals may have greater uncertainty regarding their preferences under unfamiliar price and quality settings; i.e., one might expect that $e_i^S = e_i + h_i^S$, where h_i^S captures the additional uncertainty or variability in preferences under the hypothetical conditions.

---

[2] The derivation of this log-likelihood function is available from the authors upon request.

[3] Louviere [1996] refers to this integration of RP and SP data sets having the same form as *pooling* models. A second category of models, *combining* models, integrates information from RP and SP data sets with different underlying forms. For example, Cameron [1992] combines RP fishing trip demand data with discrete choice SP data on whether individuals would take any fishing trips if a lump-sum access fee was imposed. Similar efforts have been undertaken by Larson [1990] and Huang, Haab, and Whitehead [1997].

Hypothesis Testing

There are two primary reasons for jointly estimating the RP and SP demand equations. The first and least ambitious is to garner improved efficiency in the parameter estimates by recognizing the likely correlation between the two data sets. It is hard to argue with this objective regardless of the school of thought from which one starts.

The second reason for combining the two models is to allow for hypothesis testing. Unfortunately, both the tests conducted and the interpretation of the results often depends upon the perspective brought to exercise. For example, the most common hypothesis tested is one that imposes complete consistency between the two models; i.e.,

$$H_0^1 : \beta^R = \beta^S \text{ and } \sigma_R = \sigma_S . \qquad (6)$$

Traditionally, rejection of this hypothesis has been widely viewed as a repudiation of the SP data and methodology, though this is clearly an "RP Lovers" perspective. An equally valid viewpoint is that the discrepancies arise due to problems with the RP data; i.e., the "SP Lovers" perspective. These viewpoints also influence which model one falls back on for welfare analysis in the event that H_0^1 is rejected. Obviously, "RP Lovers" and "SP Lovers" would fall back on their respective preferred models. The "agnostics", on the other hand, might argue that discrepancies between the two models should in fact be expected. The RP data is bringing the discipline of the market to bear on the SP data and the SP data is examining regions of consumer preferences outside of range of consumers' experiences in the marketplace. From this perspective, welfare analysis is best served by relying upon the combined model using both sources of information.[4]

A modification to the overall consistency test allows for differences in the underlying variation in the two models, while constraining the demand parameters themselves to be equal; i.e.,

$$H_0^2 : \beta^R = \beta^S . \qquad (7)$$

Allowing for differences in the distribution of preferences in the revealed and stated preferences models has been suggested by a number of authors, including Haab, Huang, and Whitehead [1999] and Cameron *et al.* [1999].

In addition to testing for overall consistency between the two models, one can, as suggested by Cameron [1992b], focus on specific weakness of each model. For example, one might generalize the SP demand equation as follows:

$$q_i^S = f^S\left(p_i^S, \tau_y y_i^R; \beta^S\right) + \varepsilon_i^S . \qquad (8)$$

This replaces y_i^S in equation (4) with $\tau_y y_i^R$, where τ_y is a parameter to be estimated. The argument that respondents to SP questions tend to downplay their budget constraints is then reflected by the expectation that $\tau_y > 1$. The hypothesis test in this case would correspond to restricting (2) and (8) such that

[4] A weaker form of "agnosticism" might argue for presenting the welfare estimates obtained from the unconstrained RP and SP models as bounding the true values. An alternative ("atheistic") perspective is, of course, to view the rejection of H_0^1 as a rejection of both data sources, suggesting that neither model be used for welfare analysis.

$$H_0^3 : \beta^R = \beta^S \text{ and } \sigma_R = \sigma_S, \text{ while } \tau_y \text{ remains unrestricted.} \quad (9)$$

This hypothesis test essentially treats the RP data as the base against which the SP data are tested.

An alternative approach is to reverses the roles of the RP and SP data. If the analyst believes the SP data are correct, but the RP data are subject to error, then the SP data can be used as the basis for a validity test of the RP data. As noted above, Randall [1994] has argued forcibly that the price term in RP data is poorly measured and is likely the cause of significant bias. An external validity test of the RP data can then be performed by substituting p_i^R in the RP models with $\tau_p p_i^R$, where $\tau_p > 0$, and generalizing equation (2) by using

$$q_i^R = f^R\left(\tau_p p_i^R, y_i^R; \beta^R\right) + \varepsilon_i^R. \quad (10)$$

Again, if τ_p is not estimated to be significantly different from one, external validity for the RP data would not be rejected. The corresponding hypothesis test would test the following restriction on equations (4) and (10):

$$H_0^4 : \beta^R = \beta^S \text{ and } \sigma_R = \sigma_S, \text{ while } \tau_p \text{ remains unrestricted.} \quad (11)$$

All four of the hypothesis tests identified above can be viewed as restrictions on the generalized system of demand equations in (8) and (10), with:

- $\tilde{H}_0^1 : \beta^R = \beta^S$, τ_y=1, τ_p=1, and $\sigma_R = \sigma_S$; i.e., complete consistency.
- $\tilde{H}_0^2 : \beta^R = \beta^S$, τ_y=1, and τ_p=1; i.e., consistency in demand parameters but not in terms of error variances.
- $\tilde{H}_0^3 : \beta^R = \beta^S$ and τ_p=1; i.e., when respondents answer the stated preference questions, in addition to having a different error variance, they also ignore their budget constraint. Consistency holds in all other respects.
- $\tilde{H}_0^4 : \beta^R = \beta^S$ and τ_y=1; when respondents answer the revealed preference question, in addition to having a different error variance, they also do not treat the computed travel cost term (*p*) as the cost of accessing the recreation site (analysts have calculated the incorrect price). Consistency holds in all other respects.

IV. Data

The data used to illustrate the above pooling model are drawn from a 1997 survey of 6,000 Iowa residents concerning their use of the state's wetland resources. The goal of the survey was to elicit information from respondents about their visits to, knowledge about, and attitude towards both the existing wetlands in Iowa, as well as efforts to preserve and expand these resources. Of the deliverable surveys, there was a fifty-nine percent response rate (with 3,143 surveys returned).[5]

Our analysis draws on the first section of the survey, which focused on visits to wetlands during the past year. After carefully defining what was meant by a "wetland," the survey asked respondents to recall the number of trips they had taken during the past year to wetlands in each of

[5] 594 surveys were returned by the post office as undeliverable. Details of the survey design and administration can be found in Azevedo [1999].

fifteen possible zones (see Figure 1). These zones, defined along county lines for convenience of the survey respondent, were designed to reflect major types of wetlands within the state.[6] Responses to this first question provide the basis for the revealed preference variable q_i^R. Individuals were then asked how their pattern of usage would change if the cost of visiting wetland zones near their residence (X, Y, and Z) increased.[7] In particular, they were asked to "…[c]onsider all of the recreation trips you made to wetlands in zones X, Y, and Z in 1997. Suppose that the **total cost per trip of each of your trips** to these areas had been $B more (for example, suppose that landowners charged a fee of this amount to use their land or that public areas charged this amount as an access fee)."[8] They were then asked to detail how their behavior would have changed with the increased cost, both in terms of reductions in visitations to areas X, Y, and Z, and changes in visits to other zones within the state. These questions provide the basis for constructing the SP quantity variable q_i^S.

While the surveys provide direct information on the trip quantities, the travel costs themselves must be constructed. The first step in the process was to establish travel time (t_i^z) and travel distance (d_i^z) for visits to the wetland zones (z = 1,…,15). One of the survey questions asked the respondent to place an X on a map, similar to Figure 1, locating their most recent wetland visit. The longitude/latitude coordinates for the visitation points in each of the 99 counties were averaged to find the mean visitation point in that county. Travel time and distance, for each respondent from their residence to each of the 99 mean visitation points, was then constructed using PC Miler, a software package designed for use in the transportation and logistics industry. This gave us a data set with 99 travel times and distances for each respondent in the database. Finally, zonal travel times (t_i^z) and distances (d_i^z) were calculated as a weighted average of their respective county travel county level values.[9]

For the purposes of this paper, we have focused our attention on a subset of the survey sample; i.e., those households in the Prairie Pothole Region of north central Iowa (zones 4, 5, and 8). We consider the aggregate number of trips to this region, with $q_i^k = q_i^4 + q_i^5 + q_i^8$ (k = R, S). Travel times and distances were formed as weighted averages of the zone specific values, where the weights used for individual i were the average percentage of trips to each zone among individuals in i's zone of residence. On average, for the 278 households with completed surveys in the Prairie

[6] For example, zones 4, 5, and 8 represent the prairie pothole region of north central Iowa. Pothole wetlands are the result of glacial activity and characterized by depressions in the land, most of which are less than two feet deep and filled with water for at least part of the year. In contrast, riverine wetlands dominate regions 1 through 3 and 13 through 15, and are associate with marshy land near rivers and streams.

[7] The wetland zones (X, Y, and Z) were assigned to each individual based upon the region of the state in which they lived. Specifically, the fifteen zones in **Figure 1** were grouped into five "megazones," reflecting regional wetland areas. The megazones were defined as the Missouri River Region (1,2,3), the Prairie Pothole Region (4,5,8), the Iowa River Corridor Region (9,10,11), the Mississippi Region (13,14,15), and the remainder of the state (6, 7, 12). The zones (X, Y, Z) for a given survey respondent consisted of those zones defining the megazone in which the respondent resided.

[8] The value of $B varied across the individuals surveyed, with bid values $5, $10, and $15 each randomly assigned to 20% of the sample and bid values of $20, $30, $40, and $50 each assigned 10% of the sample.

[9] Each county's weight was determined by the percentage of trips within that county's zone taken to that county. For example, zone three is made up of Pottawattamie, Mills, and Fremont counties. There were 92 trips taken to zone three. 41 trips were taken to Pottawattamie County, 24 trips were taken to Mills County, and 27 trips were taken to Fremont County. Therefore, the weight for Pottawattamie County was 0.45, the weight for Mills County was 0.26, and the weight for Fremont County was 0.29.

Pothole region, 8.2 trips (q_i^R) were actually taken. Respondents indicated that they would average only 2.7 trips (q_i^S) with the average hypothetical price increase (\$B) of \$27 per trip.

V. Model Specification

The parametric specifications for the demand functions in (8) and (10) were chosen to be of a simple linear form, with

$$q_i^R = b_0^R + b_p^R t_p p_i^R + b_y^R y_i^R + e_i^R \tag{12}$$

and

$$q_i^S = b_0^S + b_p^S p_i^S + b_y^S t_y y_i^S + e_i^S . \tag{13}$$

Clearly, given the linear form in equation (12), the parameters b_p^R and t_p are not separately identified. b_y^S and t_y are likewise not separately identified in equation (13). To deal with this issue, and to simplify the exposition below, we reparameterize the model as follows:

$$q_i^R = \tilde{b}_0^R + \tilde{b}_p^R p_i^R + \tilde{b}_y^R y_i^R + e_i^R \tag{14}$$

and

$$q_i^S = k_0 \tilde{b}_0^R + k_p \tilde{b}_p^R p_i^S + k_y \tilde{b}_y^R y_i^S + e_i^S . \tag{15}$$

In this case, the k_j's measure the discrepancies between the RP and SP demand equation parameters, with $k_j = 1 \; \forall j$ and $s_R = s_S$ corresponding to complete consistency between the two models.

Finally, the trip prices p_i^R and p_i^S must be specified. As suggested by McConnell and Strand [1981], we parameterize these prices as:

$$p_i^k = c_i + l^k L_i \quad k = R, S \tag{16}$$

where l^k is a parameter to be estimated, $c_i \equiv 0.21 d_i$ denotes the travel cost (computed as roundtrip travel distance times 21¢ per mile), and $L_i = w_i t_i$ denotes full time costs (computed as the individual's wage rate times the roundtrip travel time). As noted above, it is standard practice in the recreation demand literature to fix l^k at some value, typically between 0.25 and 0.5, rather than estimating its value.

Substituting (16) into equations (14) and (15) results in the following system of demand equations:

$$q_i^R = \tilde{b}_0^R + \tilde{b}_p^R \left(c_i + l^R L_i \right) + \tilde{b}_y^R y_i^R + e_i^R \tag{17}$$

and

$$q_i^S = k_0 \tilde{b}_0^R + k_p \tilde{b}_p^R \left(c_i + k_l l^R L_i \right) + k_y \tilde{b}_y^R y_i^S + e_i^S . \tag{18}$$

where $k_l \equiv l^S / l^R$. The hypothesis tests corresponding to those outlined above become:

- Complete Consistency. $\tilde{H}_0^1 : k_j = 1 \; \forall j$ and $s_R = s_S$.
- Heteroskedasticity Hypothesis. $\tilde{H}_0^2 : k_j = 1 \; \forall j$.
- Price Hypothesis. $\tilde{H}_0^3 : k_0 = 1, \; k_l = 1, \text{ and } k_y = 1$.

- Income Hypothesis. $\tilde{H}_0^4 : k_0 = 1,\ k_p = 1$ and $k_I = 1$.

VI. Results

The system of demand equations in (17) and (18) was estimated using standard maximum likelihood procedures, with the likelihood function in (5) correcting for the potential censoring at zero for either or both RP and SP trip data and accounting for possible correlation between the two error terms. The resulting parameter estimates are provided in Table I for the unconstrained model and for each of the constrained specifications outlined above. For the stated preference parameters, we provide both the estimated values for the k_j's and the implied levels for the stated preference parameters $\tilde{b}_j^S \equiv k_j \tilde{b}_j^R$ in square brackets.

Focussing first on the unconstrained model, we note that all of the parameters have the expected signs. The price coefficient for the revealed preference data is just above negative one and statistically significant. While the income coefficient is small and statistically insignificant, it is positive and with a small standard error. Interestingly, the coefficient on the wage rate, which is traditionally fixed at level between 0.25 and 0.50, is also close to zero. Given these parameter estimates, the probability that l^R exceeds 0.25 is less than 0.1%. Indeed, our estimate of l^R is not statistically different from zero at any reasonable confidence level.

The stated preference parameters differ from their revealed preference counterparts in a number of areas. In particular, both the intercept and price terms are significantly different, with both k_0 and k_p significantly less than one. This indicates, for example, that the price response in the SP data set is much smaller than in the RP counterpart. The remaining k_j's are all greater than one, though not significantly so. The estimate of k_I suggests that greater weight is placed on the wage rate in the stated preference choices than in their revealed preference counterpart, with the point estimate for $l^S = 0.88$. The underlying variability of preferences revealed by the two data sets is similar, with $k_s \equiv s_S / s_R$ estimated to be close to one and precisely measured. This result is in contrast to other efforts integrating RP and SP data sets that found the source of inconsistency between the two data sets to be in the degree of preference dispersion (e.g., Cameron *et al.* [1999] and Haab, Huang, and Whitehead [2000]). Finally, we note that there is substantial correlation between the two data sources, with $r = 0.72$ and significantly different from zero.

While parameters of the demand equations are important, they are typically used in the policy arena as an intermediate step to the computation of welfare impacts. As one might expect, the substantial differences in price coefficients between the RP and SP models lead to substantial differences in the associated welfare measures. In Table I, we provide estimates of the average consumer surplus associated with wetland visitations. The surplus derived from the SP model (CS_S) is over two and half times that obtained using the revealed preference model (CS_R).

Turning to the constrained models, we find that each of the model restrictions is rejected at any reasonable significance level. Consistency between RP and SP data is not borne out by the data, either in its strict form or when we allow for different levels of variances for two data sources. Indeed, even in its unconstrained form, the demand systems suggest $s_R = s_S$ (with $k_s = 1.04$). Less

restrictive forms of consistency are no more successful. The price hypothesis, which an SP-Lover might argue for, suggests that the source of the inconsistency between the two demand systems lies in the formation of the trip price. Yet this hypothesis is soundly rejected. Similarly, the income hypothesis, which our RP-Lover might suggest, argues that the source of the inconsistency lies in the lack of attention paid to income by stated preference survey respondents, but this hypothesis is also rejected.

The problem, of course, is where do we go from here? We have two sources of data with conflicting conclusions regarding the value of wetland use. One can always hope that the two valuations lead to the same policy conclusions. If they do not, one's "school of thought" regarding the RP and SP controversy will play a key role. Moderate RP- and SP-Lovers are likely to stop at the unconstrained model, taking solace in the improved efficiency achieved by jointly estimating the two demand systems, and relying on their respective valuations (CS_R and CS_S). Indeed, they may never even estimate the various constrained models. The zealots within each group, however, will rely on tests as a source of vindication for their respective positions. For the RP-Lover, the various consistency tests only serve to solidify their suspicion of the stated preference approach. However, these same results can be employed by the SP-Lover to argue against revealed preference models by simply reversing the presumption as to which is the correct model!

In contrast, the agnostic's approach to the results in Table I is likely to be substantively different. Rather than falling back on the unconstrained model in the face of the various rejections of consistency, they are more like to embrace the specification under $\tilde{H}_0^1$ or $\tilde{H}_0^2$. According to the agnostic, the rejection is expected. Indeed, the agnostic might even be disappointed if the constrained parameter estimates were not significantly different from the unconstrained ones, as that would indicate that the gains sought by combining data sources did not occur! That is, the SP data was not carefully enough designed to "fill in" missing pieces from the RP data and/or the RP data had sufficient measurement error or other problems such that it could not impose market discipline on the SP data. The agnostic's preferred strategy, would be to jointly estimate the models, imposing consistency across all parameters, or possibly allowing for difference variances (the complete consistency or heteroskedasticity hypothesis in Table 1). By so doing, all information available on consumers' preferences is best utilized.

VII. Summary and Conclusions

This paper has investigated the combining of revealed and stated preference data to jointly estimate the parameters of consumer preferences for a public good. We have presented a simple model for combining two different data sources and demonstrated how several hypotheses concerning these data sources can be formalized and tested in such a model. In the wetlands application discussed here, consistency between the RP and SP data was consistently rejected. In interpreting the lack of consistency found between the estimated parameters, we identified and discussed several "schools of thought." Analysts who identify with RP-Lovers will be inclined to take parameter and welfare estimates from RP data as the most accurate representation of the underlying true parameters, interpreting consistency tests with SP data as evidence for or against the validity of SP as a method. In contrast, SP-Lovers will view welfare and parameter estimates from SP data as the best available estimates and interpret consistency tests with RP data as referendums on RP's validity. In contrast, those identifying with an "agnostic" approach will be inclined to see

consistency tests as relatively unimportant, but instead prefer to combine data sources to estimate a single set of parameters and welfare estimates, thus enjoying the strengths of each method. Clearly in each of these three approaches, the role and interpretation of consistency tests in nonmarket valuation is closely related to the prior point of view of the analyst.

The fundamental debate concerning the legitimacy of SP methods relative to RP discussed in the introduction will continue to be much discussed. However, we are not optimistic that consistency tests between RP and SP data such as those described here will play a key role in settling the disagreement between those who see SP data as inherently flawed and those who see it as the last, best hope for nonmarket valuation. It is simply too easy to interpret the results of these tests to be consistent with one's prior views.

On the other hand, combining data sources holds considerable promise for those who are willing to credit each method with strengths and weaknesses. By designing questionnaires to elicit RP and SP data to take advantage of their respective strengths and then combining the data sources in estimation, analysts may be able to provide policy makers with more efficient and accurate estimates of the value of public goods. This research agenda is only begun and will require significant effort on the part of environmental economists to identify the parameters under which improved welfare estimates can be obtained through such methods.

References

Adamowicz, W., J. Louviere, M. Williams. "Comparing Revealed and Stated Preference Methods for Valuing Environmental Amenities" *Journal of Environmental Economics and Management*, May 1994, 271-292.

Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner, and H. Schuman. "Report of the NOAA Panel on Contingent Valuation." *Federal Register,* 58(10), 1993, 4601-4614.

Azevedo, C. D., "Linking Revealed and Stated Preference Data in Recreation Demand Modeling." Ph.D. dissertation, Iowa State University, 1999.

Bockstael, N., I. Strand, and M. Hanemann. "Time and the Recreation Demand Model." *American Journal of Agricultural Economics*, May 1987.

Cameron, T. "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods." *Land Economics*, August 1992a, 302-317.

Cameron, T. "Nonuser Resource Values." *American Journal of Agricultural Economics*, December 1992b, 1133-37.

Cameron, T. A., G. L. Poe, R. G. Ethier, and W. D. Schulze. "Alternative Nonmarket Value-Elicitation Methods: Are Revealed and State Preferences the Same?" Working Paper, UCLA, December, 1999.

Carson, R. T., N. E. Flores, K. E. Martin, and J. L. Wright. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics*, February 1996, 80-99.

Cesario, F. J. "Value of Time in Recreation Benefit Studies." *Land Economics*, February 1976.

Cummings, R. G., D. S. Brookshire, and W. D. Schulze. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.* Totowa, N.J.: Rowman & Allandeld, 1986.

Cummings, R. G., S. Elliott, G. W. Harrison, and J. Murphy. "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy*, June 1997, 609-21

Cummings, R. G., and L. O. Taylor. "Unbiased Value Estimates for Environmental Goods: Cheap Talk Design for Contingent Valuation Method." *The American Economic Review*, June 1999, 648-665.

Desvousges, W. H., F. R. Johnson, R. W. Dunford, S. P. Hudson, K. N. Wilson, and K. J. Boyle. "Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy," in *Contingent Valuation: A Critical Assessment*, edited by J. A. Hausman, New York: North Holland, 1993, 91-164.

Diamond, P. A. and J. A. Hausman. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives*, Fall 1994, 45-64.

Feather, P., and W. D. Shaw. "Estimating the Cost of Leisure Time for Recreation Demand Models." *Journal of Environmental Economics and Management*, July 1999, 49-65.

Haab, T. C., J. C. Huang, and J. C. Whitehead. "Are Hypothetical Referenda Incentive Compatible? A Comment." *Journal of Political Economy*, February 2000, 186-196.

Hanemann, W. M. "Valuing the Environment through Contingent Valuation." *Journal of Economic Perspectives*, Fall 1994, 19-43.

Huang, J., T. Haab, and J. Whitehead. "Willingness to Pay for Quality Improvements: Should Revealed and Stated Preference Data be Combined?" *Journal of Environmental Economics and Management,* November 1997, 240-255.

Kahneman, D., and J. L. Knetsch. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management*, January 1992, pp. 57-70.

Kemp, M. A., and C. Maxwell, "Exploring a Budget Context for Contingent Valuation." in *Contingent Valuation: A Critical Assessment*, edited by J. A. Hausman, New York: North Holland, 217-70.

Kling, C. L. "The Gains from Combining Travel Cost and Contingent Valuation Data to Value NonMarket Goods." *Land Economics*, August 1997, pp. 428-39.

Larson, D. "Testing Consistency of Direct and Indirect Methods for Valuing Nonmarket Goods," University of California, Davis, draft manuscript, December 1990.

Leamer, E. E. "Let's Take the Con out of Econometrics." *The American Economic Review*, March 1983, pp. 31-43.

Loomis, J., A. Gonzalez-Caban, and R. Gregory. "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" *Land Economics*, November 1994, 499-506.

Louviere, J. "Combining Revealed and Stated Preference Data: The Rescaling Revolution," paper presented at the AERE conference, June 1996.

McConnell, K., and I. Strand. "Measuring the Cost of Time in Recreation Demand Analysis: An Application to Sportfishing." *American Journal of Agricultural Economics*, February 1981, 153-156.

McConnell, K., Q. Weninger, and I. Strand. "Joint Estimation of Contingent Valuation and Truncated Recreational Demands," Chapter 7 in *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*, edited by J. A. Herriges and C. L. Kling, Cheltenham, U.K.: Edward Elgar Press, 1999, 199-216.

McFadden, D. "Contingent Valuation and Social Choice" *American Journal of Agricultural Economics*, November 1994, 689-708.

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

Portney, P "The Contingent Valuation Debate: Why Economists Should Care." *Journal of Economic Perspectives*, Fall 1994, 3-17.

Randall, A. "A Difficulty with the Travel Cost Method." *Land Economics*, February 1994, 88-96.

Figure 1. Wetland Zones

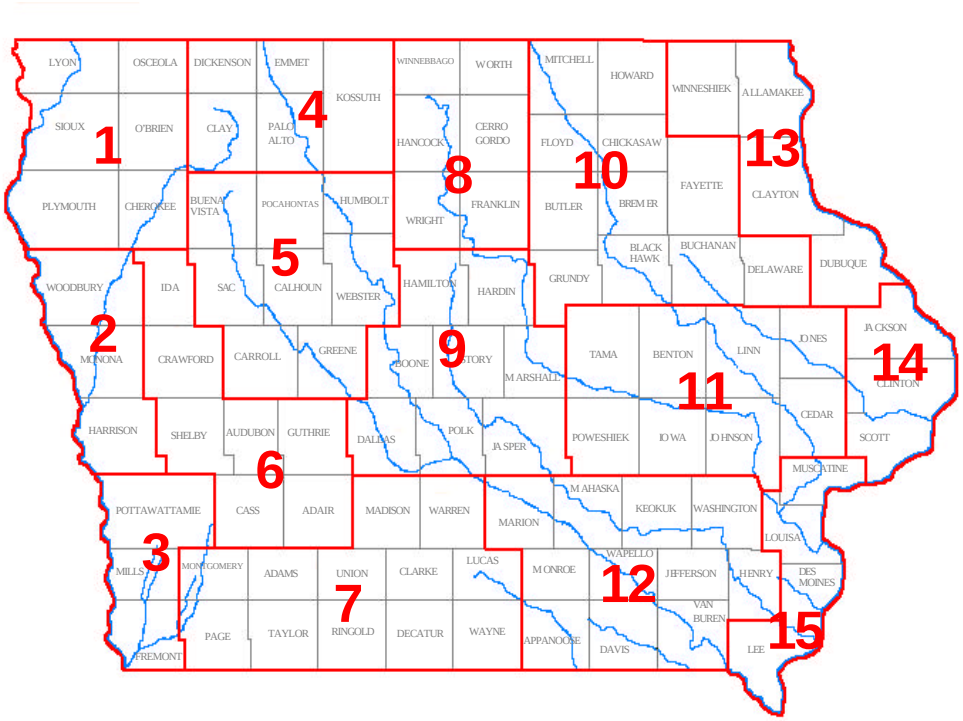


Table I. Parameter Estimates (Standard Errors in Parentheses)

Parameter	Unconstrained	Complete Consistency $\tilde{H}_0^1$	Heteroskedasticity Hypothesis $\tilde{H}_0^2$	Price Hypothesis $\tilde{H}_0^3$	Income Hypothesis $\tilde{H}_0^4$
b_0^R	23.38** (2.91)	14.52** (1.94)	14.22** (2.08)	13.47** (2.40)	13.77** (2.26)
b_p^R	-0.94** (-0.13)	-0.47** (-0.04)	-0.46** (-0.06)	-0.42** (-0.09)	-0.44** (-0.06)
l^R	0.06 (0.06)	0.44** (0.11)	0.47** (0.14)	0.53** (0.19)	0.50** (0.15)
b_y^R	0.08 (0.05)	0.21** (0.05)	0.22** (0.05)	0.22** (0.05)	0.22** (0.05)
s_R	13.84** (0.73)	14.34** (0.78)	14.48** (0.80)	14.47** (0.84)	14.47** (0.85)
k_0	0.48[a] [11.22] (0.14)	1.00	1.00	1.00	1.00
k_p	0.38[a] [-0.36] (0.07)	1.00	1.00	1.04 [-0.44] (0.08)	1.00
k_l	14.69 [0.88] (15.55)	1.00	1.00	1.00	1.00
k_y	3.33 [0.27] (2.18)	1.00	1.00	1.00	0.93 [0.21] (0.12)
k_s	1.04 [14.39] (0.09)	1.00	0.98 [14.21] (0.08)	0.99 [14.40] (0.08)	0.98 [14.24] (0.07)
r	0.72** (0.04)	0.72** (0.04)	0.72** (0.04)	0.73** (0.04)	0.73** (0.04)
-log L	1114.37	1127.56	1127.50	1127.32	1127.35
P-values		<.001	<.001	<.001	<.001
CS_R	99.61	197.39	203.50	222.95	212.25
CS_S	264.65	197.39	203.50	213.52	212.25

** and * denote significance at 99% and 95% confidence levels respectively, a and b denote significant departures from 1.00 at 99% and 95% confidence levels.

Discussion of Session II Papers

by Kelly Brown, U.S. EPA, National Center for Environmental Economics

Introduction

The title of this session is "Validity of Stated Preference Methods." I will begin my comments by providing a brief overview of how I see this session contributing to the overall theme of the workshop. Then I will discuss how each of the papers in this session contributes to this theme, being sensitive to the fact that the presenters did not necessarily choose this particular session and therefore may see their research fitting into the picture in a slightly different manner. My task is to comment on this research from a policy perspective and therefore these comments will focus on the usefulness of the research to environmental policy decision making.

Validity refers to the degree to which a method produces the desired results. For our purposes, validity is the degree to which stated preference methods produce a measure of the economic value associated with an environmental good or service of interest. One of the reasons for devoting an entire session to this topic is because research has suggested that perhaps the current state of the art does not provide a valid measure of economic values associated with non-market goods; stated preference results may not reflect individuals' true willingness to pay for environmental goods or services. This is a serious issue for policy makers attempting to make decisions over devoting resources to clean water or clean air, for example, or trying to convey the total benefits of a particular environmental regulation to the public or fellow government agencies charged with the task of collecting such information.

As we are all well aware, the hypothetical nature of stated preference methods does not necessarily provide us with a result that has been proven to consistently represent true willingness to pay. The question then becomes if, and how, we can use stated preference methods to obtain meaningful measures of economic value or how to modify the methods so as to achieve this result. Both approaches are seen in the research presented in this session. Swallow, et. al. conduct real and hypothetical experiments using different payment mechanisms to determine if the disparity between treatments is affected by the payment mechanism. Herriges and Kling, as well as Stevens, et. al. (presented by John Halstead) compare stated preference results to results from surveys using other valuation techniques to determine if, and how, values differ according to the valuation mechanism. Finally, Mansfield, et. al. analyze the use of public goods as an alternative to monetary compensation for public bads.

Each of these papers brings us closer to making better use of stated preference results, without actually supporting or debunking the method as a whole. I will now turn to a discussion of each of the papers in this session.

Real versus Hypothetical Treatments

Swallow, et. al. examine how validity might be achieved by modifying the payment mechanism used in the survey. To briefly recap, they conduct two sets of experiments. The first experiment is an in-person survey of University of Rhode Island students regarding their values associated with an adopt-a-pond program. Swallow, et. al. use a provision point mechanism with a variety of rebate plans to elicit values. Students are randomly assigned to either a real or hypothetical treatment and within each treatment students are randomly assigned to one of three

rebate plans. The rebate options allow them to assess if, and how, various payment programs affect the disparity between real and hypothetical results. It is this hypothesis that provides the most useful contribution to assessing the validity of stated preference methods.

Unfortunately, despite the very preliminary nature of their results, it does not appear as though the introduction of the various rebate plans "corrects for" the hypothetical bias. This, of course, would be a lofty goal, at best. The degree to which respondents find the payment mechanism realistic is an important aspect of survey design. However, the fact that the results do not appear to move stated preference results closer to a "true" willingness to pay indicates that it is not the payment mechanism that is causing hypothetical and real results to diverge (a familiar result), but that there is some other feature of the method that is at the root of the differences. That is, if Swallow, et. al. were to have found no significant difference between real and hypothetical responses under the provision point mechanism, then there would be evidence that the hypothetical bias might be driven by the payment mechanism, as opposed to say, the hypothetical nature of the survey. Unfortunately, for the stated preference practitioners, this was not the case.

Swallow, et. al. also conduct a second experiment to examine willingness to pay values for a new wetland conservation area in two different Rhode Island towns. Individuals are randomly assigned to one of six treatments distinguished by their real and hypothetical combinations. Again, this experimental design does not introduce any new aspects to the survey methodology. The most useful aspect of this study design is the treatment of 'no' responses in which individuals are asked to further delineate their views in terms of not wanting to pay for the wetland at the specified costs, not valuing wetland protection, and a desire to save their money. Analysis of 'no' responses is merited as researchers are not always sure what a no, or yes, response means. The results of this research are too preliminary to analyze, however, the technique looks promising for helping researchers better understand the variance underlying responses.

Stated Preference versus Other Valuation Methods

Next, I turn to research that attempts to extract meaningful information from stated preference results (as opposed to altering the mechanism itself, as seen above). First, I will discuss Herriges and Kling's research. They address validity by comparing stated preference results to results from a revealed preference survey. The idea behind this approach to assessing validity is to compare real market data with hypothetical market data in hopes of gleaning some information about the hypothetical market. Any discrepancies between the two indicate, perhaps, some flaws, in one, or either method.

Herriges and Kling develop a recreational demand survey that includes a hypothetical question about how behavior would change in response to a change in the price of travel to a state wetland. In other words, they collect real data (how many times did you go to the swamp) and then they have a measure of willingness to pay in the event that a parameter changed (cost). By estimating a joint model they are able to test the consistency between the revealed preference and stated preference data, where the revealed preference data is reflected by questions about previous behavior and the stated preference data is reflected by responses to hypothetical changes in behavior.

Their results reject a hypothesis of consistency between the values revealed by the revealed and stated preference questions. That is, the data do not provide similar information on values.

They also test several other hypotheses related to constrained models and find that these are also rejected. Contrary to the disappointment with this result expressed by Joe Herriges during his presentation, I do not find these results surprising. Revealed preferences provide information on behavior that has taken place in a previous time period, where as stated preferences indicate how an individual anticipates behaving in the future, under different conditions. These are two very different notions and methods trying to compare the two should not necessarily expect to similar results. The researcher does not know if the individual is anticipating a large raise, better health, or merely wants to get out more in the future. Likewise, the researcher does not know if the individual had anticipated participating in more recreational pursuits in the past, but did not do so because of a family emergency or a flat tire. It is not at all unreasonable to observe individuals' thinking about the future in a different light than the past, hence anticipating different behavior than that which has been exhibited.

What does this mean for stated preference methods? Research of this approach addresses the best (or worst) of both worlds. That is, the revealed preference questions "discipline" the stated preference questions by providing a real market context, while at the same time allowing the individual to incorporate other factors into their valuation decision, such as anticipated future changes in behavior or non-use values. The within-sample approach to valuation provides a nice range of values that should not necessarily be rejected because they differ. From a policy perspective I think the revealed preference/stated preference option is a practical alternative that provides a range of values. Researchers may even expect stated preference results to provide an upper bound on values because they incorporate the non-use valuation components.

Stevens, et. al. (presented by John Halstead) also examine the validity of stated preference methods by comparing results from traditional contingent valuation surveys to results from a conjoint analysis. This study employs both a conjoint question and a contingent valuation question regarding willingness-to-accept (and willingness-to-pay in one treatment) for reduced visibility in exchange for lower electric bills. The authors find that contingent valuation (CV) results and conjoint analysis (CA) results yield different results. I focus my comments on the differences as a result of the treatment (as opposed to some of their other hypotheses regarding survey mode and the like). First, they find that CV-CA depending on a somewhat arbitrary measure of indifference. That is, when they impose the restriction that a YES vote is equivalent to situations where people rated B>A they find no difference between the results from the two methods. This clearly needs some more work because such arbitrary distinctions are difficult to adopt on a universal basis. One suggestion I have is as follows. CA provides information on how individuals feel about both the status quo and the change, whereas CV provides information on the change, but not the status quo, assuming that the status quo is acceptable. I propose asking individuals if they are willing to accept the status quo and then use this information to reveal something about their indifference. In other words, consider the following, rough, experimental design.

CV:	status quo	change	
	a. YES	YES	=>1 or 2
	b. YES	NO	=> 3
	c. NO	YES	=> 1
	d. NO	NO	=> 3 or 2
CA:	1. A<B		=> YES
	2. A=B		=> indifferent
	3. A>B		=> NO

In a CV survey there are four possible outcomes. An individual can have positive or negative preferences for the status quo (as indicated by YES and NO under the status quo column above), but only states a positive or negative preference for the proposed change. Likewise, in the CA there are three possible outcomes: B is strictly preferred to A (option 1 under CA, above), A is equivalent to B (option 2), or option A is strictly preferred to B (option 3 above). There is no one-to-one mapping between the CV and CA, even though the analysis attempts to make such a comparison. In the CV questions, researchers only observe a YES or NO to the change. Indifference is typically not allowed. Whereas, in the CA questions, indifference is allowed, however, it is arbitrarily considered a YES or a NO.

By asking individuals in the CV context, about the status quo, researchers are gaining more information about underlying preferences and can map the indifferent individuals (those who answer YES or NO to both the status quo and the change) to those who are indifferent to the CA choices.[1] In addition, I would argue against arbitrarily lumping indifferent individuals in the CA analysis into the YES (or NO) category. This design helps to expand the available information and may provide more insight into the differences, if any, between CV and CA results.

The Stevens, et. al. study also provides some useful information on the choices individuals are given when valuing visibility benefits. The relatively few people who actually agreed to accept the compensation does indicate that actual values may be higher than provided by the change in electric bills and other forms of compensation should be considered. Also, it is clear from the follow-up questions that people have a difficult time untangling visibility from the other issues associated with poor air quality and this should not be ignored.

Alternative Compensation

Finally, turning to the Mansfield, et. al. paper, these authors offer a unique alternative money in compensating for public bads. That is, they introduce the idea of offering a public good as opposed to monetary compensation for public bads to restore utility to individuals. Their results indicate that people are more willing to accept public goods than cash in the presence of public bads and this could serve to be useful in order to gain support for politically unpopular projects, such as siting of landfills. I have a few comments on the survey design itself. The paper does not indicate if individuals were told the total cash award that would be made, as well as their individual award. If given the choice between $1000 and a park, individuals might think that a park is worth more than $1000, not actually accounting for the fact that each individual would receive $1000, or considering whether their own personal use is valued at $1000.

[1] Obviously, individuals who answer YES to both CV questions are expressing very different preferences from those who answer NO to both CV questions. This would need to be addressed in the analysis.

In addition, this method does not provide information on the actual values associated with the good (or bad) in question. It is important to determine the value of the public bad and then determine a reasonable valued alternative to compensate, which is more burdensome than coming up with a single value in the first place. However, it is perhaps more equitable to use public goods as opposed to cash for such compensation because use then is not determined by some ability to pay.

Summary

Where do we stand on the issue of validity? None of these studies provides any overwhelming support or disapproval of stated preference methods. These studies do, however, move the state of the art forward in terms of the ways in which questions are asked to derive values, be it with alternative payment mechanisms or tradeoffs with other public goods. In addition, revealed preference/stated preference combinations will help us move closer to identify ranges of useful values. This research also suggests that it is useful to continue moving the state of the art forward both in terms of tailoring the method itself and expanding alternative ways to conduct stated preference research.

Discussion of Session II Papers

by Anna Alberini, University of Maryland

I. Introduction

The papers presented in the session "Validity of Stated Preference Methods" investigate various aspects of the validity of stated preference data and methods. Specifically, they examine:

(i) Willingness to Pay (WTP) v. Willingness to Accept (WTA) compensation measures (Stevens et al paper, Mansfield et al paper);

(ii) The effects of different survey format, different elicitation approaches (conjoint analysis v. dichotomous choice contingent valuation), populations at different distance from the amenity to be valued (Stevens et al paper);

(iii) Stated preference data v. revealed preference data (Azevedo et al paper);

(iv) Hypothetical v. actual WTP under settings with various degrees of incentive compatibility (Swallow et al paper).

Regarding (i) and (iv), a large body of theoretical literature has been developed that explains the empirical results, and so the ideal output of the research is to discriminate between theories, provide support for or against each theory, and/or identify situations where the existing theories are insufficient to explain the observed behaviors.

For example, we expect to see that WTA for a degradation in environmental quality is greater than WTP to avoid the degradation or obtain a similar improvement. This is the finding of much earlier empirical work, and is consistent with the theory of prospects (Tversky, 1979) and the work by Hanemann (1991) showing that the difference between WTA and WTP depends on the elasticity of substitution between a public good and its private substitutes. In empirical work presented in this session, one would therefore expect to see results compatible with these findings. One would also expect to see that if some of the factors advocated by theory (e.g., elasticity of substitution between the public good and private substitutes) are removed or attenuated (when valuation concerns only private good, or when private substitutes are available), the difference between WTP and WTA should become smaller.

Regarding (ii) and (iii), one would expect to see that, provided that survey instruments are carefully developed and pretested, written, self-administered questionnaires give the same WTP results as computer-assisted in-person surveys. Similarly, one would hope to see that stated preference data and revealed preference (e.g., trips to wetlands, contingent on a certain cost per trip, and actual trips) subsume the same underlying utility functions.

If these are the expectations, is the research presented in this session confirming our expectations? Because most research projects are still underway, it is still too early to say. If the expectations are not borne out in the data, it will be important to understand the reasons why.

The remainder of this discussion is organized as follows. First, I briefly discussed each of the individual papers in sections II, III, IV and V. Finally, I summarized what I have learned from the research projects here presented and offer some lessons for future research.

II. Comments on the paper "Compensating for Public Harms: Why Public Goods Are Preferred to Money" by Carol Mansfield, George Van Houtven and Joel Huber

The purpose of this paper is to determine whether the siting of facilities imposing environmental disamenities can be made more acceptable to a community by offering public goods in lieu of cash compensation.

The paper begins with a nice summary of theories and empirical reasons why willingness to accept to forgo the siting of noxious facilities has been exceedingly high, and why in some studies the rate of acceptance of the proposed siting did not increase—and in fact, decreased—with cash offers. Reasons include the facts that:

cash offers may be perceived as bribes, and as such they carry a stigma;

cash offers crowd out people's willingness to accept the siting of a noxious facility for the good of the community;

accepting a cash offer may be seen as accepting personal responsibility for the negative consequences of an action, while accepting a public good may be seen as distributing such a responsibility over a much larger group of people.

Accepting a public good may be seen as prioritizing other people's well-being more highly than one's own, which is a desirable feature.

All of these explanations, of course, build nicely on the theory of prospects (Tversky, 1979), which suggests that people value differently losses from gains, and on Hanemann's theoretical model which implies that the difference between WTP and WTA depends crucially on the elasticity of substitution between a public good and its private substitutes. They are also compatible with the discussion offered in Mitchell and Carson (1989), who point out that people may report very high WTA values when they are described scenarios that contradict their perception of who has the property right over environmental quality.

In addition, the scenarios constructed by the authors for this research suggest additional reasons why people may hold very high WTA values, or refuse cash payments altogether:

(a) Fear that the agreement whereby cash is paid to individuals may be not be binding for the other party: if the latter does not wish to honor its commitment, an individual has little recourse, but a community who has been denied the public good promised to them will find it easier to put pressure over that party, either informally or through formal channels;

(b) Fear that acceptance of a cash payment opens the door to having to make more concessions to the party seeking to build the noxious facility, an issue potentially related to (c) below;

(c) Uncertainty over what one is really committing to, especially if the sponsors of the survey are not known.

The Experimental design devised by the authors consists of hypothetical scenarios describing noxious facilities being proposed for siting near the respondent's home, in exchange for either (a) cash

compensation or (b) another public good. For each scenario, neither the cash amount nor the public good were varied to the respondents.

In addition, the authors construct a neutral market choice, whereby the respondent must choose between houses located in neighborhood A or B, the houses and neighborhoods being identical in all respects, except for the property tax and the level of a public good. The public bad is the same at both locations, or absent at both locations.

Finding 1. The authors' prior implies that one should expect that the percentage willing to accept public good, among those respondents who were offered the public good as a compensation for accepting siting, is greater than the percentage of respondents willing to accept cash. In practice, depending on the sample and survey format, the former is in some cases greater than and in other cases less than the latter, implying that prior expectations cannot be confirmed or refuted.

It is difficult to draw conclusions from the percentages reported in table X for 3 reasons. First, it is unclear that respondents necessarily understood the scenarios as intended by the researchers.[1] Second, there are many differences in the demographics in the various areas where the study was conducted, and the mall intercept recruiting frame loses the researcher's ability to control for the population. Third, the cash amounts offered as compensation is not varied to the respondents. I believe that more conclusive findings would have been possible, had the cash amounts been varied to the respondents (holding the scenario the same). Doing so would allow the researchers to show if the percentage of respondents willing to accept the dollar amount increases with the dollar amount, or remains virtually unchanged over a wide range of dollar figures.

It is also difficult to say what aspects of a public good prompt the respondents to accept it in exchange for allowing the siting of the facility in their neighborhood. This point and the previous one suggest that perhaps future research might want to conceive this study in terms of choice experiments, where people choose between A, a scenario that offers cash as compensation, B, a scenario that offers a public good as compensation, and a third option allowing them to reject the facility. Respondents would be asked to evaluate several triplets, with varying cash amounts and attributes of the public good. This approach allows the researchers to find out what tradeoffs respondents are making, and would improve the efficiency of the estimates of the coefficients of the underlying utility function. It would allow to compute marginal rates of substitution between different attributes of the public good, and would allow the researcher to estimate the WTA associated with each proposed scenario.

Finding 2. The neutral market choice is in my opinion the most promising part of the experimental design. It avoids the "status quo bias" likely to affect responses in the hypothetical siting scenarios described to the respondents. Future research could attempt to estimate WTA separately from the hypothetical scenario and neutral market choice. Provided that this is done carefully, holding the other aspects of the hypothetical siting scenario and neutral market choice the same (including the payment vehicle, which should be described in terms of savings on property tax in both), any differences in WTA would be ascribed to the formation of a "reference point" and the related "status quo bias."

[1] To elaborate on this point, in the Superfund cleanup scenario people may have been thinking of increased health risks posed by the cheaper barrier method, as opposed to complete cleanup. In the scenario where the neighbors of a farm are offered a recreational facility in exchange for allowing the farm to become a livestock operation, a recreational facility might be an unrealistically expensive option to appease neighbors for the odors associated with the livestock operation. I was also surprised to read about a scenario describing the burning of waste yard in Greensboro—shouldn't waste yard be composted, or taken to landfills, or incinerated at a facility which captures the smoke, rather than just burned creating smoke that bothers residents?

Once again, this goal could be accomplished only if the dollar amounts are varied to the individuals, holding all else the same. Varying the dollar amounts across individuals also helps rule out another possible cause for the key finding reported by the researchers--that the percentage of people that in the neutral market choice choose cash over the public good *is* influenced by the presence or absence of a public bad. The researchers interpret this finding as evidence that the marginal utility of a public good is altered by the presence of a public bad.

To illustrate, the authors ask respondents to choose between identical houses at two locations with different level of public goods and the same level of public bad (either the specified level, or none at all).

The utility from choosing one of the two locations is:

(1) $U = a + b \cdot PG + g \cdot PB + d\,(t - \text{tax saving}) + e$

with α the utility from the house *per se*, which remains identical across the two locations; PG is a dummy for the presence of the specified public good; PB is a dummy for the presence of the specified public bad, t is the total property tax liability, and the tax savings (e.g., $500 a year) is specified in the survey, and applies only to one of the two locations. Here, γ and δ are negative, while β is positive.

When comparing the choice between house A and house B, where both have the public bad, but only the former has the public good, the difference in the underlying utility levels is:

(2) $\Delta U = b + d \cdot \text{tax saving} + e_A - e_B$

Notice that (2) is the same whether or not the public bad is there. Assuming that the error term $(e_A - e_B)$ has the same variance whether or not the public bad is present, the probability of choosing A over B is the same in the situations with and without the public bad, and is equal to:

$$(2)\ \Pr(A) = \Pr(\Delta U \geq 0) = \Pr\left(\frac{e}{s} \geq \frac{b}{s} + \frac{d}{s} \cdot \text{tax saving}\right) = \Pr\left(\frac{e}{s} \leq -\frac{b}{s} - \frac{d}{s} \cdot \text{tax saving}\right),$$

where $e = e_A - e_B$, and σ is the standard deviation of ε.

If the marginal utility of the public good is affected by the public bad (as hypothesized by the authors), one can write:

(3) $U = a + b \cdot PG + b' \cdot PG \cdot PB + g \cdot PB + d\,(t - \text{tax saving}) + e$

in which case the difference in utility levels is $\Delta U = b + d \cdot \text{tax saving} + e_A - e_B$ for the situation with no public bads and $\Delta U = (b + b') + d \cdot \text{tax saving} + e_A - e_B$.

Clearly, the fraction of the sample who chooses one A differs across the treatment with and without public bad if b' is different from zero. But b' could be zero, and the percentages still be different if the variance of $(e_A - e_B)$ is different when the public bad is present. Here, for instance, choosing A (the house with the public good and no tax credit) is more likely in the situation with the

public bad if the variance of the error term, σ, is lower than that for the situation without public good. Is this reasonable to believe? Once again, this could be determined if one were able to estimate separately the β, γ and δ from σ, which can be done only if the tax credit is varied to the individuals.

Finally, this research has attempted to document the existence of some effects, but has not conclusively pinpointed to the reasons why we see them happen. It is likely that they are the result of several concurrent causes, and that it will be extremely difficult to devise an experiment where such causes can be disentangled. The paper needs to acknowledge that.

III. Comments on the paper "Ask a Hypothetical Question, Get a Valuable Answer?" by Christopher Azevedo, Joe Herriges, and Cathy Kling

This paper begins with a discussion where valuation economists are placed into one of three categories: Revealed-Preference (RP) lovers, Stated Preference (SP) lovers and "agnostics" who believe that stated-preference data and revealed-preference data should be used for joint estimation of the parameters influencing utility and of welfare effects, without necessarily considering one type better than the other.

An example is then provided where RP data documenting individual trips to wetland locations in Iowa are combined with the same individuals' statements as to the number of trips that they would take if the price of a trip increased by $X. Joint estimation is carried out that allows for correlation between the SP and RP data.

The authors specify 4 hypotheses explaining differences between stated and revealed preferences: namely, that:

(a) all coefficients are identical;

(b) the variances of the error terms are different;

(c) the coefficient of disposal income is different;

(d) the coefficient of price is different.

The model is first estimated in a completely unrestricted fashion, and then each of the above restrictions is imposed, and rejected. The conclusion I draw from these results is that the statistically significant differences between the parameters of the SP and RP demands for trips offer little hope that SP and RP data can be reconciled, at least in this example.

In general, the paper is well written, but the discussion at the beginning and the end of the paper should be a little more balanced. I believe that most valuation researchers are aware of the limitations and advantages of using stated v. revealed preference data, and do not endorse blindly one approach over the other, although because of their training, background and data availability they sometimes elect to work with only one of these types of data.

For instance, I believe that many of the valuation researchers that have relied primarily on RP data are fully aware of the empirical and theoretical difficulties associated with valuation using RP data: consider for instance multiple-day trips, multiple-site trips, the difficulties in specifying the choice set (when using models of site choice), the participation problem, the definition of the appropriate market, etc.

In addition, studies based on RP data sometimes rely on untested assumptions about individuals' perceive environmental quality or risk and react to it—consider for instance compensating wage studies, where the individual is routinely assumed to know and react to *objective* workplace risks measured at the *industry* or *occupation* levels. RP studies also pose econometric problems—for instance, in compensating wage studies it may be very difficult to disentangle other factors influencing wages from the workplace risk, and this may have huge effects on the estimate of value of a statistical life. Last but not least, actual data may preclude consideration of certain changes in environmental quality at one location. These are very real and valid concerns and limitations, and many researchers known for their seminal work in RP modeling have turned to supplementing RP data with questions about what respondents would do under hypothetical circumstances.

The most important finding of the paper is that RP and SP seem to be driven by completely different set of parameters—the differences are not just limited to specific slope coefficients (such as that of price or disposable income) or the variances of the error terms. This is an interesting result, and not necessarily one that invalidates the use of stated preference data, especially if one recognizes—as nicely pointed out by the previous discussant, Dr. Brown—that in this study we have actual trips taken in the past and trips that *would be taken in the future* under specified conditions. Before we draw firm conclusions from this study, I would like to have more information about the following:

(a) The likelihood function appears to be that of a bivariate tobit model. Were restrictions imposed to reflect the fact that most respondents presumably stated a lower number of trips at the higher price than the number of trips at the current price? In other words, the stated-preference trips are truncated from above at the actual trips, but I believe the model does not impose such truncation.[2]

(b) It is interesting that when one relaxes one of the restrictions of the fully restricted model (in moving from col. (B) to col. (C), (D) and (E)) the final estimates of the coefficients for which the constraint was relaxed are effectively equal to the constrained value (one).

(c) It almost appears like the estimation algorithm used the fully constrained parameter values as the starting point, and failed to move away from that. Did you try different estimation algorithms (e.g., Newton, BFGS, etc)? Did you try different starting points?

(d) If the average number of trips taken by the individual is 8.2 and 2.7 for an increase of $27 in the price of a trip, is a bivariate tobit appropriate? Would the results be robust to a bivariate Poisson or negative binomial model?

(e) Are your results robust to other subsets of your 6000 person survey? Other wetland areas?

(f) Might the results change when the models include other individual characteristics that influence the demand for trip, the variances of the error term, and a person's ability to respond to the hypothetical questions?

---

[2] To illustrate, consider individuals with non-zero actual and stated trips. Their actual trips would be written as $R_i = \mathbf{x}_i b + e_i$, where R is the number of actual trips, **x** is a vector of variables influencing the demand for trips (price per trip, income, other individual characteristics), β is a vector of unknown parameters and ε is an error term. Stated-preference trips would be written as $A_i = \mathbf{z}_i g + h_i$, where the vector z reflects the different price per trip, γ is a vector of coefficients, and η is an error term potentially correlated with ε. If $A_i < R_i$, then $h_i < (\mathbf{z}_i g - \mathbf{x}_i b) + e_i$. This implies specifying a truncated distribution for η.

(g) Is there anything from your other questions that could help us understand why the differences between actual and hypothetical trips? If it is not the measurement of price and income, what else could it be that was not explicitly hypothesized in this study?

IV. Comments on the paper "The Value of Visibility: A Comparison of Stated Preference Methods" by Thomas Stevens, John Halstead et al.

This paper presents the results of a study based on split samples which attempts to answer the following questions:

(a) holding all else the same, do values from on-site stated-preference valuation studies differ from off-site studies?

(b) Holding all else the same, does WTA differ from WTP?

(c) Holding all else the same, do computer-assisted, in-person surveys results in different values than mail surveys based on written questionnaires?

(d) Holding all else the same, does conjoint analysis (CJ) produce the same values as dichotomous-choice contingent valuation (CV)?

Clearly, answering all of these questions results in a very complex study design. As shown in the chart, samples are split over a variety a criteria, and must include sufficient respondents to be able to administer different levels of visibility loss, so that WTP can be tested for scope. In practice, at this stage of the research, the "cells" spanned by the experimental design contain too few respondents for proper comparisons to be meaningful. The problem is compounded by the fact that effectively estimation of WTA or WTP is by dichotomous-choice observations—however desirable these are from a number of respects, their informational content about WTA/WTP is necessarily limited.

(a) *Holding all else the same, do values from on-site stated-preference valuation studies differ from off-site studies?*

The regression model for WTA run by the authors pools all data, controls for survey mode and location, and finds no evidence of a statistically significant difference.

Here, however, the appropriate comparison would be between groups A and C, and between groups B and D. Unfortunately, the paper does not present sufficient details so that conclusions can be drawn from such a comparison. In addition, it is unclear how respondents were recruited to participate in the off-site personal interviews. There is some concern that participants in the off-site and on-site computer-assisted interviews differ from age and income, but the differences are in different directions, depending on whether one looks at the CJ or CV study.

(b) *Holding all else the same, does WTA differ from WTP?*

Table 5 suggests that, if anything, WTP is greater than WTA ($511 v. 331). But these figures are dollars per month, and additional analyses are necessary before they can be deemed as conclusive.

(c) *Holding all else the same, do computer-assisted, in-person surveys results in different values than mail surveys based on written questionnaires?*

The problem with this aspect of the study is that the authors are not simply testing for difference in values resulting from computer-assisted in-person surveys and the use of self-administered surveys based on written materials. What they are actually examining is how laptop-assisted, in-person surveys compare with self-administered-surveys *compounded with the self-selection* typically expected of mail surveys.

The authors should first test for the presence of such effects, at least to the extent that it can be captured in observable characteristics of the respondents, such as their incomes, ages, education levels, etc. Cameron et al (forthcoming) demonstrate how one can check for the presence of self-selection biases using zip-code level averages for income, age etc., which should be available for those persons who were sent the questionnaire, but did not return, and for those who did.

Even if the completing and returning of the mail questionnaires is not associated with any particular characteristic of the recipient, it is important than in comparing WTA across the authors keep in mind that their on-site and off-site samples are different in terms of demographics. The two groups' mean or median WTA may differ just because of the different ages and incomes. For a proper comparison, the authors should fit a regression relating WTA to age, income etc. for each of the two groups, and then statistically compare the coefficients of these variables across groups, or predict WTA using the coefficient from one group, but the individual characteristics typical of the other group.

(d) *Holding all else the same, does conjoint analysis produce the same values as dichotomous-choice contingent valuation?*

Most recent high-quality applications of CJ analysis for valuation purposes have tended to be so-called "choice experiments." In other words, respondents are to choose between A and B, where A and B are environmental amenities described by a set of attributes, including cost to the respondent. Respondents are to choose between A and B, and, in what I consider the best applications, also between A, B, and a "do nothing, pay nothing" option. This allows the researcher to estimate WTA or WTP for each commodity. Another advantage is that the same respondent can be asked to evaluate several pairs of commodities, which increases the number of observations, holding the number of respondents the same (of course, one should control for possible correlation between the responses provided by a respondent).

Here, instead, the researchers ask people to rate A and B, where A is always the status quo and B is a reduced-visibility situation accompanied by a reduction in the electricity bill paid by the respondent. The rating is between 0 (absolutely unacceptable) and 10 (perfectly acceptable).

In practice, however, the authors end up using the ratings only in an ordinal sense: If A rated higher than B, it is taken to imply that the respondent answered his WTA question as a "no."

But is this correct? If A, the status quo, is rated 10 and B, the proposed change, is rated 0, then I would agree with their interpretation that B is rejected at the stated price. But what does it mean when A is rated 4 and B is rated 3? The authors treat these ratings in the same fashion as in the previous answer, although here it would appear that option A is not liked very much at all in the first place, and B is liked less than A, but not very much less. (Were there any respondents that rated the status quo so poorly? Does it even make sense to ask people to rate the status quo?)

At a minimum, the authors should try to experiment with ordinal logit models, or some other models that account for "liking A a lot better than B." To conclude, I do not believe that asking people to *rate* visibility levels (accompanied by reduced electricity bills) is the right way to do the comparison between CJ and CV, especially if the ratings are recoded as dichotomous choice observations, and if interpretation of the ratings themselves could be criticized as arbitrary.

In practice, the regression model run by the authors cannot provide firm conclusions about whether the CJ and CV approach give different results. This is because, depending on how the "ties" are treated, the coefficient of the CJ dummy is negative, large and significant in one specification, and virtually zero (and insignificant) in the other.

The authors also note the estimated median WTA is very different, depending on whether a "tie" (i.e., A and B are rated the same) is treated as an acceptance of the proposed change or whether only when the rating of B is strictly greater than that of A is this response treated as an acceptance. The model should be re-estimated to account for the fact that a "tie" indicates indifference—effectively, this produces a continuous observation on WTA.[3]

The authors should also make sure that median WTA is calculated separately for each of the levels of visibility loss examined in this study.

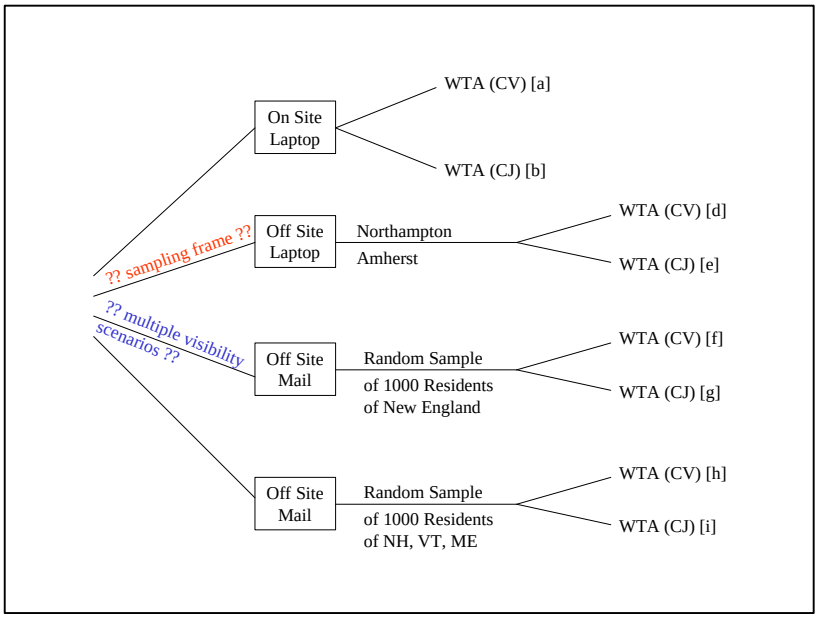


Figure 1. Experimental design in the Stevens et al study.

Finally, the authors state that "That the CVM and conjoint models can produce different results should not be too surprising. Although few comparisons of these techniques have been published, most previous empirical comparisons suggest substantial differences (see Stevens et al., 2000). There are several reasons for this. First, when compared with the CVM, many conjoint questions provide more information

[3] The authors seem to be aware of the problems associated in modeling their ratings. However, they should be careful not to confuse ratings of the alternative scenarios with measures of the *uncertainty* associated with responses to CV questions, as they discuss on page 16 of the paper. What is being measured here is not uncertainty about what one likes, but how much one likes or dislikes a given scenario.

about substitutes. Second, from a psychological viewpoint, respondents may react more differently when choosing among options than they do when making dollar evaluations (Irwin et al, 1993; Brown, 1984)."

It is unclear that this reasoning applies to the present study, since substitutes are not mentioned to the respondent, and we do not know whether respondents were spontaneously considering substitutes when answering either the CV or the CJ questions.

V. Comments on the paper "Toward Comparing Stated Preference and Real Money Contingent Values: Wetlands Valuation" by Stephen Swallow, Michael Spencer and Laurienne Whinstanley

Using two different valuation approaches (standard dichotomous choice and "choice experiments," a variant of conjoint analysis), and two different populations (students recruited at a University campus; state residents) and survey methods (questionnaire taken in the classroom, and mail surveys), this paper examines the relationship between values from hypothetical choices and values from choices involving actual payments. The setting is a donation mechanism ("adopt a pond" for water quality monitoring purposes), and three experimental treatments are devised: one where excess funds are not rebated back to the donors, one where excess funds are rebated back to the donors using a uniform proportional rule, and one where excess funds are retained and used to cover administrative costs of the program (the "extended benefit" mechanism).

Unsurprisingly, WTP is highest for the uniform proportional rebate scheme. Respondents involved in the real money experiment appear to dislike the "extended benefit" mechanism, perhaps because administrative costs are often associated with wasteful use of funds. While willingness to pay for the "adopt a pond" program declines sharply with the real dollar amount requested of the respondent, participants in the "hypothetical" experimental treatment are relatively insensitive to the dollar amount. The authors show that hypothetical WTP exceeds actual WTP by a factor of two or three, depending on the particular experimental treatment.

In general, I remain unconvinced that studies based on the donation mechanism can shed much light on the relationship between actual and hypothetical WTP, because of the lack of incentive compatibility. Perhaps the authors believe that if the donation mechanism encourages free riding behaviors, there is no reason why these should differ across hypothetical and actual donations. In my opinion, there is no particular reason why they should be equal, either.

I also do not find that incorporating various assurances about the use of the excess funds truly helps understand differences between hypothetical and actual WTP. It is important, in my opinion, that experiments be designed to avoid the use of the donation mechanism and to implement the same degree of incentive-compatibility that can be expected of dichotomous-choice, referendum contingent valuation.

VI. What are the lessons that I have personally learned from the papers I have read and seen presented at this conference?

In brief, keep your study design and experimental treatments as "clean" as possible. This means…

1. In designing experimental treatments intended to prove or disprove effects, it is important to keep the design as simple as possible. Unless the available budget and sample sizes are very large, a complicated design may end up with too few observations in each cell to be able to draw firm conclusions.

2. When testing for an effect using split samples, one needs to make sure all of the other aspects of the scenario (payment mechanism, etc.) and of the population are as close as possible across the two subsamples.

3. It is important to avoid drawing samples of convenience from different populations—it is difficult to say what the combined effect of the different population (residents of New England) and the different sampling frame (mall intercept, return of mail questionnaire) will be.

4. Until we know more about them, it is probably better to avoid using elicitation formats that are not clearly interpreted (e.g., ratings) and that can be fed into statistical models of WTP/WTA responses only by making arbitrary assumptions.

References

Cameron, Trudy A., Douglass W. Shaw, and Shannon Ragland (forthcoming), "Non-response Bias in Mail Survey Data: Salience v. Endogenous Survey Complexity," in Joseph Herriges and Catherine Kling, eds., *Valuing the Environment Using Recreation Demand Models*, Edward Elgar Publishing, Ltd.

Hanemann, W. Michael (1991), "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review*, 81(3), 635-647.

Kahneman, D. and Tversky, A. (1979), "Prospect Theory: An Analysis of Decision under Risk" Econometrica, 263-291.

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

Question and Answer Period for Session II

Barbara Kanninen, University of Minnesota, addressed a comment to Stephen Swallow. She suggested that estimating a universal calibration ratio may ultimately be the wrong goal. The Office of Management and Budget (OMB) may have done economists a disservice by giving them the idea that 2 could be the natural calibration ratio. Perhaps economists should search for a technique or calibration model, which would yield different ratios based upon the biases and other aspects of each specific situation.

Swallow agreed in principle. He remarked that his work had come up with a calibration ratio in the wetlands situation of about 6, and between 2 and 3.5 for ponds. His experimental design used real wetlands parcels, presenting the parcels' attributes, such as location and size, to the respondents. However, the experimenters could describe a single parcel in different ways — for example, offering the opportunity to help preserve 30 acres of a 70 acre wetland or offering an opportunity of helping preserve the whole 70 acres. The design allowed the experimenters to come up with real or hypothetical values for these different attributes from the same survey. Others have been working on how to model that difference, to attribute it to education or other factors. That is a worthwhile aim, but beyond the scope of his present work.

Richard Carson, University of California, San Diego, remarked that he saw a serious problem with hypothetical bias. For example, in a survey, when incentives are reduced to hypothetical fund-raising, "free riding" will bias the numbers.

Swallow agreed. His numbers reflect some free riding and other departures. But that doesn't mean stated preference (SP) is wrong, just that it is different. One of the weaknesses in his existing data is that much of the wording in the SP work was parallel to wording in the revealed preference (RP) study. Maybe we should be comparing RP numbers with SP numbers generated in somewhat different ways.

V. Kerry Smith, North Carolina State University, raised two points. First, regarding the Iowa wetlands study, if the price of visiting wetlands increases, people can compensate by visiting closer wetlands. Did the study questions require people to assume their trips would be to the same area, or did the study lump all possible wetlands trips into the same demand function?

Second, in Carol Mansfield's work renters may weigh the possibility of public bads reducing rents while homeowners think about falling property values. That ought to be detectable in comparing the responses of renters and owners.

Joseph Herriges responded to the first point saying his analysis used a simple aggregate model for the prairie pothole region. The next step in analysis is to break it down to three separate zones.

Carol Mansfield said when they regress willingness to accept, the home ownership coefficient is positive and significant. Homeowners were more likely to say yes to both scenarios, an unexpected result. Perhaps it had to do with the nature of the public bads used in the study. The point needs further study.

Anna Alberini observed that some of the results could be explained by variances being different when respondents were offered the public good with the public bad and when respondents heard no mention of the public bad whatsoever.

F. Reed Johnson, Triangle Economic Research, observed to Joe Herriges that the design of the Iowa study necessarily embodied some prior assumptions about preferences. For example, should you weight one real trip as equivalent to ten repetitions of one conjoint? Also, the data in the RP model are going to be riddled with errors and variables. Might these errors and variables swamp underlying similarities between the RP and SP data?

Joe Herriges agreed that what you get from the data depends in part on what you bring to them. Both the approaches of the RP lover and the SP lover will have flaws. The weight you give depends on your prior, and the correct answer is hard to pin down.

Michael Hanemann, University of California, Berkeley, said that based on experience gained in a survey of sport fishing, the number of trips taken could be insensitive to price, but strongly affected by time constraints and other factors. Choice of destination was much more sensitive to price than number of trips was.

This leads to the issue of functional form. He did not believe that demand was linear and in his study used a cubic function to model demand. Also, SP and RP may differ in how they are non-linear.

Further, he pointed out that the number of trips is always an integer, but we are plugging it into a continuous model. If the price goes up, the model does not pick up the integral nature of the variable. But it is possible that SP responses can reflect more subtle changes in demand.

One study where willingness to accept (WTA) gave useful results involved electricity pricing. The trade-off was lower electricity rates if the utility discontinued assistance with energy conservation. In contrast, using WTA in cases dealing with purely public goods raises issues of property rights, asking someone to give up something they do not entirely own.

Finally, he observed that market choice can be worrisome in contexts where the choice offered leaves much unspecified. People want to know details. If you offer them a choice of homes as in the Mansfield study, they will want to know details about the neighborhoods, such as the quality of the schools, which the study does not offer. If you don't tell them, they may make their own assumptions about missing details.

Carol Mansfield replied that they told their respondents that the two choices were identical except for the factors described. However people may take negative factors as signaling other problems with the neighborhood. The issue needs a closer look.

Richard Carson remarked that he had just finished a study of successes in siting hazardous waste facilities. One case involved a pure reduction in property tax, but the others were mostly mixtures of public goods and some type of reduction in tax. It may be good to incorporate these mixed options in future studies. Without money in the picture, you can't answer the question, have you made the public whole?

Kerry Smith observed that EPA needs an inexpensive way to get information about public preferences. It is not necessary to match SP and PR responses from the same individuals. It would be acceptable in principle to stack a small SP sample on, say, the large recreational surveys that EPA does, as long as there were opportunities for restrictions across parameters. Similar sampling is done in other fields.

Julie Hewitt, U.S. Environmental Protection Agency, addressed a question to John Halstead. People have some control over the amount of electricity they use. When people said they were willing to accept higher prices, do you think they were weighing that they could offset part of the rate increase by using less?

John Halstead said the study implied that people would have the same consumption but in hindsight should have stated it explicitly.

Hewitt asked if Halstead could check for variances by income.

Halstead replied that it would be more illuminating to check by state, since New Hampshire has much higher power rates than Vermont or Maine.

STATED PREFERENCE: WHAT DO WE KNOW? WHERE DO WE GO?

PROCEEDINGS
SESSION THREE

APPLICATIONS OF STATED PREFERENCE METHODS TO ECOSYSTEM AND HEALTH ISSUES

A WORKSHOP SPONSORED BY THE US ENVIRONMENTAL PROTECTION AGENCY'S NATIONAL CENTER FOR ENVIRONMENTAL ECONOMICS AND NATIONAL CENTER FOR ENVIRONMENTAL RESEARCH

October 12-13, 2000
Doubletree Hotel, Park Terrace
Washington, DC

Edited by Sylvan Environmental Consultants for the
Environmental Law Institute
1616 P Street NW, Washington, D.C. 20036

ACKNOWLEDGEMENTS

Sections of this report, indicated as "summarizations," were prepared by Sylvan Environmental Consultants for the Environmental Law Institute with funding from the National Center for Environmental Economics. ELI wishes to thank Matthew Clark of EPA's Office of Research and Development and Kelly Brown, Julie Hewitt, Nicole Owens and project officer Alan Carlin of National Center for Environmental Economics.

DISCLAIMER

These proceedings are being distributed in the interest of increasing public understanding and knowledge of the issues discussed at the Workshop and have been prepared independently of the Workshop. Although the proceedings have been funded in part by the United States Environmental Protection Agency under Cooperative Agreement CR-826755-01 to the Environmental Law Institute, it may not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Session III Proceedings
Table of Contents

DO WETLANDS KILL TREES?: KNOWLEDGE AS AN INPUT IN ECOSYSTEM VALUATION

by
John Hoehn
Frank Lupi
Michael Kaplowitz

Michigan State University

Prepared for EPA National Center for Environmental Economics Workshop: *Stated Preference: What do we know? Where do we go?* Washington D.C., October 12, 2000. [ed. note: Draft]

Wetlands ecosystems are valued for a range of ecological services. These services are protected by national, state, and local regulation. The primary federal wetland protection statute is Section 404 of the Clean Water Act (33 U.S.C. §1344). Under this statute, the U.S. Army Corps of Engineers, in conjunction with the U.S. Environmental Protection Agency (EPA), administers a review and permitting process for the "discharge of fill material" in "waters of the United States." Since 1989, the guiding principle of federal wetland policy is the "no net loss" of wetlands criterion (Gaddie and Regens 2000). To implement this principle, the wetland permit process encourages potential dischargers to minimize and avoid wetland impacts wherever possible. Where wetlands are impaired or destroyed, wetland mitigation is required.

Mitigation refers to actions taken to recreate, restore, or protect wetlands of an equivalent type and function to those being impaired or destroyed (Denison and Schmid 1997). Since wetlands vary by type, ecological functions, and the services they yield to humans, the means for judging the equivalency of destroyed and mitigated wetlands is both problematic and central to successful implementation of the "no net loss" policy (National Research Council (U.S.). Committee on Characterization of Wetlands. 1995; Mitsch and Gosselink 1993). Substantial effort has been made to define and measure wetland equivalencies using engineering principles and biophysical characteristics (Bartoldus 1999). However, the economic equivalency of wetland services has received less attention. Absent an understanding of the economic tradeoffs, wetland mitigation may leave economically important services unprotected and under provided.

In this paper, we report initial research results regarding the development and application of a framework for measuring the relative economic values of wetland ecosystems. These initial results stem from the first eight months of a three-year, U.S.E.P.A. funded project. We begin by reviewing the ecological characteristics of wetland ecosystems and past efforts to value wetlands. We then derive a model that leads to three approaches to estimating wetland ecosystem values in stated choice experiments. The relative performance of these valuation approaches depends on the distribution and extent of ecological knowledge among respondents. Knowledge of a particular form is an essential input into accurate ecosystem valuation.

The second part of the paper examines the knowledge base that residents of central Michigan might use in valuing wetland ecosystems. Residents were contacted using random digit dialing and were asked to participate in a group discussion about natural resource issues. Each group involved 6 to 8 residents. Each group interview was conducted by a moderator using a prepared discussion guide.

Discussion participants demonstrated better than expected general knowledge of wetland ecosystems, but their detailed knowledge of wetland functions and services was uneven. Participants recognized habitat for plants and animals as a key wetland function. A smaller portion identified maintenance of water quality and water storage as important wetland functions. Misperceptions were also revealed. For example, several respondents thought that trees do not grow in wetlands and that wetlands kill trees despite the fact that wooded wetlands are common in Michigan. When asked to interpret and discuss photographs of wooded wetlands, these participants said that wetlands were killing the trees.

I. Wetlands Ecosystems and Valuation Research

Wetlands are transitional types of ecosystems that occupy a spectrum between land and water ecosystems. Their exact definition has been controversial (National Research Council (U.S.). Committee on Characterization of Wetlands. 1995). The operational definition used in Federal wetlands permitting regulations builds on two essential wetland characteristics: (i) the land is composed of soils that are water-saturated during part of the vegetation growing season and (ii) the land supports plants that are typical of saturated soils (Smith et al. 1995). Using this definition, wetlands may have covered about 12 percent of the area of the continental United States during colonial times. Since that time, human activity in the United States has converted approximately 45 percent of wetlands area to other uses (Heimlich, Carey, and Brazee 1989).

Wetlands ecosystems vary greatly in type, ecological function, and services to human beings. Wetland types include bottomland swamps, tidal marshes, cattail marshes, vernal ponds, fens, and bogs. Ecological functions of wetlands include water storage, maintenance of surface and groundwater flows, biochemical cycling, retention of water-suspended and dissolved materials, accumulation of peat, maintenance of characteristic biological energy flows, and maintenance of characteristic habitats.

Wetland types and functions provide services that affect human well-being. The water storage function, for instance, may result in service to human beings by retaining floodwaters. Maintenance of groundwater flows may contribute to stable sources of potable water. Wetland habitats may offer recreational opportunities, open space amenities in otherwise densely settled areas, and potential non-use services such as maintaining biodiversity.

The objective of wetland mitigation is to replace wetlands destroyed by permitted activities through the creation, restoration, or protection of equivalent wetlands. The ratio of mitigated wetland area to impaired wetland area is called the mitigation ratio. Mitigation ratios typically vary by wetland type. For instance, in Michigan, recent rules require compensatory mitigation of 1.5 acres for each acre lost when the wetland being lost is a common type. When the destroyed acreage is a rare wetland type, 5 acres of mitigation are required for each acre lost (MCL §324.30319). At the Federal level, the Army Corps of Engineers makes adjustments in the mitigation ratios to account for the type and duration of impacts, the rarity of the impacted wetlands, and the methods used in mitigation (U.S. Army Corps of Engineers-Charleston District 1996).

Wetland mitigation ratios are analogous to the in-kind prices of impaired wetlands. Such ratios represent an agency's in-kind valuation of mitigation activities relative to the lost wetland type or function. A question then arises regarding the adequacy of such prices. For instance, a mitigation ratio that is satisfactory on engineering or biological grounds, may not be acceptable in terms of preventing the loss of economic services and values. For instance, a particular wetland may be ecologically common in a region or state, but rare in terms of its recreational services and open space amenities by virtue of its location in an urban area. Hence, using Michigan's rules to make the point, the statutory mitigation ratio for replacement of a particular cattail marsh might be set at 1.5 to 1 on statewide ecological

grounds, whereas the particular wetland's economic value to its urban area might warrant a rare wetland ratio of 5 to 1.

The economic literature suggests the importance of considering relative economic values in mitigation pricing. Many studies estimate the value of specific wetlands and thereby demonstrate the economic value of wetlands. However, most studies shed little light on the relative value of different wetlands types, functions, and wetland services (Heimlich et al. 1998). A handful of studies do document commercial and recreational values associated with some wetlands (Loomis et al. 2000; Costanza et al. 1998; Bergstrom and Stoll 1993). Other research suggests that wetlands may provide open space amenities (Mahan, Polasky, and Adams 2000; Opaluch 2000). Some recent studies imply that the economic services of wetlands, including recreation, water quality, and flood control services are well recognized by ordinary citizens (Azevedo, Herriges, and Kling 2000). Especially interestingly in terms of mitigation ratios, Mullarkey (1997) estimates that an acre of naturally occurring wetland is 6 times more valuable to respondents than an acre of mitigated wetland.

II. Key Economic Features of Wetland Ecosystems

Wetlands mitigation, to varying degrees in different cases, attempts to account for differences in wetland types, functions, and services. In the context of mitigation, economic values are useful to the extent that they allow for differences across wetland ecosystem types, functions, and services. In an economic sense, a wetland is not a generic, fungible economic commodity. Rather, a wetland is a Lancastrian, multi-attribute bundle that may vary in three major dimensions: type, function, and service. A research design for wetland ecosystem valuation would vary these attributes and assess how value changes with changes in ecosystem type, function, or service.

A second feature of wetland ecosystems that bears on the economics of wetland values is that wetland attributes occur in specific patterns and types. Ecosystems share a general pattern of species relationships. At the foundation of an ecosystem food web are plants that convert energy and nutrients into food. Plant consumers and predator relationships are build upon the vegetative foundation. The specific pattern of species relationship varies with the type and scale of an ecosystem (Miller 1999). That is, a fen does not support the same species and relationships as a bog. Nor does a small wetland of a particular type support the higher order predators that a larger wetland of the same might (Osborn 1996). Since the species mix and interrelationships may vary with type and scale, it is possible that the economic value of wetland types may differ from individually valued sets of wetland functions and services.

A third feature of wetland ecosystems that impacts the economics of wetland values is the uncertainty associated with incomplete knowledge. Knowledge of wetland ecosystems, their functions and services is incomplete on the scientific level (Miller 1999). That is, science may not be able to characterize a full list of relevant wetland attributes nor may science be able to help restore these attributes once there are impaired. In turn, ordinary citizens have incomplete and possibly inconsistent knowledge of the science of wetland ecosystems and functions. Given the evolving nature of science, a useful economic research design for ecosystem valuation might describe how wetland values change with

specific changes in respondents' baseline knowledge of wetland types, functions, and services.

III. A Research Design for Wetland Ecosystem Valuation

The research design outlined below takes an initial step toward a rigorous framework for valuing wetland types and services in stated preference experiments. The goal is a research design that shows the relationship between the value of wetland types and the value of wetland services. As our research program advances, we plan to extend the framework to describe the derived demand for wetland functions. Additionally, we seek a wetland valuation design that makes explicit the role of respondents' knowledge in valuation.

To simplify the exposition, we outline a framework that addresses two wetland types. Wetland acreage of type 1 is represented by *A1*. Wetland acreage of type 2 is represented by *A2*. Each wetland type yields different sets of wetland services. Wetland type 1 yields services of a single kind that we represent with the symbol *S1*. Wetland type 2 yields services of the first kind, *S1*, as well as services of a second kind, *S2*. The total amounts of services available from acreages of type 1 and 2 are:

$$(1) \qquad S1 = A1 + A2$$
$$S2 = K(A2)$$

where K(A2) is an increasing, concave function that maps the acreage of type 2 into a levels of services S2. Equation (1) might correspond to a situation where both wetlands provide open space amenities but only type 2 wetlands support habitat with significant biodiversity.

The next step in the valuation model is to link economic services with human well-being. Human well-being is represented by a utility function, U,

$$(2) \qquad U = U(S1,S2,M),$$

where the level of well-being depends on the levels of the two services and an economic measure of income, M. The link between wetland acreage and well-being comes from the combination of equations (1) and (2). Substituting equations (1) into (2) shows the relationship between economic well-being and wetland acreages,

$$(3) \qquad U = U[A1 + A2, K(A2),M]$$
$$= u[A1,A2,M]$$

where u() is utility function defined on wetland acreage rather than services. This latter utility function leaves the relationship between acreage and services implicit.

In economic terms, a no-net loss policy would leave economic well-being unchanged by compensating for a reduction in type 2 acreage with an increase in type 1 acreage and visa versa. For small changes in acreage, the amount of compensatory mitigation required to offset the loss of type 2 acreage is derived by taking the total differential of the second line of equation (3) with respect to *U*, *A1*, and *A2*. To keep well-being constant, *dU* is set equal

to zero and the differentials rearranged. By this method, the following economic mitigation ratio is derived,

$$(4) \qquad P_{A2\,A1} = u/\ A2 \,/\ u/\ A1$$

$P_{A2\,A1}$ is the utility-theoretic mitigation price of a small reduction in type 2 acreage, measured in terms of a compensating increase in type 1 acreage. In terms of the utility function, this mitigation price is the ratio of the marginal utility of type 2 acreage, $U/\ A2$, and the marginal utility of type 1 acreage, $U/\ A1$.

Each of the marginal utilities in equation (4) is potentially measurable in stated choice experiments. In a choice experiment, respondents would be presented with alternative policy choices involving wetland acreage of type 1 and type 2. The choice data for acreage could then be used to statistically estimate the marginal utilities. Similar experiments could be conducted for choices involving wetland services such as open space and biodiversity. The problem then becomes how to link the estimated marginal utilities of services to the mitigation choices characterized in terms of acreage.

The link between the mitigation price of acreage, $P_{A2\,A1}$, and the mitigation price for services of type 1 and 2, $P_{S2\,S1}$, may be derived by taking the total differential of the first line of equation (3) with respect to *U*, *S1*, and *S2*. Setting *dU* equal to zero leads

$$(4) \qquad \begin{aligned} P_{A2\,A1} &= 1 + K'\ u/\ S2 \,/\ u/\ S1 \\ &= 1 + K'\, P_{S2\,S1} \end{aligned}$$
$$1$$

where *K*' is the marginal productivity of acreage of type 2 in producing services of kind 2, as understood and known by choice experiment respondents.

Several features of the mitigation price as stated in equation (4) are notable. First, we can expect the mitigation price of acreage to be greater than one when the in-kind price of services is positive.

Second, the mitigation price is a function of preferences as represented by the marginal utilities and by the perceived technical relationship between acreage and the second kind of service. This technical relationship is represented by K' in equation (4). The marginal utilities of acreage estimated in stated preference experiments are conditioned on respondents' knowledge of K'. If respondents' knowledge is inconsistent with wetland science, the mitigation prices may be inconsistent with wetland science as well.

Respondents' knowledge plays a central role in accurate estimation of the marginal utilities of acreage. If this knowledge is inconsistent with wetland science, there seem to be two ways to bring the mitigation prices in line with the science. First, it may be possible to bring respondents' knowledge in line with scientific knowledge using educational tools such as carefully worded text, photographs, and diagrams. Whether such informational devices can be effective is an open hypothesis that warrants appropriate tests.

A second way to bring mitigation prices in line with the science is to design stated preference experiments to elicit the mitigation price of services, $P_{S2\,S1}$. The wetland service preference information, $P_{S2\,S1}$, may be combined with a scientific estimate of K' to calculate a facsimile acreage mitigation price based on scientific information,

$$(4) \qquad S_{A2\,A1} = 1 + k' P_{S2\,S1}$$

where k' is the scientific measure of the marginal productivity of type 2 acreage in producing services of the type 2 kind.

The analysis of the economic model of ecosystem values leads to three alterative valuation approaches shown in Table 1. Each approach varies in its information requirements regarding individuals' preferences and the ecological relationship between acreage and services. One approach sets up the choice experiments in terms of acreage tradeoffs for different wetland types. Such an approach mixes preference with ecological knowledge in the structure of the mitigation prices. All else equal, it results in a valid estimate of mitigation prices if respondents' knowledge is adequately complete and consistent with science.

Table 1. Valuation Approaches

Choice Experiment Design		Limitations
1.	Tradeoffs in terms of acreage of different wetland types	Confounds preferences and ecological knowledge; Biased if respondents' knowledge is incomplete or inconsistent
2.	Tradeoffs in terms of final wetland services	Incomplete service list; miss value of whole
3	Tradeoffs by acreage type, but make systematic effort to provide scientific information	Perceptions may not be sensitive to scientific information

The second approach sets up the wetland ecosystem choice experiments in terms of tradeoffs in ecosystem services. Such an approach would compliment the preference information from respondents with information on ecological relationships from science. It would yield a mitigation price based on science that the researcher deems appropriate and acceptable. The science portion of the valuation may also be modified as scientific information changes. A drawback to this approach is that the list of relevant services identified by the research and specified in the model may be incomplete resulting in a partial valuation. In addition, such an approach may not capture the value associated with the pattern of ecological relationships represented by wetland types.

A third approach to wetland ecosystem valuation is based on wetland types. This approach modifies the first approach by attempting to bring respondents' knowledge in line with scientific knowledge. This approach would try to assess respondents' baseline knowledge and to develop information tools that would alter the baseline so that respondent's knowledge was consistent with scientific knowledge. Respondents would engage in choice experiments once they received a systematic exposure to the

information treatment. A key issue for the success of this method is whether respondents are sufficiently sensitive to the new information. If not, the new information may have little effect and the choice experiment results would mirror those of the first approach.

The availability of three different approaches to valuing wetland ecosystems offers the opportunity for cross-corroboration and hypothesis testing. For instance, the second approach based on scientific information might be used to set reasonable upper bounds on the valuation estimates derived from the first approach. Further, the second approach might be used to set up hypotheses regarding the effects of information treatments on the mitigation price.

IV. Knowledge Base of Michigan Respondents

Qualitative research is helping us learn what it is that people value about wetland ecosystems. This step will be used to help the researchers determine the functions and services that should be the focus of the valuation effort. Furthermore, the qualitative research also gives insights into the general state of people's knowledge about wetland ecosystems, their functions, and types (Kaplowitz 2000). We have also been exploring ways of communicating to respondents about wetland functions, "what wetlands do."

To this point, the qualitative research has conducted three group discussions with participants recruited from the general population of adults in the Lansing, Michigan. Each discussion group involved 6 to 8 participants. Participants were initially contacted using random selected telephone numbers. Because of election year resistance to participate in political focus groups, participants were asked to participate in a group discussion of "natural resource issues in Michigan." They were not told that we would be discussing wetlands.

Basic outline of group interviews

Each group interview lasted for roughly two hours. Sessions were held in a facility on the campus of Michigan State University. All of the sessions were conducted by the same moderator who used the same discussion guide for each session. The moderator used non-directive prompts to encourage participants to participate and elaborate their responses. The discussion guide and the sessions had five basic sections, with the first three taking roughly 45 minutes and the last two sections taking roughly 45 minutes. The balance of the time was used for breaking the ice, taking a "snack" break, or completing university paperwork.

The five substantive sections of the discussion guide and sessions were:

1. Introduce participants, identify each participant's top three natural resource issues, and discuss.

2. General background questions about wetlands to explore what participants know about wetlands and to learn about their experiences with wetlands and the things that wetlands do.

3. Photographs of both wetland and non-wetland ecosystems projected on a screen to determine how people judge what is and is not a wetland, to see if

people can distinguish wetland and non-wetland plant communities, and to see if people know about different types of wetlands.

4. Verbal, written, and graphic presentation of different wetland functions including flood control, wildlife habitat, and sediment retention. The functions and definitions for this section were taken from scientific literature on wetlands.

5. Some questions about wetland mitigation and about replacement of impaired wetlands. In the later two focus groups, there were additional questions about replacing wetlands lost due to a highway project were used

Knowledge of wetland functions

Participants evidenced knowledge of wildlife habitat functions of wetlands. The participants also rated the wildlife habitat functions highly in terms of their relative importance vis-à-vis other wetland ecosystem functions. Almost all participants rated wildlife habitat as extremely important, the highest category, on their function ranking worksheets. This finding is consistent with other research on wetlands (Azevedo, Herriges, and Kling 2000; Swallow et al. 1998; Stevens, Benin, and Larson 1995).

Participants had mixed knowledge of some of the other functions of wetland ecosystems such as water quality, groundwater recharge and flood control. Often there were a few respondents in each focus group that were aware of and knowledgeable about one or more of these "non-habitat" functions. However, every group had a majority of participants who seemed much less aware of these types of functions and who did not seem very knowledgeable about them.

Interestingly, several of the scientifically recognized wetland functions prompted negative feedback from participants. Several individuals rejected the importance of functions such as pollution interception and waste treatment. These individuals expressed strong opinions that wetlands should not be used for these functions. In several instances, participants voiced their concern that environmental laws are supposed to provide for pollution cleanup and waste treatment; wetlands need not perform such functions. Note that these functions appear prominently in much of the literature describing wetland functions. After further discussions, most of these participants felt that it would be all right to create new wetlands for purposes such as waste treatment. This feedback seems to illustrate the potential difficulty of relying solely on scientific descriptions of wetland ecosystems, functions, and services.

What do photographs communicate?

As a part of the group sessions, photographs of various wetlands were shown to the participants. This exercise was intended to probe participants' knowledge of wetland types, wetland vegetation, and general understanding of wetland ecosystems. The participants' discussions of the images yielded some interesting insights about what photographs can communicate to people. For example, at one point we showed a photograph of a fen (a particular wetland type) that did not have visible water and had grasses and vegetation that

was browning. In response to this image, some respondents noted that it did not look healthy and that it was not supposed to be that way. One participant said the photograph showed an area that "I would say [was] scorched by fire." In reality, the photograph contained a moderate amount of shadow that was mistaken as evidence of fire. This photograph clearly communicated something other than what had been intended, and the cue that caused the misperception, the shadows, is unlikely to be absent in future photographs of fens and other ecosystems.

Another example of the power of photographs to (mis)communicate was found when the blurry background in a photograph of a non-wetland meadow was "seen" by a respondent to be water. It is important to note that the focus group participants were viewing these images on a large projection screen at levels of resolution that are likely quite higher than what would be feasible in a typical survey application. The conclusion that can be drawn from these experiences is that photographs do communicate information, both intended and unintended, and that they must be pre-tested along with other potential survey elements. This will hold for web-based surveys as well as other mediums.

Wetland misperceptions: wetlands kill trees and trees don't grow there

As a part of the group interviews, participants were shown a variety of photographs that depicted different wetland types in different settings as well as photographs that did not show wetlands. Part of the group interview probed for whether or not each of the photographs depicted a wetland. In each of the groups, several respondents commented on the notion that trees do not grow in wetlands and that wetlands kill trees. In fact, some participants used their perceived presence of dead trees in the photographs to distinguish wetlands from non-wetlands. Therein lies the source of the paper title. The so-called "dead tree" comments occurred in all three of the sessions and they occurred in relation to different photographs of forested wetland areas. It is interesting to point out that in Michigan where the participants live over two-thirds of the wetlands are forested. Another factor that may have played a role in this perception was that one of the wetland photographs showed some prominent trees that had been attacked by Dutch Elm disease. However, two of the sessions raised comments about wetlands and dead trees in conjunction with photographs of forested wetlands shown before the image of the wetland with the diseased trees. Thus, the photograph with the dead elms did not cause the perception, though it may have amplified the perception for some individuals. One conclusion that emerges from these examples is that it seems vital to the design of an accurate valuation instrument that researchers be aware of respondents' perceptions (and mice-perceptions) about the good being valued. Establishing such information is a key step in the development of methods of communicating with respondents about the good to be valued and the context of the valuation.

Knowledge of mitigation

In all three of the group sessions, some questions were asked to about wetland mitigation and about the replacement of impaired wetlands. These questions were aimed at revealing peoples' understanding and acceptance of wetland mitigation. In the later two sessions, additional questions were asked in the context of a scenario in which the government would be replacing wetlands impaired by a highway project. This scenario was

developed to force people to consider, to add realism, and reinforce the idea of trade-offs. The scenario was also used to learn more about one possible context for stated preference wetland valuation. The comments and discussion surrounding these portions of the group sessions revealed a general skepticism that wetland mitigation could adequately replace what might be lost due to a wetland impairment. This skepticism is related to the unique challenges posed by ecosystems as well as the role of knowledge as an input into ecosystem valuation.

Another finding from this section of the group interviews was that there was some confusion over the meaning of wetland mitigation, especially wetland replacement. Some individuals took the concept quite literally and inferred that it would mean transferring plants and animals from one site to the mitigation site. For example, one participant asked, "How are they going to transfer all those frogs?" Again, this serves as another example of how indispensable to survey design it is for researchers to have a grasp of respondents' baseline knowledge and understanding.

Perhaps the main finding from what was learned about peoples' knowledge of mitigation relates to the general skepticism about replacing all functions of a specific wetland. The following are examples of the kinds of comments we received in discussions on wetland mitigation:

"I don't know if you can come out equal."
"Really replacing or just duplicating parts you see?"
"Like substituting oleo for butter."
"Could they truly get back all that was lost?"

It appears that such skepticism consists of two elements. The first related to a disbelief that certain functions, or services, of wetlands could actually be replaced. The second related to a feeling by several individuals that wetland replacement would not adequately compensate for impairments because wetlands are complex. That is people acknowledge that even though many functions might be replaced, there is more to the wetland than the specific functions that get replaced. Both elements of peoples' skepticism raise issues that are fundamental to ecosystem valuation. The former element raises questions about whether we want to elicit people's beliefs in the underlying production relationship, K(.), at the same time we elicit economic choices and values. As illustrated above in the table, this can lead to a co-mingling of values and knowledge about how final services are derived from the "replacement" wetland ecosystem. The second element speaks to the notion that an ecosystem is more than a bundle of listed functions or services.

V. Conclusions

The valuation framework outlined above identifies three approaches to valuing wetland ecosystems and wetlands mitigation. The three approaches show that the economic value of wetlands is derived from the value of wetland services; wetlands are valued when they yield valuable services. This linkage between wetlands and wetland services has an important implication for stated choice experiments. If respondents' knowledge is inconsistent with wetland science, stated choice experiments may yield incomplete or inaccurate valuations.

Knowledge of the linkage between wetlands and wetland services plays a slightly different role in each of the three valuation approaches derived above. The first valuation approach takes respondents' knowledge as given. It elicits a valuation conditioned on respondents' baseline knowledge. The second approach elicits a valuation of wetland services and then uses scientific knowledge to compute a wetland valuation from the estimated value of services. The third approach attempts to bring respondents' knowledge in line with scientific knowledge using systematic information treatments. It elicits wetlands values conditioned on respondents' updated knowledge base.

The reported qualitative research was intended to explore the knowledge base of likely respondents in order to assess the feasibility of the three valuation approaches. Initial findings show that Michigan residents are more cognizant of wetlands than expected, but that their knowledge is uneven. Most respondents had some prior knowledge of wetlands functions such as provision of wildlife habitat, maintenance of groundwater flows, and floodwater retention. However, some functions identified by wetland science, such as retention of polluted run-off and waste treatment, were rejected as illegitimate by some respondents. A portion of these respondents thought that pollution retention would harm the ability of a wetland to support wildlife and other functions. Others thought that current environmental laws should lead to cleanup of pollution at the source, rather than letting pollution flow into a wetland.

The qualitative research also underscored the difficulties of using photographs to communicate wetland knowledge. The initial hypothesis was that photographs might be an effective means of communicating differences in wetlands types and functions. Photographs, however, seemed to be an inaccurate communication device. When shown a photograph of a fen, some respondents correctly interpreted dark areas as shadows, while others interpreted the same dark areas as evidence of impairment and, perhaps, fire. When shown photographs of wooded wetlands, some respondents concluded that the wetlands were killing the trees, even though healthy wooded wetlands are a common wetland type in Michigan.

The evidence thus far underscores the role of knowledge as an input in valuing wetland ecosystems. The empirical results show that respondents have some baseline knowledge of wetlands, but that this baseline knowledge may be incomplete or inaccurate in certain dimensions. In this context, each of the three valuations approaches may be useful in posing and testing hypotheses about wetlands values and the effect of knowledge. For instance, if respondents' baseline knowledge is incomplete, values estimated via the second approach may be larger than values estimated via the first approach. Thus, the three valuation approaches may offer the means of testing and corroborating wetland values.

References

Azevedo, Christopher, Joseph A. Herriges, and Catherine L. Kling. 2000. Iowa wetlands: Perceptions and values: Center for Agricultural and Rural Development.

Bartoldus, Candy C. 1999. *A comprehensive review of wetland assessment procedures: A guide for wetland practitioners*. St. Michaels, MD: Environmental Concern Inc.

Bergstrom, John C., and John R. Stoll. 1993. Value estimator models for wetlands-based recreational use values. *Land Economics* 69 (2):132-137.

Costanza, Robert, Ralph d'Arge, Rudolf de Groot, Stephen Farber, Monica Grasso, Bruce Hannon, Karin Limburg, Shahid Naeem, Robert V. O'Neill, Jose Paruelo, Robert G. Raskin, Paul Sutton, and Marjan van den Belt. 1998. The value of the world's ecosystem services and natural capital. *Ecological Economics* 25:3-15.

Denison, Mark S., and James A. Schmid. 1997. *Wetland mitigation: Mitigation banking and other strategies for development and compliance*. Rockville, MD: Government Institutes Inc.

Gaddie, Ronald K., and James L. Regens. 2000. *Regulating wetlands protection*. Edited by M. E. Kraft and H. Wilson, *Suny series in environmental politics and policy*. Albany: State University of New York Press.

Heimlich, R.E., M.B. Carey, and R.J. Brazee. 1989. Beyond swampbuster: A permanent wetland reserve. *Journal of Water Conservation* 44 (5):445-450.

Heimlich, Ralph E., Keith D. Wiebe, Roger Claassen, Dwight Gadsby, and Robert M. House. 1998. Wetlands and agriculture: Private interests and public benefits. Washington, D.C.: Resource Economics Divisions, Economic Research Service, U.S. Department of Agriculture.

Kaplowitz, Michael D. 2000. Identifying ecosystem services using multiple methods: Lessons from the mangrove wetlands of Yucatan, Mexico. Agriculture and Human Values 17 (2):169-179.

Loomis, John B., Paula Kent, Liz Strange, Kurt Fausch, and Alan Covich. 2000. Measuring the total economic value of restoring ecosystem services in an impaired river basin: Results from a contingent valuation survey,. *Ecological Economics* 33 (1):103-117.

Mahan, Brent L., Stephen Polasky, and Richard Adams. 2000. Valuing urban wetlands: A property price approach. *Land Economics* 76 (1):100-113.

Miller, G. Tyler. 1999. *Environmental science: Working with the earth*. Belmont, CA: Wadsworth Publishing Company.

Mitsch, William J., and James G. Gosselink. 1993. *Wetlands*. 2nd ed. New York: Van Nostrand Reinhold.

Mullarkey, Daniel J. 1997. Contingent valuation of wetlands: Testing sensitivity to scope. Dissertation, Agricultural and Applied Economics, University of Wisconsin - Madison, Madison, WI.

National Research Council (U.S.). Committee on Characterization of Wetlands. 1995. *Wetlands: Characteristics and boundaries*. Washington, D.C.: National Academy Press.

Opaluch, James. 2000. Strength of preference indicators. Paper read at EPA Environmental Policy and Economics Workshop Series, October 2000, at Washington, D.C.

Osborn, Ronald G.. 1996. Ecological criteria for evaluating wetlands. Fort Collins, CO: National Biological Service.

Smith, Daniel, Alan Ammann, Candy Bartoldus, and Mark Brinson. 1995. An approach for assessing wetland functions using hydrogeomorphic classification, reference wetlands and functional indices. Washington, D.C: U.S. Army Corps of Engineers.

Stevens, Thomas H., Samuel Benin, and Joseph S. Larson. 1995. Public attitudes and economic values for wetland preservation in new england. *Wetlands* 15:226-231.

Swallow, S., M. Spencer, C. Miller, P. Paton, R. Deegen, L. Whinstanley, and J. Shogren. 1998. Methods and applications for ecosystem valuation: A collage. In *Proceedings of the first workshop in the environmental policy and economics workshop series*, edited by B. Kanninen. Washington, D.C.: Office of Research and Development and Office of Policy, U.S. Environmental Protection Agency.

U.S. Army Corps of Engineers-Charleston District. 2000. *Compensatory mitigation*. 1996 [cited October 27 2000]. Available from http://www.sac.usace.army.mil/permits/sop96-01.pdf.

STRENGTH OF PREFERENCE INDICATORS AND DISCRETE CHOICE MODELS

Presented by James J. Opaluch, Environmental & Resource Economics, University of Rhode Island
Co-authored with Steven Swallow and Thomas Weaver, Environmental & Resource Economics, University of Rhode Island

Summarization

Professor Opaluch presented a study of using strength of preference indicators along with discrete dichotomous choices to improve the accuracy of estimation of environmental values.

Because of limited funding, the study used data sets gathered for other projects.

Though theory assumes that preferences are ordinal, people are usually able to state the strength of their preferences. Many economists are skeptical of this information, doubting its consistency and equating its use with using cardinal utility scales and interpersonal comparisons of utility.

The study sought to determine whether strength of preference indicators contain useful information. In the context of practical policy analysis with real people, could these fuzzy strength of preference indicators improve measurements of preference?

The policy context was the siting of a new landfill in Rhode Island. The planners wanted public input for the siting process. Using focus groups, researchers had developed a survey to score potential sites.

The survey offered respondents pairs of choices. For example, in one pair, choice *A* sited the landfill in an area with high quality groundwater and normal wildlife habitat, occupying 100 acres of marsh, 300 acres of woods, and 100 acres of farmland, and costing each taxpayer $360 per year. Choice *B* was identical except the site included 400 acres of woods and no farmland. Researchers presented respondents with diagrams and descriptions of the choices and asked respondents to indicate their preference by pasting a large orange sticker labeled “Landfill” on one of the two diagrams. (The use of a sticker avoided confusion over whether respondents were marking the site they preferred as a landscape or the site they preferred to be used for the landfill.) Then, researchers asked respondents to indicate the strength of their preference on a five point scale, with one end of the scale labeled “Weakly prefer Site” and the other end labeled “Strongly prefer Site.”

In focus groups, the researchers found that the respondents needed significant background information about the need for landfills. Without that information, respondents were likely to tacitly reject all the options and not truly reveal the differences in preference that the survey tried to measure. The researchers presented this information

in a professionally produced six-minute video, which proved much more effective than written or even live oral presentations.

The goal of the research was to see if the information in the answers to the strength of preference questions could be used with the discrete choice information to get better estimates of public preference than one could get with the discrete choice information alone.

The study had 12,000 observations in the full data set from the original survey. The researchers calculated the preference coefficients for a standard discrete choice model using just the discrete choice data (the full sample binary logit approach) and using the discrete choices with the strength of preference data (the full sample ordered logit approach). The researchers assumed that these two sets of results represented the true preferences of respondents.

Then the researchers took small random samples of those 12,000 observations and tested whether using the strength of preference data with the discrete choice data (the small sample ordered logit approach) could yield a better prediction of the true preference than using the discrete choice data alone (the small sample binary logit approach).

In using the strength of preference data, the researchers modeled the strength of preference indicators as quasi-cardinal measures. They took strength of preference statements to be comparable across respondents, but they did not assign equal magnitudes to the utility differences between alternatives.

For each choice pair, there were ten possible selections: five rankings (from strongly prefer to weakly prefer) for choice *A* and five similar rankings for choice *B*. Rather than assign cardinal values of rank to the selections, the researchers calculated ordered logit coefficients describing the probability that respondents would pick one choice over an adjacent choice. In effect, these coefficients represent cut-offs in utility between selections. The researchers found that the differences in coefficients between adjacent selections got larger towards the endpoints, perhaps because people psychologically tended to save the more extreme rankings for use on possible future choices. For any given pair of options, the increases in coefficients from weakly prefer to strongly prefer were roughly symmetrically.

The small random samples that the researchers took from the full sample ranged from 272 observations (a little more than 2%) to 2176 observations (nearly 20%). They took 100 random samples for each of eight sample sizes.

They estimated the preference coefficients for each sample using binary logit and ordered logit approaches, and calculated the difference from the "true" values calculated from the full sample. As a measure of how good the small sample estimate was, they calculated the mean of the squared difference between the estimated and "true" values.

When the "true" values were the ones calculated from the full sample using the ordered logit approach, the ordered logit approach on the small samples consistently beat the binary logit approach, by factors ranging roughly from three to four (comparing the

sizes of the mean squared errors). The improved performance was greater on the smaller samples.

When the “true” values were the ones calculated from the full sample using the binary logit approach, the ordered approach still beat the binary approach for samples of fewer than 1360 observations. It converged to being roughly equal in accuracy on samples of 1360 or more observations.

Prof. Opaluch concluded that it can be worthwhile asking strength of preference questions, particularly when using small samples. Common language indicators do seem to contain real information. People seem to have some shared understanding about the meaning of these quasi-cardinal strength-of -preference rankings. He noted gains of from 100% to 400% in sampling efficiency, depending on sample size and what one considered to be the true values.

Discussion of Session IIIa Papers

by Daniel Hellerstein, USDA Economic Research Service

Strangely enough, or perhaps Julie and Nicole have access to my hard drive, I have recent experience with the subject matters of both these papers. Besides, as is typical of most ivory tower academics, the gist of these works is understanding concepts and tools, rather then analysis of a particular policy. I take this as a cue, that I allow my comments to range beyond those of a dedicated policy discussant.

Let's start with a big thought. Is there a "concept" that lurks in the background of both papers? Perhaps yes: let's call it "accounting for respondent uncertainty". For Swallow and Opaluch, it's accounting for respondents' strength of preferences; for Hoehn and Lupi it is accounting for how the respondent thinks about proposed alternatives.

Enough philosophy. Let's consider each paper.

Swallow, Opaluch, and Weaver Paper

The Swallow, Opaluch, and Weaver (SOW) paper is part of a continuing groundswell of work that uses ordered models, such as the ordered probit and ordered logit; these include works by:

Alberini, Boyle, Welsh: Maine fishing
Haefele and Loomis: Forest/pest management
Hellerstein et al: Grassland birds preferences for organic agriculture in Germany

Paralleling this work is an interest in allowing respondents to express uncertainty about their answers, as exemplified by papers by Champ and by Poe and Welsh.

So what's the deal? I think there are two somewhat competing notions that underlie both of the above:

a) Let people express uncertainty: the task of placing a dollar value on possibly obscure changes in environmental quality isn't easy; expecting people to be sure of their decisions is just not realistic. In such a world, forcing a yes/no decision is traumatic and is likely to lead to mistakes: perhaps people will yea say, perhaps people will back off from a hasty commitment to something that (most of the time) they do prefer.

b) Make 'em take it seriously: since it's too easy for respondents to act nicely in a hypothetical setting, strength of preference measures are useful as reality checks on people's intentions — they act as a proxy for making them pay up.

Dealing with the latter story is pretty clean cut — one chooses a cut-off that really means "yes I prefer A to B". For simple CV questions, this may mean treating anything less then a "DEFINITELY YES" as a "NO".

In the contingent-response world examined by SOW, this is complicated by the variety of attributes that may change. However, a null hypothesis that "only dollars matter" suggests that only a "strongly prefer" on a more expensive alternative would be treated as a choice of this (more expensive) alternative.

Yet the "make 'em take it seriously" notion seems inefficient, with real information arbitrarily collapsed. Moreover, it's too conservative — it doubts that a probably yes is anything but a polite no. Is that what you mean by a probably yes? Herein lies the appeal of the ordered estimators (such as the ordered probit and ordered logit) — they offer a systematic means of dealing with real uncertainty.

Actually, in the contingent-response world, it can be argued that (for example) a strongly prefer versus a weakly prefer measures the magnitude of a preference, rather then respondent uncertainty. That is, a respondent knows what she likes better, but wants to be able to qualify her statement (by saying, "I only like it a little bit better" versus "I like it a lot better").

Perhaps, but the same interpretation can be given to a "definitely yes" versus a "probably yes" response to a dichotomous choice question. That is, in either the dichotomous choice or the contingent response framework, either of these two notions may be operative.

I like the idea of ordered models, but (as pointed out by SOW) much of the profession is skeptical; a skepticism based on the subjectivity inherent in these rankings. For example, holding tastes constant (including the random component of taste that the ordered models seek to control), a "decisive" individual may say "DEFINITELY YES" (or "STRONGLY PREFER"), whereas a more cautious fellow may say "UNSURE" (or "SLIGHTLY PREFER"). Lacking a way of classifying individuals into such categories, the concern is that the ordered estimator will be ill defined and subject to bias.

SOW deal with this problem explicitly. First, in their appendix they show that "thresholds" that are randomly distributed around a (threshold specific) mean can yield a familiar ordered probit model. I'm a bit concerned that this framework will yield greater uncertainty about the underlying error variance (hence a larger confidence interval for WTP), but I suspect that this is both unavoidable and of minor significance.

More importantly, they use split sample designs to see what does a better job — the test being what estimator yields the best answers, where the "truth" is the WTP derived from a larger sample. This truth also depends on what estimator is used for the large sample.

One would expect that if the "truth" is a simple world where strength of preferences are merely a conversational ploy by respondents, then a simple model (the standard logit) would be best. Conversely, if strength of preferences is really related to underlying utility, then the ordered model should be a winner.

Somewhat surprisingly, the ordered model (ordered logit in their case) comes out as a winner — it does a better job in both worlds! This suggests that adding extra information

doesn't hurt, and can help. These results do give some breathing room to practitioners wishing to use ordered models.

Unfortunately, from a policy perspective this does not let us off the hook of choosing between the "let people express uncertainty" versus the "make them take the task seriously" interpretation of strength of preference. Basically, several of us (myself, and Alberini) have found that WTP numbers computed from ordered models are similar to WTP numbers that arise from using "unsure" as a cut-off in binary choice models. Use of these UNSURE models inflates WTP values (often by a factor of 2 or 3) in comparison to using DEFINITELY YES. Although SOW use a somewhat different framework (how much they "prefer" rather then how "certain" they are), the concern still holds.

I don't know the answer to this one. I suspect it's tied into the public goods nature of environmental goods. On one hand, there is evidence (such as Champ's work and the experience of market research) that only the most definitive people will actually pony up the money when offered the hypothetical good. On the other, for public goods (such as the existence values often the subject of stated preference work), the concern for free riding is likely to cause people to hesitate on actually paying a stated amount, even if this amount is the personal value of the proposed level of the environmental good. In this world, ordered models may be doing a good job of capturing these concerns.

Hoehn and Lupi Paper

Rather then worry about how certain people are about their responses, Hoehn and Lupi (HL) are most concerned about how certain people are about the proposed alternatives.

But first, allow me to digress and consider the problem of the value of wetlands. Recently, we undertook a literature review to determine wetland values by wetland type and region of the country. We examined several broad studies, including:

Heimlich: 33 studies
Woodward and Wui: 35 studies (meta analysis, and graphical analysis)
Brouwer et al: 30 studies (meta analysis)
Bardecki: review of 277 papers

The general conclusion from these four reviews is that the prospects for benefits transfer are not strong. Some cautious statements can be made about what functions are valuable (such as flood protection versus maintenance of biodiversity), and there seems to be some indication that valuation methodology is not highly significant. However, the general finding is that the range of values is extreme, with a coefficient of variation well above 1.0.

Despite these general findings, we attempted to come up with some number for the varying functions provided by different types of wetlands in different parts of the nation.

The table will give some sense of the paucity of information: with filled-in squares representing cells for which some study has some kind of value.

Thumbnail Sketch: Availability of wetland benefit measures

(Shaded squares = study of value available)

Key

Regions	
I	N. Crescent
II	E. Uplands
III	S. Seaboard
IV	Heartland
V	Miss. Portal
VI	Prairie Gateway
VII	N. Great Plains
VIII	Basin & Range
IX	Fruitful Rim

Services	
A	Wildlife habitat
B	Recreation
C	Flood protection
D	Storm buffers
E	Water quality
F	Com. fish/shell
G	Timber etc.
H	Aesthetic/OS
J	Non-use

Wetland types	
Est	Esturine
PF	Palustrine forested
PE	Palustrine emergent
PS	Palustrine shrub

I	Est	PF	PE	PS
A		■	■	■
B	■	■		
C		■		■
D				
E		■	■	■
F	■			
G				
H		■		■
J		■	■	■

II	Est	PF	PE	PS
A				
B		■		
C				
D				
E				
F				
G				
H				
J		■		

III	Est	PF	PE	PS
A				
B				
C				
D				
E				
F	■			
G				
H				
J		■		

IV	Est	PF	PE	PS
A				
B		■	■	■
C				
D				
E				
F				
G				
H				
J				

V	Est	PF	PE	PS
A				
B	■			
C	■			
D	■			
E	■			
F	■			
G				
H				
J				

VI	Est	PF	PE	PS
A				
B				
C				
D				
E				
F				
G				
H				
J				

VII	Est	PF	PE	PS
A			■	
B		■	■	■
C				
D				
E				
F				
G				
H				
J				

VIII	Est	PF	PE	PS
A				
B				
C				
D				
E				
F				
G				
H				
J				

IX	Est	PF	PE	PS
A		■	■	■
B	■			
C		■		
D				
E	■	■	■	■
F	■			
G				
H				
J				

The point of this digression is that the we don't have a very good inventory of values of wetland service flows and hence welcome contributions that may broaden our knowledge base, especially if this contribution allows us to readily do benefit transfer.

So what do HL bring to this problem? Right now, they don't have a lot of tested hypotheses, but the notion they are exploring is of great interest: that how people think about environmental values may differ from what us inside-the-beltway analysts would find convenient. And they do this for both "wrong" and "right" reasons.

The wrong reason is that they don't believe the science -- they refuse to accept our policy scenarios as possible.

The right reason is that their real concerns are not being addressed -- that what we ask them to value is not what they really care about, or not all that they care about.

The wrong reason is actually sort of interesting. One could argue that people value what they perceive, not necessarily what is really out there (the sizzle is more important than the steak). Perhaps, but let's assume that the long run isn't that long, so that the populace will thank us (eventually) for using the correct facts when they are known. That still leaves the second problem -- what do people really care about? And that's a much tougher nut to crack.

I digress and relate our experience with our survey that valued grassland birds. We were interested in the value of more birds and tried our best to conjure a scenario where only grassland bird numbers changed. Despite our efforts, our screener questions revealed that of the respondents who would be willing to pay at least $1, over half did it because "of the overall effects that supporting bird populations would have on the environment", whereas about 20% did it "just for the birds".

Are these people wrong? Or are they properly skeptical, seeing through our tricks to the truth that one can't change birds without effecting a lot of other things?

As a policy matter, this has some profound effects. Let's call it the dilemma of hedonics -- that even if we had a vector of prices for environmental goods and services, it wouldn't allow us to do good benefits transfer.

For example, let's consider the non-market value of conservation programs. If there was a finite set of environmental goods and services that people care about, then all we need to do is find unit values of these goods and services. Then, come the next ambitious conservation program ginned up by the farmers, enviros, or whatever; we'd be able to do a simple vector product to yield the program's value (assuming the science was there to tell us the size of the change in environmental goods and services).

HL suggest that it won't be that simple -- that the "sum is greater then the parts", hence adding up the (appropriately weighted) parts isn't enough. Interestingly, this is in contrast to John's earlier work, wherein WTP(A) + WTP(B) > WTP(A+B)!

So what do we do? Perhaps there are arcane, yet important, parts that need to be included in the bundle of more obvious goods and services. Yet, being "arcane", these won't be easy to measure or to communicate to respondents — Alternatively, a collection of indices (say, species diversity indices) may capture much of what people care about and also be sufficiently related to physical changes to allow benefits transfer. Or, perhaps a collection of wetland "types" can be identified — sort of a multinomial approach to valuation, as opposed to an hedonic approach (an interesting possibility, but one that may sacrifice our ability to measure the value of small changes).

Let me end on a last philosophical point — that the information given to, and by, respondents can matter — that the way we interpret this information can have real impacts on measured values of the size, and scope, of the benefits due to environmental improvements.

EVALUATING CONTINGENT VALUATION OF ENVIRONMENTAL HEALTH RISKS: THE PROPORTIONALITY TEST[1]

James K. Hammitt
Center for Risk Analysis
Harvard School of Public Health

718 Huntington Ave., Boston, MA 02115
James_Hammitt@Harvard.edu

February 2000

[1] Published in the *Association of Environmental and Resource Economists Newsletter* 20(1): 14-19, May 2000.

The rate at which people are willing to substitute money for mortality risk can be estimated using revealed- or stated-preference methods. Revealed-preference methods are generally considered more credible since it is reasonably assumed that people's choices about real risks are more thoughtful and better informed than their responses to survey questions about hypothetical risks. However, revealed-preference estimates of the value of mortality risk can only be obtained in settings where the alternatives that an individual passes up can be identified and the differences in risk, cost, and other important dimensions can be estimated. Unobserved differences between individual risks and actuarial risk estimates can produce misleading results.

Most revealed-preference estimates of the monetary value of mortality risk have been obtained by comparing workers' pay and on-the-job fatality risk (Viscusi, 1993). After controlling for education and other factors that influence employment opportunities, these studies find that workers in high-risk jobs receive higher wages than those in safer jobs. For example, workers facing an annual occupational-fatality risk of 3 in 10,000 may receive $500 more in annual wages than workers with otherwise similar jobs in which the risk is only 2 in 10,000.

The rate of compensation for risk is commonly expressed as a "value per statistical life" (VSL). In this example, the VSL is $5 million (= $500 ÷ 1/10,000). Since workers who prefer the safer, lower-risk job are willing to give up $500 per year for the risk reduction, 10,000 such workers would together be willing to give up $5 million per year to prevent one expected death among them.

Are the estimates of VSL obtained from occupational-risk studies appropriate for evaluating the benefits of environmental and public-health regulations? A number of factors suggest they may not be.

First, the target populations may include different types of people. Wage-risk studies by necessity reflect the preferences of workers in high-risk jobs, who are generally healthy, male, and young adults. In contrast, environmental and public-health regulations may primarily benefit children or the elderly, or people who are unusually susceptible to pollution due to chronic lung disease, HIV-impaired immune systems, or other factors.

Second, wage-risk studies are based on the preferences of people who accept high-risk jobs, who implicitly reveal a greater willingness to accept risk for money than otherwise similar people who do not accept these jobs.

Third, the types of mortality risks differ. Wage-risk studies are largely based on fatal-accident risks. The mortality benefits of environmental regulations more often come in the form of lower risk of cancer or other fatal disease, which people may value differently.

If the results of wage-risk studies are of limited application to environmental risks, contingent valuation (CV) may be a valuable alternative. CV is an extremely flexible method. One can ask almost any sort of question about a hypothetical choice between alternative situations varying in risk and monetary consequences and experience suggests that most survey respondents will answer. Moreover, the questions can be targeted to the population

most likely to benefit from a specific environmental regulation—the elderly, those with chronic disease, or others with relevant characteristics.

Evaluating Contingent Valuation

Does contingent valuation yield valid estimates of WTP to reduce mortality risk? The fact that respondents will answer survey questions does not in itself imply that those answers are either thoughtful or informed. Other criteria are required to evaluate CV results.

One criterion is the extent to which the values estimated from CV studies agree with estimates from revealed-preference approaches. Some comparisons have been made which show rough consistency between CV and revealed-preference estimates. Yet the value of those comparisons is limited by the fact that revealed-preference estimates can only be obtained for goods with which consumers have experience. These comparisons do not provide direct evidence about the validity of CV estimates in cases where CV is most needed—for novel or unfamiliar goods.

A second criterion is the consistency between CV estimates and theoretical predictions about which factors should, and should not, affect willingness to pay (WTP). For mortality-risk reduction and many other goods one would expect that WTP for a benefit would be larger for people with higher incomes, all else being equal. By contrast, WTP should not depend on logically inessential aspects of the question such as whether the risk reduction is described as a change in probability (from 0.0003 to 0.0002), frequency (from 3 in 10,000 to 2 in 10,000), or odds (from 1 in 3,333 to 1 in 5,000).

Proportionality of WTP to the Risk Reduction

One theoretical prediction that has received much attention is "sensitivity to scope," that is, the extent to which estimated WTP depends on the size of the risk reduction or other good. CV has been criticized on the grounds that estimates of WTP are inadequately sensitive to differences between the items that are valued.

In some applications to environmental quality, respondents may indicate virtually the same WTP for protection of substantially different wilderness areas or numbers of wildlife (Diamond and Hausman, 1994). When respondents indicate they are willing to pay the same amount for improvements of widely differing magnitude, it raises a concern that they are simply expressing general support for environmental protection rather than valuing the specified improvement.

For environmental quality, while it is reasonable to expect that WTP should be larger for a greater improvement there seems to be no clear answer to the question of how much larger is enough. For small reductions in mortality risk, however, there are good reasons to assert that WTP should be nearly proportional to the reduction in risk. Indeed, near proportionality between WTP and change in mortality risk appears to be a necessary (but not sufficient) condition for CV-based estimates to be considered valid measures of VSL. If estimated WTP is not proportional to the magnitude of the risk reduction, the estimated VSL will be strongly sensitive to the arbitrary choice of how large a risk reduction is presented in the CV instrument.

The individual's VSL describes the rate at which he would pay for infinitesimal reductions in risk. It is not constant but depends on income and baseline risk. As the individual buys successive increments his VSL will fall as both his remaining income and his risk decline. But under standard models of decision making, both effects should be small.

The standard model of WTP for reductions in current mortality risk is based on the assumption that individuals seek to maximize their expected state-dependent utility of income

$$U(p, w) = (1 - p)\, u_a(w) + p\, u_d(w) \qquad (1)$$

where p is the probability of dying in the current period and $u_a(w)$ and $u_d(w)$ are the utility of income w conditional on surviving and not surviving the period, respectively (Drèze, 1962; Jones-Lee, 1974; Weinstein et al., 1980). Holding expected utility constant yields

$$VSL = \frac{dw}{dp} = \frac{u_a(w) - u_d(w)}{(1-p)u_a'(w) + pu_d'(w)}. \qquad (2)$$

The numerator is the difference in utility between surviving and dying and the denominator is the expected marginal utility of income. Under the conventional and reasonable assumptions that $u_a(w) > u_d(w)$ and $u_a'(w) > u_d'(w) \geq 0$, VSL increases in risk. Risk aversion in both states ($u_a''(w) < 0$, $u_d''(w) \leq 0$) is sufficient for VSL to increase with wealth.

The effect of risk on VSL—the "dead-anyway effect" (Pratt and Zeckhauser, 1996)—reflects the difference in the marginal utility of income depending on whether or not the individual survives the period. The effect is largest when the marginal utilities are as different as possible, that is, for $u_d'(w) = 0$. In this case, decreasing the mortality risk p by Δp decreases VSL by the proportional change in survival probability $(1 - p)/(1 - p + \Delta p)$. For the usual case where the baseline risk p is a few percent or less, the proportional decrease in VSL is approximately equal to $1 - \Delta p$.

While theory implies the dead-anyway effect is small, it places no obvious constraints on the income effect. Thus, we must turn to empirical estimates. These suggest the income elasticity of VSL is no greater than one.

The primary sources of information on VSL -- studies of compensating wage differentials--typically do not provide information about the income elasticity because income (or wage) is the dependent variable and cannot also be used as an explanatory variable. One approach to estimating the income elasticity is to conduct a meta-analysis of compensating-wage-differential studies where the populations differ in income, risk, and other factors. Liu et al. (1997) used this approach to evaluate the relationship between estimated VSL, average income, and fatality risk for the 17 compensating-wage-differential studies listed in Viscusi's (1993) review article for which these variables were available. They estimated an income elasticity of 0.5.

Income elasticity can be estimated from revealed-preference studies for goods other than employment and from CV studies. For example, Blomquist (1979) estimated an elasticity with respect to the present value of future earnings of 0.3 in his study of seat-belt use. CV studies do not always find a statistically significant relationship with income or report sufficient information to calculate an elasticity. Jones-Lee et al. (1985) estimated a value of 0.3 in a study of transportation risk. Evans and Viscusi (1990) estimated an income elasticity of 1.0 for nonfatal injury risk.

The available evidence suggests that the income elasticity of VSL is no greater than one, and may be substantially smaller. If so, the effect of changing income on the proportionality of WTP to risk reduction is small whenever WTP is a small share of income.

How large a departure from proportionality is consistent with the standard model? Consider an individual with annual income of $40,000 (the approximate average for US households) facing a 28 in 10,000 chance of dying in the next year (the approximate average for US residents aged 25-54). Assume the individual's VSL is $5 million (a standard estimate). How much more would he pay to reduce his risk by 2 in 10,000 than by 1 in 10,000?

For this individual, WTP_1 to reduce mortality risk this year from 28/10,000 to 27/10,000 is equal to the risk increment $\Delta p_1 = 1/10{,}000$ times some VSL intermediate to its initial value VSL_0 (= $5 million) and its value VSL_1 at the final position where his risk is 27/10,000 and his income is ($40,000 - WTP_1). Since WTP_1 is less than $\Delta p_1 \bullet VSL_0 = \$500$, his final income will be greater than $39,500. If his income elasticity is no greater than one, the income effect alone yields $VSL_1 > 39{,}500/40{,}000 \bullet VSL_0 = \$4.9375$ million. The dead-anyway effect decreases this value by a factor no smaller than 0.9999, to $4.9374 million. Thus WTP_1 is between $500 and $493.74.

Similarly, the individual's WTP_2 to reduce his risk from 28/10,000 to 26/10,000 is equal to $\Delta p_2 = 2/10{,}000$ times some VSL between VSL_0 and its value VSL_2 once he has paid for the larger risk reduction. In this case, his final income will be greater than $39,000, the dead-anyway effect reduces VSL by a factor of no less than 0.9998, and so $VSL_2 > 39{,}000/40{,}000 \bullet 0.9998 \bullet VSL_0 = \$4.874$ million. WTP_2 is between $1,000 and $974.80. Dividing the lower bound on WTP_2 by the upper bound on WTP_1 implies that the individual will pay at least 1.95 times as much to reduce his risk by 2/10,000 as he will pay to reduce it by 1/10,000.

The near-proportionality of WTP to change in mortality risk depends on several factors. First, the effect of reduced income cannot be too large, which implies that it is unreasonable to expect near-linearity if the payments are a substantial fraction of income (or if the income elasticity of VSL is much larger than current estimates suggest).

The dead-anyway effect is always small unless the risk change is a substantial fraction of the individual's total survival probability. Note that the effect depends on the individual's total mortality risk rather than the level of risk from any specific cause. Whether the risk reduction to be valued involves a small or large fractional change in a particular risk (for

example, road accidents) is irrelevant, except perhaps if the marginal utility of income if one dies depends strongly on the cause of death.

Near-proportionality does not depend on the assumption that the individual maximizes his expected utility. Most alternative theories of decision making under uncertainty are locally linear in the probabilities (Machina, 1987) which is all that is required. Under rank-dependent expected utility, for example, the individual would evaluate his position using

$$V(p, w) = [1 - \pi(p)]\, u_a(w) + \pi(p)\, u_d(w) \qquad (3)$$

where $\pi(p)$ is a smooth, monotonically increasing function with $\pi(0) = 0$ and $\pi(1) = 1$ (Quiggin, 1993). Holding V constant yields

$$VSL = \frac{dw}{dp} = \frac{p'(p)\left[u_a(w) - u_d(w)\right]}{\left[1 - p(p)\right]u_a'(w) + p(p)u_d'(w)}. \qquad (4)$$

Compared with the standard expected-utility result shown in equation (2), the numerator is multiplied by $\pi'(p)$ and the expected marginal utility in the denominator is calculated using the transformed probabilities. This formula will yield qualitatively similar results to the standard model so long as $\pi'(p)$ does not change sharply between the initial and final risks.

In contrast, near proportionality need not hold under theories of decision making such as prospect theory (Kahneman and Tversky, 1979) that allow for thresholds in the way people evaluate probabilities. For example, if an individual perceives an annual mortality risk of 27/10,000 as equivalent to zero but a risk of 28/10,000 as different from zero, then he would pay something to reduce his risk from 28/10,000 to 27/10,000 but nothing for the further reduction to 26/10,000. Thus, his WTP for the larger and smaller risk reductions would be equal.

Although such a result is possible, probability thresholds seem to be an *ad hoc* and context-specific rationalization. Depending on how the question is framed, the existence of probability thresholds could also yield a much greater than proportional relationship between WTP and risk change. If an individual views a reduction of 1/10,000 as negligible but a reduction of 2/10,000 as meaningful, WTP for the smaller reduction might be zero while WTP for the larger one would be positive.

Another possible reason for non-proportionality in CV studies is that respondents may not report their values for the numerical risk change specified in the question. As suggested by Viscusi (1985, 1989), they may instead combine the stated risk reduction with their own prior estimates of how effective the hypothetical program might be to form a revised, posterior estimate of the risk reduction. Even if the respondents' reported values are proportional to their posterior risk estimates, they may not be proportional to the risk reductions specified in the survey. In this case, it is impossible to estimate the respondents' marginal rate of substitution for money and risk unless the posterior risks they value can be ascertained.

The argument for near proportionality of WTP to change in risk does not require that the individual be willing to pay the same amount to reduce different risks, since it concerns WTP to reduce the same type of fatality risk by different amounts. An individual might be willing to pay different amounts to reduce his risk of dying in a traffic accident and from cancer by 1 in 10,000. Nevertheless, he should be willing to pay nearly twice those amounts to reduce each risk by 2 in 10,000.

The State of the Field

Hammitt and Graham (1999) reviewed the results of every CV study we could find that was published since 1980 and estimated WTP for reductions in numerically specified health risks. We sought to determine whether estimates of WTP were proportional to the risk reduction.

Of the 25 studies we identified, only 14 provided information on how estimated WTP varied with the magnitude of risk reduction. Eight studies involved fatality risks. Of these, WTP was statistically significantly related to the magnitude of risk reduction in six cases and not significantly related in two. In every case, WTP varied much less than proportionately to the risk reduction. Some of these studies asked the same respondents to value larger and smaller risk reductions and found that many reported they would pay the same amount for both reductions.

For example, Jones-Lee et al. (1985) elicited British respondents' WTP to reduce fatality risk on a foreign bus trip by 4/100,000 and 7/100,000 (from an initial level of 8/100,000). Mean WTP are £137 and £155, respectively. Because estimated WTP is not proportionate to the risk reduction, dividing WTP by the risk change yields different estimates of VSL—£3.4 million and £2.2 million, respectively. Moreover, median WTP for the two risk reductions are equal (£50) and 42% of the respondents indicated the same WTP for both risk reductions. (Eight percent indicated greater WTP for the smaller risk reduction.)

Six of the 14 studies evaluated nonfatal risks and revealed a similar lack of sensitivity to the magnitude of benefit. WTP was significantly related to the risk change in five studies but was always much less than proportional to the magnitude of the change.

One reason that CV studies usually yield estimates of WTP that are inadequately sensitive to the risk reduction may be the difficulty of accurately communicating small risk changes to survey respondents. Except for the studies by Loomis and duVair (1993) and Hammitt and Graham (1999), there has been little formal testing of the effect of risk-communication methods in CV.

In recent work, Corso et al. (1999) found evidence that difficulties in communicating small changes in risk may be a major contributor to the generally inadequate sensitivity of CV-estimated WTP to the magnitude of risk reduction. These authors elicited WTP to reduce respondents' annual automobile-accident fatality risk by 5/100,000 and 10/100,000 from randomly chosen subsamples of respondents. Respondents were further randomized to one of three groups presented with a visual risk-communication aid (a chart with 25,000 dots, logarithmic or linear risk ladder) or to a control group that received no visual aid.

Table 1 presents regression models estimated separately for each of the four groups. The models assume WTP is lognormally distributed and include only an intercept and a dummy variable "Large risk reduction" which is equal to one if the respondent was offered the larger risk reduction and zero otherwise.

As shown in the table, sensitivity to scope varied markedly with the visual aid used. In the control group, median WTP for the larger risk reduction is 1.10 times larger than for the smaller reduction. The estimates are not significantly different (the coefficient on the dummy variable "Large risk reduction" is not significantly different from zero) and so the hypothesis that WTP is insensitive to risk reduction cannot be rejected. As a result, the estimates of VSL obtained by dividing estimated WTP by the risk reduction differ by a factor approaching two. In contrast, for the group presented with the dots, median WTP is nearly proportionate to the risk reduction (the coefficient on the dummy variable is not significantly different from log(2) = 0.693) and the hypothesis that WTP is proportionate to the risk reduction cannot be rejected. For this group, the estimated VSL is virtually the same for the subsamples valuing the smaller and larger risk changes. Results for the two groups presented with risk ladders fall between these extreme cases.

Table 1: WTP as a Function of Risk Reduction
(Models estimated separately by subsample)

	No aid	Linear	Logarithmic	Dots
Intercept	5.448	5.630	5.333	5.067
	(0.141)	(0.145)	(0.145)	(0.141)
Large risk reduction[a]	0.097	0.318	0.503	0.658
	(0.198)	(0.202)	(0.198)	(0.209)
Sample size	277	288	264	275
Reject insensitivity?	no	no	yes**	yes***
Reject proportionality?	yes***	yes*	no	no
Median WTP (small,	$232	$279	$207	$159
large)	$256	$383	$342	$306
Ratio of WTP	1.10	1.37	1.65	1.93
Median VSL (small,	$4.6	$5.6	$4.1	$3.2
large) (millions)	$2.6	$3.8	$3.4	$3.1

Notes: Standard errors in parentheses.
[a] Dummy variable equal to one if respondent offered larger risk reduction.
*, **, *** = significant at 10%, 5%, 1%.

Conclusion

Contingent valuation is an extremely flexible method for eliciting preferences about health risks. There are few alternatives for obtaining empirical estimates of the value of reducing mortality risk to a specified population. For CV to fill this need, investigators need to develop methods for conducting CV studies that yield demonstrably valid results. An important criterion for evaluating validity is consistency with other information, including the predictions of reasonable theories of decision making and valuation of health risk. In particular, VSL estimates from studies that do not demonstrate the near-proportionality

between estimated WTP and risk reduction implied by theory must be viewed with some skepticism.

Acknowledgments

This essay is an expanded version of "Valuing Lifesaving: Is Contingent Valuation Useful?" *Risk in Perspective* 8(3), Harvard Center for Risk Analysis, March 2000. Magnus Johannesson, John Loomis, and Jason Shogren provided helpful comments and the US EPA provided financial support. The views expressed may not represent Agency views or policy.

References

Blomquist, G., "Value of Life Saving: Implications of Consumption Activity," *Journal of Political Economy* 87: 540-558, 1979.

Corso, P.S., J.K. Hammitt, and J.D. Graham, "Evaluating the Effect of Visual Aids on Willingness to Pay for Reductions in Mortality Risk," Harvard Center for Risk Analysis, October 1999.

Diamond, P.A., and J.A. Hausman, "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* 8(4): 45-64, 1994.

Drèze, J., L'Utilitè Sociale d'une Vie Humaine, *Revue Française de Recherche Opèrationelle* 6: 93-118, 1962.

Evans, W.N., and W.K. Viscusi, "Estimation of State-Dependent Utility Functions Using Survey Data," *Review of Economics and Statistics* 73: 94-104, 1990.

Hammitt, J.K., and J.D. Graham, "Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?" *Journal of Risk and Uncertainty* 18: 33-62, 1999.

Jones-Lee, M., "The Value of Changes in the Probability of Death or Injury," *Journal of Political Economy* 82: 835-849, 1974.

Jones-Lee, M., M. Hammerton, and P. Philips. "The Value of Safety: Results of a National Survey," *The Economic Journal* 95: 49-72, 1985.

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

Liu, J.-T., J.K. Hammitt, and J.-L. Liu, "Estimated Hedonic Wage Function and Value of Life in a Developing Country," *Economics Letters* 57: 353-358, 1997.

Loomis, J.B., and P. duVair, "Evaluating the Effect of Alternative Risk Communication Devices on Willingness to Pay: Results from a Dichotomous Choice Contingent Valuation Experiment," *Land Economics* 69: 287-298, 1993.

Machina, M.J., "Choice under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives* 1: 121-154, 1987.

Pratt, J.W., and R.J. Zeckhauser, "Willingness to Pay and the Distribution of Risk and Wealth," *Journal of Political Economy*, 104: 747-763, 1996.

Quiggin, J., *Generalized Expected Utility Theory: The Rank-Dependent Model*, Kluwer Academic Publishers, Dordrecht, 1993.

Viscusi, W.K., "A Bayesian Perspective on Biases in Risk Perception," *Economic Letters* 17: 59-62, 1985.

Viscusi, W.K., "Prospective Reference Theory: Toward an Explanation of the Paradoxes," *Journal of Risk and Uncertainty* 2: 235-263, 1989.

Viscusi, W.K., "The Value of Risks to Life and Health," *Journal of Economic Literature* 31: 1912-1946, 1993.

Weinstein, M.C., D.S. Shepard, and J.S. Pliskin, "The Economic Value of Changing Mortality Probabilities: A Decision-Theoretic Approach," *Quarterly Journal of Economics* 94: 373-396, 1980.

AGE, HEALTH, AND THE WILLINGNESS TO PAY FOR MORTALITY RISK REDUCTIONS: A CONTINGENT VALUATION SURVEY OF ONTARIO RESIDENTS

Presented by Maureen Cropper, World Bank and University of Maryland
Co-authored with Alan Krupnick, Resources for the Future, Anna Alberini, University of Maryland, Nathalie Simon, EPA, Bernie O'Brien, McMaster University, Ron Goeree, McMaster University, Martin Heintzelman, Resources for the Future

Summarization

Dr. Cropper presented the results of one of several surveys she and her colleagues have done looking at the impact of age and health status on people's willingness to pay (WTP) to reduce the risk of dying. The reason they are interested in these two questions, she said, is because most of the benefits from air pollution control programs and other health and safety programs that save lives are in proportion to the existing distribution of death. So it matters for policy decisions what happens to people's willingness to pay for risk reductions as they get older. People with certain pre-existing health conditions, like chronic heart and lung diseases, benefit most from regulations like those reducing air pollution, so it is also important to know how willingness to pay varies with health status.

The researchers conducted a contingent valuation survey looking for the dollar amount people were willing to pay for an abstract product to reduce the risk of death over the next ten years. They were also interested in what people would pay for a product that would reduce their risk of dying beginning at age 70. Cropper said she would not focus on those results because of time but said they are also important for policy.

The survey took place in Hamilton, Ontario, Canada, with people between the ages of 40 and 75. The researchers asked people if they had particular chronic illnesses, focusing on heart and lung diseases, and to fill in Standard Form-36 (SF-36). (SF-36 is a quality-of-life survey of 36 questions used routinely by the medical community to measure physical, mental, and emotional health.) The survey was administered at a centralized facility by computer so that people could go through it at their own pace. To insure they did the survey completely, a researcher read each screen to them.

To represent the chances of surviving versus the chances of dying, researchers used a visual aid that worked well with focus groups. This was a grid made up of 1,000 squares, some blank, some filled in, illustrating the baseline risk of dying over the next 10 years and changes in the risk of dying if one bought the product. In focus groups they found that people had a hard time understanding their risk of dying over a short time period. It became a more real concept in the context of the relatively long period of ten years.

They asked people to value reductions in their risk of dying over this ten-year period for either a 1 in 10,000 or 5 in 10,000 annual risk reduction. To do this they asked them what they would pay for an abstract product that would reduce their own personal risk. They used external and internal scope tests and were concerned whether WTP for different size risk changes would increase on average in proportion to the risk change.

The study was set up in two groups or waves. In the first wave, people got a 5 in 1,000 risk change first, in the second wave they got a 1 in 1,000 change first. The researchers compared these answers in an external scope test. Each individual was presented with both sizes of current risk reduction and a future risk reduction of 5 in 1,000 starting at age 70. So, Cropper explained, they were able to do internal scope tests, as well.

She went on to detail the structure of the survey. They started by asking people about their health, if they had conditions such as chronic bronchitis, asthma, high blood pressure, etc. They next talked about chance and communicated information on the risk of death with visual aids and explained the idea of baseline probabilities. They allowed people to do various exercises, adding and erasing squares on a thousand-square grid to visually demonstrate the benefits of using the product and the risks of death. They then tested their comprehension.

Next, they told the respondents what the risk of death was for people of their race and gender over the next ten years. They discussed what people were currently doing to reduce their own risk, such actions as prostrate cancer screening, controlling cholesterol, etc., and told them what were the quantitative reductions in risk for these interventions.

They then posed the WTP questions, asking people if they would buy the product, which would not be covered by health insurance but was shown to be safe and effective. They used a bid structure with payments (in 1999 Canadian dollars) made annually over ten years from which people would receive annual risk reductions. They asked one follow-up question. If people told them “no” twice, that they were not willing to pay the offered amounts of $100 and $50, they asked if there was any amount they would pay. They recorded those amounts and used them later in their estimation of a WTP function.

They included bid amounts that were large enough so people could have a valued statistical life, or VST, of $7.5 million Canadian, which is comparable to the United States VST of $6 million used by EPA. They then asked a series of debriefing questions: did people believe the baseline risk, or believe in the product’s effectiveness, etc. The answers were used as co-variants in looking at factors explaining the differences in the WTP responses. Finally, they gave SF-36 to determine physical and mental limitations.

One concern, Dr. Cropper said, was the low response rate, but she added, they were comforted by the fact that, in regard to health and income, the people in the survey looked like the people of Ontario. They were not, however, as old a group as they would have liked, the average age was 54 and only 9 percent were over 60.

A higher percentage of people were willing to pay the stated amounts for the larger risk change of 5 in 1,000. There were a lot of people in both waves that said they would pay nothing for the product. Unfortunately, they did not debrief them as to why. In their current

surveys, she said, they have included questions to find out why they are getting such large numbers of zeroes.

The big question, she thinks, is do people pass the proportional scope test? Is WTP on average for the 5 in 10,000 annual risk reduction 5 times as large as that for the 1 in 10,000 reduction? If it is, the mean VSL should be the same for the two size risk changes. Regardless of the model used for the data, she said, showing the results from the double-bounded, Turnbull, and spike models, they did not arrive at proportionality.

She then showed some of the results of the co-variants for age and health status. Putting the age variables from wave one into the WTP equation, with dummy variables for the age categories, they ended up with statistically significant results. When they imposed quadratic and linear functions, the results were statistically insignificant.

Next, she showed the results of the re-estimation of the spike model, which they viewed as their best model because it incorporated all the responses. Displaying a table showing the mean WTP and the standard error for the various age groups, she pointed out the statistically significant difference between the $418 mean WTP for the over-70 group and the values for the other groups. Between the peak value for the age group between 51 and 70 and the value for the group over 70 there was a one-third decline. Dr. Cropper cautioned that this was a tentative result.

What was more surprising, she said, was the lack of significance of the effects of chronic health conditions and the SF-36 scores on WTP. To illustrate her point, she used a table of coefficients indicating what happened when they put the chronic-condition variable dummies, one at a time, into a regression that included income, education, age, debriefing variables, and also some of the summary scores from SF-36. The resulting figures showed the maximum impacts of each of the disease categories on WTP. Pointing to the p-values, she said that the only thing that came in as possibly significant was the cancer variable. People with cancer were willing to pay $270 more than people without cancer. Her co-authors would interpret this as more significant then she would, Dr. Cropper said. She cautioned that the results were from only 26 people and that they had obviously dealt well enough with cancer to be able to come in to do a survey at a centralized facility. To say that people with cancer are willing to pay more to increase their life expectancy based on these results she thinks is premature.

What was statistically significant, she noted, was the higher willingness to pay of the respondents with higher mental health. When they looked at the SF-36 data, they found that the p-values relating to people's mental states were significant. People with fewer signs of depression were willing to pay more for risk reduction. Mental status, not physical status, seems to matter in WTP, she concluded.

Dr. Cropper cautioned that she didn't want to claim too much for the survey. One problem was the low response rate. Another weakness was the failure of WTP for the 1 in 10,000 risk reduction to vary with age and health in the wave two results. She also thought that the failure of the proportionality test was a problem.

Her tentative conclusions are that there is some evidence of willingness to pay for a reduction of the risk of death going down after age 70 and that physical health status itself does not seem to have a significant effect on the willingness to pay.

She and her colleagues are currently doing a similar study based on data collected from 1,350 respondents in the US via Web-TV and plan to do another centralized facility survey in Prince George's County, Maryland.

Discussion of Session IIIb Papers

by Steve Crutchfield, USDA Economic Research Service

This participant's remarks are not available.

Question and Answer Period for Session III

Kelly Brown, U. S. Environmental Protection Agency, asked James Opaluch to clarify that the survey he described was actually done and that it was the source of the previously collected data in his study.

Opaluch confirmed that it was an actual survey and that they randomly selected from the sample.

Richard Carson, University of California, San Diego, remarked that in a typical choice experiment you might have ten choice sets and 100 respondents, giving you 1000 choices. This seems enough to give you good results from the binary estimator. The range over which the ordinal estimators do a better job than the binary estimators seems limited to fairly small sample sizes.

Opaluch agreed their study suggests that when analyzing 200 to 1000 choices, the ordinal estimators do a better job but after that the two do about the same.

Carson noted that the proportionality test in the Hammitt study falls out of the textbook formulation of risk. But there are other models, and the evidence from actual markets is either absent or inconsistent with the proportionality test. He suggested that it is nice but not essential to have proportionality, whereas a violation of the general scope test (people not willing to pay more for larger risk reductions) would be disturbing.

James Hammitt replied that one could construct models where proportionality is not expected. But he was troubled by some of the implications that people might try to draw out of non-proportionality, such as arguing that it is better to break up reductions in risks into small packets rather than to consider them all at once.

Daniel Mullarkey, Economic Research Service, USDA, noted that in wetlands work he had found similar results to those of Frank Lupi and John Hoehn. People may know a lot about some wetlands functions but little about others. There is scientific uncertainty and lack of information in the area, which breeds potential for respondents to reject the scenario that you offer. He asked the panel how to screen for scenario rejection.

John Hoehn said one way is to develop tests based on the different approaches to valuation, comparing the results, looking for consistency.

Mullarkey asked about more direct screening. That is, what if you ask a respondent if the scenario was believable and the respondent said no?

Hoehn said we should try to have a general idea in advance about what wetlands services people really value. For example if people value habitat services, we should try to describe those services to people in a really salient way. We also need to better understand what it is about habitat that people value.

V. Kerry Smith, North Carolina State University, posed three questions. First, in Maureen Cropper's study, did they ask respondents about recent changes in activity

limitations? Other research suggests that changes in health or activity signal changes in quality of life perception.

Second, were the people who were unwilling to pay for reduced risk the same people who had difficulty understanding probabilities?

Third, for John Hoehn, there are data from the American Association of State Highway Transportation Officials on prices paid for wetland banking in highway projects. It would be interesting to compare those to what Hoehn found.

John Hoehn agreed.

Maureen Cropper said they asked about current health status over the last four weeks. It is an interesting question whether people become accustomed to physical limitations and if so how long it takes.

Cropper also noted that unwillingness to pay does not correlate to inability to understand the probability questions in the survey. We asked some of these people open-ended questions about their response. Typically they told us that they considered the risk change too small or that they could not afford to pay.

F. Reed Johnson, Triangle Economic Research, asked how should economists deal with ill-informed preference? Does aggressively informing respondents bias results? Can people be rationally ignorant? If a person lacks knowledge of a resource, can a change in the resource affect the person's welfare? Should we be measuring how much people are willing to pay to become informed?

Also, he noted that existence values can be negative as well as positive. In February when the wetlands behind his house are a source of chorusing frogs, his wife is pleased. Two months later when the mosquitoes come, she is not.

John Hoehn acknowledged that wetlands can have negative existence values, noting a situation in Michigan where wetlands may be converting mercury into bioactive forms.

He said that a change in a resource can often matter to an uninformed person.

Does it matter that people are uninformed? As information changes, values may change. We may not want to base policy on values that are highly unstable. So it is important to examine the basis of people's values and their sensitivity to new information. Researchers need to examine and understand the conditionality of values. Some of the unexplained results that we have talked about during this conference may reflect scenario rejection based upon the respondent's knowledge. If we apply these results without understanding them, we risk bias.

James Hammitt remarked that Dr. Johnson's questions were profound and intriguing. Should government function as a referendum or should it provide leadership or act as well-informed people would wish it to act? Hammitt said that if a study were tricked by framing effects, he would not want government to change its policy on that account.

Mike Christie, University of Wales, Aberystwyth, wondered about the possibilities and drawbacks of using web-based surveys. They offer large samples and easy data entry, but how do you control sample size and how can you aggregate the data?

John Hoehn said he was concentrating on developing a questionnaire format that takes advantage of the web. He was not focused yet on the sampling problems.

Maureen Cropper said they were using a commercial service to locate a random sample of households. Subscribers to the service get free web television in return for taking surveys. The service allows them to target by age groups. The researchers can compare the demographic characteristics of their sample against the general population. This is better than putting the survey out on the web and letting people self-select.

Jim Opaluch returned to the "tree falling in the forest" issue of whether a person can benefit from a resource the person does not know about. The answer depends on use versus non-use values. If there is a health effect, through an improvement water quality, there are values even if people don't know what is happening. We can educate people about the connection between water quality and health and measure those values.

But suppose there is a species that exists but everybody thinks it is extinct. Are we really getting value from the species? It is a difficult question.

Daniel Mullarkey noted that lack of knowledge of a good today may change and create value for the good in the future. Don't zero people out just because they don't know now.

Daniel Hellerstein observed that it is dangerous to assume ignorance is bliss, or to assume that if nobody knows, nobody will ever care.

Kerry Smith said that if you expose someone to a latent hazard such as asbestos and the risk is not discovered for ten years, it still has a value even though you cannot change the risky behavior.

Jim Opaluch replied that in that case – a health case – there clearly is a value. But with existence values the question is more difficult. The definition of existence value is the value of knowing that something exists. But is there value if it exists whether you know it or not?

Glenn Harrison, University of South Carolina, thought Opaluch might be confusing the standing issue with valuation.

Richard Carson returned to Johnson's questions of rational ignorance and the value of information. The typical person has little opportunity to influence policy. It is dangerous to draw conclusions from people's not investing in information when their ability to act on the information is limited.

Regarding Opaluch's endangered species example, value in an economic sense is only defined by offering choices. The willingness to spend money to save the species is the only sign to the economist that the species is valuable. Consider new consumer goods – cell

phones, for example. Once upon a time people didn't know they existed. If you take the strict willingness to pay view, they only acquired value when people got to make choices about them. The degree of knowledge is not relevant to thinking about value in that sort of context.

Carol Mansfield, Research Triangle Institute, observed that existence value of wetlands stems from a sense that the wetlands are functioning, not from knowledge of how exactly they function. You can get utility from ecosystem function without knowing whether specific animals or plants exist.

Patrick Welle, Bemidji State University, saw a methodological issue here about contingent valuation, dichotomous choice, and conjoint analysis. We have to be careful about how we separate out some of these attributes in choice experiments. People might not be able to imagine the choice presented and would reject the scenario.

STATED PREFERENCE: WHAT DO WE KNOW? WHERE DO WE GO?

PROCEEDINGS
SESSION FOUR

PANEL DISCUSSION: "THE NOAA PANEL AND THE SEVEN-YEAR ITCH"

A WORKSHOP SPONSORED BY THE US ENVIRONMENTAL PROTECTION AGENCY'S NATIONAL CENTER FOR ENVIRONMENTAL ECONOMICS AND NATIONAL CENTER FOR ENVIRONMENTAL RESEARCH

October 12-13, 2000
Doubletree Hotel, Park Terrace
Washington, DC

Edited by Sylvan Environmental Consultants for the
Environmental Law Institute
1616 P Street NW, Washington, D.C. 20036

ACKNOWLEDGEMENTS

Sections of this report, indicated as "summarizations," were prepared by Sylvan Environmental Consultants for the Environmental Law Institute with funding from the National Center for Environmental Economics. ELI wishes to thank Matthew Clark of EPA's Office of Research and Development and Kelly Brown, Julie Hewitt, Nicole Owens and project officer Alan Carlin of National Center for Environmental Economics.

DISCLAIMER

These proceedings are being distributed in the interest of increasing public understanding and knowledge of the issues discussed at the Workshop and have been prepared independently of the Workshop. Although the proceedings have been funded in part by the United States Environmental Protection Agency under Cooperative Agreement CR-826755-01 to the Environmental Law Institute, it may not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Session IV Proceedings
Table of Contents

Panel Discussion: "The NOAA Panel and the Seven-Year Itch"

Introduction

The purpose of this panel discussion was to explore the state-of-the-art of stated preference research, taking stock of how the field has advanced since the 1993 NOAA panel report, and to discuss the direction future research could take in order to help inform policy decisions.

Panel members were chosen because of their involvement in stated preference research or because of their involvement with the use of stated preference results in a policy context.

Prior to the panel discussion Nicole Owens, US EPA, NCEE, presented an overview of the NOAA panel report. Then, each of the panelists addressed pre-assigned questions. This was followed by an open discussion period.

The following sections of the proceedings contain a combination of notes, summaries, and statements provided by panel members as well as a summary of the open discussion period.

Panel members included:

Richard Carson, University of California, San Diego
David Chapman, DOC, NOAA
Paul DeCivita, Health Canada
Maureen Cropper, University of Maryland and World Bank
Michael Hanemann, University of California, Berkeley
Carol Jones, USDA, ERS
Randall Lutter, American Enterprise Institute
Al McGartland, US EPA, NCEE
V. Kerry Smith, North Carolina State University

Questions for Panelists:

1. What have we learned since the NOAA panel?
   (Carson, Cropper, Hanemann, Smith)
2. What remains to be done to ensure that stated preference results are valid and defensible for use in policy or regulatory settings?
   (Smith, Jones, Lutter)
3. How has your agency used stated preference research in the past and what type of stated preference research does it need for the future?
   (McGartland, Jones, Chapman, DeCivita)
4. What do you see as the three biggest stated preference research priorities?
   (All)

Panel Discussion

Recap of the NOAA Guidelines

by Nicole Owens, US EPA National Center for Environmental Economics

Before our panel turns its attention to discussing what we've learned since the NOAA panel, what remains to be done to ensure the validity and reliability of stated preference methods, and directions for future research, we thought it would be useful to briefly discuss and review some aspects of the National Oceanic and Atmospheric Administration (NOAA) Panel and its recommendations. The point of what follows is to provide some context for our panel discussion, not to critically assess the majority of the NOAA Panel's conclusions.

Federal statutory natural resource damage assessment provisions were first implemented in the 1977 amendments to the Clean Water Act. The Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (or CERCLA) contained provisions for recovery for injury to, destruction of, or loss of natural resources. During the 1980's there was controversy over limits on the use of contingent valuation in assessing damages at the same time that natural resource damage assessment cases were being brought before the courts. After the Exxon Valdez oil spill and subsequent passage of the Oil Pollution Act of 1990, contingent valuation studies gained new prominence in the natural resource damage assessment process.

It was in this context that NOAA convened an expert panel in 1992 to explore whether or not contingent valuation studies were reliable enough to measure total value (direct plus passive use) for the natural resource damage assessment process. The panel report was commissioned as part of rulemaking concerning the natural resource damage assessment and restorations regulations required by the Oil Pollution Control Act. The panel consisted of Kenneth Arrow, Robert Solow, Edward Leamer, Paul Portney, Ray Radner, and Howard Schuman. While the evaluation was conducted specifically within the context of natural resource damage assessment, the panel guidelines affected the contingent valuation method more generally. To some extent, the panel's recommendations shaped the development of the method, use of the results of stated preference studies by Federal agencies, and the direction of research in the area since 1992. This occurred despite some claims that the panel lacked knowledge of contingent valuation techniques, despite that the Panel was not asked to consider the use of contingent valuation in the regulatory process, and despite that the final version of the NOAA rule did not include any specific requirements for how to implement assessment methods.

The NOAA panel was charged specifically with evaluating the use of stated preference studies in determining nonuse values for pollution-related impacts to natural resources. Although important, nonuse valuation is a more narrow application than is relevant for many agency needs. Despite the relatively narrow focus of the Panel and criticism of the Panel, the report does have some relevance to the design of and use of results from contingent valuation studies.

Briefly, the NOAA panel concluded that stated preference studies could provide valid and reliable results and gave several specific and fairly stringent recommendations on

how stated preference studies should be designed and administered to ensure reliability and validity.

The panel's report emphasizes the importance of the scenario surrounding the valuation questions. Respondents need to understand and believe the context in which they are given. The panel also recommended that the payment vehicle must be meaningful to respondents and that respondents be reminded of budget constraints and of available substitute resources.

The panel also noted, among other things, that low response rates would make survey results unreliable, the importance of pretesting, and a preference for conservative design as well as the use of follow-up questions and checks on respondents understanding and acceptance of the scenario.

However, the panel gave three specific recommendations that were particularly controversial. These recommendations helped direct some of the stated preference research, the results of which render some aspects of these recommendations obsolete. Despite this, it is interesting to note that many surveys are still reviewed upon the bases of these recommendations.

These three recommendations are the use of split sample scope test, the use of in person interview, and the use of a referendum value elicitation format.

The results of some of the research presented over the last two days have dealt with these three recommendations. Further the results of other research since the panel have rendered some aspects of these recommendations obsolete.

One important point that we come away with is that all stated preference research should be evaluated on first principles, not just on the basis of one group's recommendations, which, at least in terms of some of the surveys EPA is involved in still happens. This view also seems to be supported by the fact that in the end NOAA did not incorporate any specific standards of performance in the regulations.

References

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

Jones, Carol Adaire. "Use of Non-Market Valuation Methods in the Courtroom: Recent Affirmative Precedents in Natural Resource Damage Assessments." *Water Resources Update*, 109(Autumn 1997):10-18.

Panel Discussion

Discussion of Questions 1 & 4

by Richard Carson, University of California, San Diego

I. What Have We Learned Since the NOAA Panel?

A. A Deeper Understanding of Relevant Welfare Economic Theory

1. WTP and WTA can be very different for a variety of reasons.

Large divergences between WTP and WTA are empirically seen in *both* survey and actual transactions.

2. "Embedding" as used in CV literature is not a well-defined concept but rather should be thought as two distinct concepts with different economic predictions: sequencing and nesting.

3. Sequence effects operate in different directions under WTP and WTA sequences.

4. Sequence effects are likely to be large.

5. Income elasticity of WTP is likely to be smaller than the corresponding income elasticity of demand.

6. Interdependent utility functions, not altruism *per se*, is the source of potential double counting.

This is avoided if either altruism is toward the good (*e.g.*, a wilderness area) or if the agent is aware that other agents will also have to pay for the good.

B. A Deeper Understanding of the Properties of Preference Elicitation Formats

1. Different elicitation formats should produce different results.

Finding "procedural invariance" would suggest non-optimizing agents.

2. Stringent auxiliary conditions are needed for incentive compatibility of a binary choice question.

3. A binary discrete choice question cannot be incentive compatible in the case of:

(a) voluntary contributions, or

(b) private goods.

4. In a double-bounded question, the two responses should not be perfectly correlated.

5. An open-ended type question should produce a substantial numbers of zeros and responses to it should be correlated with any information that is perceived related to cost.

6. In a multinomial choice context, optimal "non-truthful" preference revelation is likely to result in: (a) estimates of marginal tradeoffs between attribute levels that are correctly estimated, but (b) estimates of the "scale" parameter, hence total WTP, that are biased.

The problem is likely to be most severe if only one good will be provided.

C. Recognition that SP and RP Estimate Comparisons Are Consistent With Theoretical Predictions

1. Voting—close correspondence.

2. Voluntary contributions—SP estimates substantially higher than actual contributions.

3. Quasi-public goods—SP estimates slightly lower than and highly correlated with SP estimates.

4. Private goods—SP estimates overstate demand for new goods and overstate price sensitivity for existing ones.

May be "worse" rather than "best" case situation to compare SP and RP estimates.

D. Results Are Sensitive to Scope of Good Being Valued

1. A very large number of split sample studies now reject the scope insensitivity hypothesis.

Further, some "Exxon" scope insensitivity results do not hold up under closer scrutiny.

2. Recent meta-analyses conducted for air quality, outdoor recreation, and wetlands all reject scope insensitivity hypothesis.

3. The finding of a strong correlation between SP and RP preference estimates rejects scope insensitivity hypothesis unless both are insensitive.

4. "Internal" scope tests such as those found in multinomial choice experiments, CV studies valuing multiple levels of a good, and CV studies looking at perceptions of the probability of providing the good or the "size" of the good being provided all tend to reject scope insensitivity.

5. Survey problems, such as vague descriptions, bad payment vehicles, and failure to control for differences in "implied" probability of provision, that can result in scope insensitivity are better understood.

The low power of many statistical tests to reject substantially different estimates is now also better understood.

The greater threat now is probably someone falsely rejecting a study that shows agents do not value substantially larger increments of a good very much more.

Scope insensitivity is a serious issue with respect to valuation of low-level risk. The problem here, however, is the difficulty of risk communication (which also influences behavior toward risk) and the manner in which measures such as the statistical value of life are derived and used.

E. Other Areas of Knowledge Improvement

1. Broader understanding of the survey development process.

2. Development of robust estimation techniques/better understanding of estimation issues.

3. Repeated demonstrations of temporal reliability.

4. Greater recognition that there is not a single "critical" experiment.

II. What Are the Three Biggest SP Research Priorities?

A. Determine the best ways to reduce costs of doing studies (following NOAA Panel guidelines) while still maintaining acceptably high quality.

- Choice of survey administration mode and sample design.
- Choice of elicitation formats.
- Dropping suggested design features: temporal averaging, scope test, offered don't know.

B. *Determine the best ways to systematically conduct studies to fill in the "gaps" in ways that will facilitate doing benefit-transfers in areas where government agencies are most likely to need estimates.*

- Systematic identification of the gaps.
- Development of comprehensive long-term agency plans for filling in those gaps.
- Collection and storage of data in a manner to facilitate benefit-transfers.

C. *Sort out what are basic criticisms of neoclassical economics from what are separate issues with SP methods.*

- Are there any systematic violations of neoclassical theory that are confined to surveys?
- What modifications, if any, should government agencies take in response to these violations?
- What really are the arguments against the use of benefit-cost analysis (and of SP estimates in that context) versus what are the arguments against the use of SP estimates for any purpose?

Panel Discussion

Discussion of Questions 3 & 4

by David Chapman, National Oceanic and Atmospheric Administration

First, I would like to thank the organizers for putting together a very interesting and stimulating conference. It is clear that there is a lot of exciting work ongoing and I look forward to reading the papers as they come out.

I am from the NOAA, the National Oceanic and Atmospheric Administration. NOAA has been keenly interested in the applicability of Stated Preference (SP) methods to both policy and natural resource damage assessments for many years now.

At NOAA there are two main interests we have in using state preference methods: First is in natural resource damage assessments and, second in management decisions for areas such as National Marine Sanctuaries. There are four main statutes that NOAA operates within: The Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA); The Oil Pollution Act of 1990 (OPA); The Coastal Zone Management Act (CZMA); and the National Marine Sanctuaries Act (NMSA). Each of these statues allows NOAA to assess human induced impacts to natural resources. NOAA feels that stated preference methods are viable tools to evaluate management and damage assessment decisions.

How has your agency used SP research in the past and what type of stated preference research does it need for the future?

NOAA uses the results from SP research in both it's management of coastal areas, such as marine sanctuaries, and in NRDA. In areas such as the Florida Keys NMS NOAA is involved in making policy decisions weighing resource protection versus access for such things as recreation. In this arena, SP methods have been very useful in determining the effects of new or novel management methods.

In Natural Resource Damage Assessments, we have used SP methods to both estimate the total value of injured resources and to measure the amount of compensation, in terms of restoration actions, that are necessary to compensate for injuries. It is this latter approach of using SP methods to balance the public losses resulting from an oil spill or Superfund site, with the potential gains from alternative restoration actions that to me is really most exciting.

Since the NOAA's Blue Ribbon Panel in 1993, we have learned quite a lot about a number of the key questions the Panel raised about the applicability of CV to passive use values. And I feel confident that we are at a stage where we can design and implement CV studies that passes the Panel's recommendations. However, at least so far, those studies have been focused on measuring the lost interim use value resulting from an environmental insult. In the intervening years, NOAA has re-directed it focus on estimating the amount and types of restoration that would adequately compensate the public for degradation, or lost of use, of natural resources. Some of the newer stated preference methods such as stated choice paired comparisons seem to show promise in helping answer these technically more

complex questions. NOAA has undertaken preliminary research efforts to investigate these issues. In 1998, NOAA held a workshop on the use of Stated Choice Methods for Resource Compensation. The proceedings from that workshop will be available early in 2001.

In the Lavaca Bay Damage Assessment, NOAA participated with the Responsible Party (ALCOA) in conducting a Stated Choice study to estimate the public losses resulting from a limited closure to fishing in Lavaca Bay, Texas. This study was designed to both measure losses and gains from proposed creation and enhancement of recreational fishing facilities (piers, docks etc) by combining both RP and SP data. The results of this study were used to design compensatory restoration projects for the damage assessment.

Major Research Priorities

And this leads me to my final point: NOAA Research priorities for Stated Preference Methods.

1. Getting costs of high quality, defensible SP studies down.

Instrument Design

Sampling Costs

In-person, all the variants of phone/mail, Internet.

This is a major hurtle for any agency in applying these tools. Unless a study is affordable, we won't be able to undertake it. And it's not just an issue of – is the problem big enough to justify such an expense, we are all very budget constrained. And if we don't undertake the study, it can't be used in the decision making process, whether those are management decisions, or court proceedings.

2. Extending what we already know about combining the RP and SP data. As defined earlier, I would put myself in the agnostic category. I do feel that there is a lot of information that can be gained from both SP and RP data. And as we all recognize there are strengths and weakness to each approach to data collection. And in the end, when we are sitting in front of our computers trying to figure out way to analyze these two data sources, we really need to have a good understanding about what each source of data is really measuring. In some instances we may be very confident that they really are measuring similar preferences and combining in some linear manner in the likelihood function is appropriate, but that may not always be the case, and then we may need think of other ways to extracting information from the two sources of data. For me, using both RP and SP data has some of the most exciting applications.

My final research priority is in understanding more about the transferability of the results of from SP studies. I see this as a question on both, how do we do the best job we can with what we have, and a fair amount has been written on that such as Kerry's paper, and the book by Bill Desvousges et. al but also a question about how to design our work with Benefits Transfer in mind. I think that if we keep in mind that fact that many of these studies will be used to transfer to other situation, or at a minimum to a different time when the population demographics may have changed we may be able to increase a greater

number of government decisions in a timely and effective manner. Some of the SP, or SP/RP work seems to lend itself to very transferable or flexible results. Often, at the time a study is being undertaken, we do not know the final policy that will be proposed for implementation, or the damage scenario that might finally be proven. To ensure that the results of our studies are applicable, they have to be flexible enough address at least a reasonable range of possible outcomes.

Panel Discussion

Discussion of Questions 2 & 4

by Maureen Cropper, World Bank and University of Maryland

QUESTION 2: **What remains to be done to ensure that stated preference results are valid and defensible for use in policy and regulatory settings?**

I would like to focus this question more narrowly on the health area: What remains to be done to ensure that stated preference estimates ***of the value of avoiding mortality and morbidity*** are defensible for use in policy and regulatory settings?

But, at the same time, I would like to broaden the question to encompass ***revealed preference*** methods as well: What remains to be done to ensure that estimates ***of the value of avoiding mortality and morbidity*** are defensible for use in policy and regulatory settings?

The "big ticket" items that drive the benefits of health and safety regulation are the value of reduced mortality and, occasionally, the value of avoiding chronic illness. Most estimates of the Value of a Statistical Life (VSL) that are used for regulatory purposes come from revealed preference studies—primarily labor market studies but, increasingly, studies based on consumer behavior.

The one thing we have learned from stated preference studies—as you heard this morning—is that it is very difficult for people to comprehend risk levels and risk changes. Performing internal and external scope tests is essential, and it is also essential that WTP vary with covariates such as income. Such tests are essential if stated preference studies are to be used for policy.

However, it is also essential that revealed preference studies pass similar tests. These tests, however, are almost never performed. Only one study of which I am aware (Gegax, Gerking and Schulze) uses risk perceptions rather than objectively measured risks in a revealed preference (compensating wage) study. All other studies either assume that people correctly perceive objective risks, or they appeal to correlations between qualitative measures of risk and objective risks to justify using objective risk measures in a revealed preference study.

The failure to test the risk perceptions of people in revealed preference studies is especially surprising in view of the poor performance of subjects who are asked to value risk changes in stated preference studies. The standard justification for not performing such tests in revealed preference studies is that, while many people may not understand risk, there are a few knowledgeable people who do and who "move the market." This may be true, but it is a defense that is not allowed in stated preference studies: When a researcher in a contingent valuation study discovers that a subset of respondents who are "very sure" of their answers behaves more consistently than all respondents, it is usually lamented that policy makers cannot rely on the preferences of such an elite when performing benefit-cost analyses. But, that is exactly what may be happening in revealed preference studies.

To summarize, all studies that purport to value risk of death or illness should be required to:

(1) Provide tests of subjects' understanding of the nature and magnitude of the risks valued;

(2) See how WTP varies with (a) the magnitude of the risk change and (b) income;

(3) Investigate the sensitivity of the results to choice of functional form for econometric relationships and, in the case of revealed preference studies, to the variables one must control for to estimate WTP;

(4) employ adequate statistical methods with regard to choice of sample, number of observations, etc.

The second test, incidentally, is likely to be extremely difficult to perform in hedonic analyses. Estimating an individual's marginal willingness to pay function for risk of death requires solving the identification problem in hedonic markets. In this respect, the contingent valuation method has a great advantage over revealed preference techniques.

QUESTION 4: What do you see as the three biggest stated preference research priorities?

In the health context, to obtain values of the following commodities, which typically drive benefit-cost analyses of health and safety regulations:

- WTP to reduce risk of death today
- WTP to reduce risk of death in the future
- WTP to reduce risk of contracting a chronic disease (e.g., cancer or chronic lung disease) that entails serious morbidity and may increase risk of death.

Panel Discussion

The Health Canada Perspective

by Paul De Civita, Health Canada

PANEL QUESTION #3: How has your agency used SP research in the past and what type of stated preference research does it need in the future?

Context

Health Canada's approach to SP research is influenced as much by our policy and science needs as they are by our available budget resources.

Status

Defensible stated preference studies on health are rare in Canada. Several years ago, there were virtually no SP studies that addressed either morbidity effects or premature mortality (there are 4 wage risk studies with similar results to the US studies). So in order to do our benefits assessments we relied almost entirely on US and UK studies and on the transfer method. While we still rely on the transfer method, over the last three years HC has had the opportunity to commission two primary studies.

Two studies

The first SP study commissioned was a stated choice survey on the acute cardio-respiratory morbidity health effects specifically to be used in air pollution mitigation initiatives. The principal researchers for this survey were Reed Johnson and Bill Desvousges of Triangle Economic Research. The science and policy motivation for undertaking this survey was simply that Canadian data were unavailable.

The second survey is a mortality risk study that has just recently been completed and is undergoing an expert review. The principal researchers are Maureen Cropper (who presented the study at the conference), Anna Alberini and Alan Krupnick. There were strong science and policy needs that motivated us to sponsor the survey including understanding small risk changes and age effect.

Transfer method

Because of resource challenges, both surveys were administered in one city and, as a result, the results may not be representative of the Canadian population. In the design of both surveys, we were conscious that the results would be used in transfer method applications so we encouraged our researchers to include and report the information to allow us to undertake a defensible transfer. In fact, for the morbidity survey we have developed a protocol that allows us to employ equations transfers to generate regional specific values. We plan to do the same for the mortality risk survey.

Government economists usually enjoy working with the researchers and design an instrument from scratch. However, for the mortality risk survey, because of time and

resource constraints, we instead searched for a survey instrument that was already largely developed and adjusted it to meet our policy needs thereby saving us most of the development costs.

Needs of policy makers

With regard to the question of what kind of information do policy makers need to use SP, I believe that the answer to this question is not just a technical one – its not just a question of continuing to refine our approaches to minimize biases. The answer is mostly one of promotion and communication. Economic valuation is not an easy topic for non-economist managers to understand. Our senior policy managers are constantly being approached by stakeholders with simple sounding common sense argumentation that lead them to have doubts about applying SP results.

- We need to take action that will allow us to provide them with the confidence in the scientific integrity of our results.
- We need to continue to communicate these very complex ideas in very simple terms.
- We need to continue to draw parallels between our approaches and approaches used in other economic fields and in business for example.
- We need to be prepared to revisit issues that may no longer be interesting academically.

From my experience, senior managers are prepared to support us if they can be reassured that our methods and results are based on sound scientific principles and are generally accepted among the expert community.

Collaboration

Lastly, I would like to underline the importance of international collaboration on these issues. Sharing our respective research efforts among the US, Europe, Australia and Canada can bring about considerable benefits that include: reducing unnecessary duplication of efforts; cost sharing; cross fertilization of ideas; validation of results; and, international support.

Replication of surveys in other countries not only helps with validation of the results but it will also allows us to better understand national differences in values that would, in turn, allow us to better transfer results internationally. There are already some very good examples of these efforts.

PANEL QUESTION #4: What do you see as the three biggest stated preference research priorities?

Categories

Our three biggest research priorities can be placed in two distinct but interrelated categories: commodity/issue and methodology.

Gains / losses

One methodological issue that needs continued work is the WTP / WTA issue. We need to establish a clear framework to characterize environmental challenges as either gains or losses and then we need to develop defensible questions to elicit WTA values in contingent valuation. This is an important issue because there can be significant differences between the two values that can distort policy if WTP values are used as proxies. The explanation that the absence of budget constraints for WTA questions leaves us with only WTP questions is simply not good enough anymore. Perhaps one result of trying to capture WTA values will be to increasingly use stated choice formats. Canada, with the help of Jack Knetsch who has written extensively on this topic, is working towards articulating a framework to help us systematically characterize environmental challenges as either gains or losses.

Altruism

The second methods priority I would like to flag is evolving the SP technology to defensibly elicit altruistic values – and I am thinking more about the specific challenges for generating values for children diseases, but not exclusively. This issue has long been expressed qualitatively because of these distinct challenges and, as a result, can undermine the importance that policymakers may place on these values.

More premature mortality studies

Thirdly, while there has been quite a bit of activity on this front, we need to encourage more surveys on premature mortality. As you all know, human health benefits and in particular premature mortality are dominating our assessments and as such have also generated a lot of focus from our stakeholders. We need to create a critical mass of literature that looks at a variety of risks.

Paul De Civita, Manager
Economic Analysis and Evaluation Division
Office of Policy and Programme Services
Healthy Environments and Consumer Safety Branch
Health Canada
Room 204, Environmental Health Centre
Tunney's Pasture, pl 0801A
Ottawa, Ontario, Canada
K1A 0L2
Tel.: 613 952 4582
Fax: 613 941 3883
E-mail: paul_decivita@hc-sc.hc.ca

Panel Discussion

What Have We Learned Since the NOAA Panel?

by W. Michael Hanemann, Department of Agricultural & Resource Economics and Goldman School of Public Policy, University of California, Berkeley

Background

The NOAA Panel was formed in unusual circumstances; it was announced immediately following the conference on CV that Exxon sponsored in Washington DC in April 1992.

The conference was a brilliant public relations stunt, and it effectively framed the agenda for the NOAA Panel.

The papers presented at the conference were designed to make a point. The empirical studies were not representative of the state of the art in CV, and the conceptual discussion was not a balanced assessment of the issues. Nevertheless, they became the NOAA Panel's point of departure. The Panel's report reflects this influence.

An example is the question of sensitivity to scope, where the Panel took an issue that is a molehill and made it into something of a mountain.

Since the Panel Report

The focus has largely been constructive — refining and testing SP methods.

What is considered good practice has changed because of the Panel; at all levels, there is now widespread recognition of the importance of

- using a multidisciplinary team to design the survey;
- striving to make the scenario economically consequential;
- using a closed-ended response format;
- carefully testing the questionnaire;
- including debriefing questions in the field version of the study; and
- obtaining a reasonably high response rate.

Approaches to SP have been broadened and new formats have been introduced, including choice experiments ("contingent behavior"), rating, new types of payment card, one-and-a-half bound, deliberative polling, MAUT, etc.

In consequence there is some blurring of the boundaries — e.g. choice experiments combine features of CV with revealed preference.

There Is Now Recognition of the Commonalties Between RP and SP

Instead of exaggerating the differences and viewing them as mutually exclusive, as at the Exxon Conference, researchers are more open to combining approaches and are more aware of the similarities.

The key commonality is that RP and SP both focus on individual preference and behavior with respect to specific narrow commodities, and they both confront the irreducible complexity of human cognition and behavior. They are both forms of what may be called *Disaggregated Choice Analysis.*

Distinguish two types of demand analysis:

> (A) Aggregate data on very broadly defined commodities (housing, food, transportation, etc)
>
> (B) Disaggregated data on very specific commodities (e.g. 16 oz bottle of Hunt's low-cal, organic, tomato ketchup in a plastic, easy-pour container)

Those who worked with (B) tend to be well aware of

> (i) Profound heterogeneity in behavior among individuals whom economic theory would consider identical (same prices, same income, etc);
>
> (ii) Preferences are complex. They depend on a variety of attributes that can vary with the situation; and
>
> (iii) Real behavior is by no means as simple or straight forward as in economic theory.

Theorists and those who worked with (A) tend to be clueless.

Many of the Things We Have Learned Were Known Already to Practitioners of Travel Cost

We knew that there is no such thing as procedural invariance.

> E.g., When you try to measure the number of times somebody visited a beach, or how many hours they spend watching TV, there is not invariance with respect to either the mode of asking the question or the context in which the question is answered.

Looking at the demand for Boston area beaches in my Ph.D., I found

- Both perceptions of quality and preferences for quality were context-dependent.

- Objective measures of beach and water quality played little role in explaining people's behavior, while their subjective perceptions had a significant impact.

- Possibility of cognitive dissonance — behavior might shape perceptions, rather than the other way around.

Complexity of Preference

People care not just about what they pay and what they get, but also about

- whether they are overpaying;

- who else is paying;

- whom they pay; and

- what is being done with the money they pay.

These are all potential arguments in a Generalized Lancaster utility function.

The quantity units are subjective. Whether an orchestra has 20 or 30 violinists may hardly matter; whether it has 2 or 3 might matter hugely. With subjective response to quantity, as with attributes, one is dealing with psychophysics.

Context Matters For Both Preference and Perception — They Are Both Context-Dependent

The same words can mean different things (Harold Pinter).

The meaning — of words and commodities alike— is implicit in the situation. It is also socially determined — there is a shared understanding of implicit meaning.

Therefore, different attributes matter, and the same attributes get different weights, in different situations.

This has implications in both RP and SP for

- the design of survey instruments,

- the analysis and interpretation of survey data, and

- the extrapolation of results from the survey data analysis.

The goal must be to understand and model the context-dependence of preferences and of survey response strategies. Use this knowledge to design surveys, analyze the survey data, and match context when extrapolating the survey results.

This makes benefit transfer harder, but it is necessary for good science.

Some Lessons for SP

Concreteness and realism matter. Avoid a scenario that is overtly hypothetical or counterfactual. Avoid a scenario that is incomplete (leaves important details unspecified). In both cases, respondents may deal with this by making "best case" assumptions.

Emphasize making the payment right now.

Choose the right context — since sequence matters, choose the sequence that is relevant.

People don't want to overpay; therefore,

- Avoid open-ended
- Seriously problem for payment cards
- Prefer one and a half bound.

One can detect and correct for yea-saying — this should be done.

Certainty of response should be investigated and accounted for.

Use simple split-sample survey experiments to investigate respondent behavior in surveys.

Issues

- How can one impart realism and concreteness to choice experiments? How does a lack thereof influence outcomes in them?
- Are multiple pieces of data from the same respondent as good as less data from more respondents? I am dubious, because of both the correlation among successive responses, which reduces the amount of information, and also the heterogeneity among individuals, which is undersampled.
- How reliable are self-administered surveys (mail, internet)? I am concerned about both selection bias and the loss of quality that comes from the presence of an interviewer.

Research Areas

1. Survey mode

   Test mail-telephone against in-person.
   Test internet against mail-telephone, phone and in-person.

2. Introduce careful debriefing into choice experiments. Use survey experiments to test whether respondents accept the scenario and are valuing what the researcher assumes.

3. Design experiments to test whether and how(1) economic consequentiality, and (2) hypotheticality/unrealism affect survey responses.

4. Investigate — through data collection, statistical analysis and Monte Carlo simulation — the tradeoff between more subjects and more questions per subject.

Panel Discussion

Discussion of Questions 2, 3, & 4

by Carol A. Jones, Associate Director for Research, Resource Economics Division, USDA Economic Research Service

Question 2: What remains to be done to ensure that stated preference results are valid and defensible for use in policy or regulatory settings?

I am assuming that the panel addressed this question from the ***methods*** perspective in response to the first question. I will focus on issues associated with ***applications*** of the methods to policy and regulatory settings.

A. First, we must recognize that the requirements for validity, precision *will depend upon the specific context at hand:*

1) What levels of validity, precision are required by the decision-making context?

- Burden of proof is different for litigation (the context for the NOAA regulations for natural resource damage assessments) vs. regulatory contexts: in litigation, one must establish the "weight of the evidence", whereas for challenges to regulations, the agencies must meet an "arbitrary and capricious" standard

- Budget constraints may be different: the potential for cost-recovery when the government wins in litigation may lead to a relaxation of the otherwise very stringent budget constraint the government faces

- [Of course most legal cases are settled not litigated, so the trade-off is somewhat of a moving target.]

2) What is the value of additional refinement of the analysis?

- The key questions include: will improved estimates change the policy conclusion *or* will they improve the likelihood and timeliness of a reasonable settlement?

- There may be trade-offs between unbiasedness and variability of estimates:

    - E.g., for cost-benefit analysis, if the data support doing the project based on a downward biased estimate of benefits and/or upward biased estimate of costs, then what is the value in further refining the estimate?

        - Analogously if the expected biases work toward overstating the net benefits of a choice and they do not support doing the policy, one has to ask if there is a benefit of additional information.

    - Of course it's not so simple when the choice structure is not a simple 0,1 option.

B. In order to value to what extent policies generate benefits, we must first have the capability of relating policies to specific outcomes — *in order to know what to value.*

As an example we take the case of an agricultural policy with environmental implications. An essential ingredient to analysis is developing modeling architecture that creates linkages between economic models and environmental models. In this case, linkages need to be made among 3 sets of models:

- Economic models of private decision-making in response to policies (e.g., farmer management of nutrients in response to TMDLs, with outputs that may include quantity of nutrients transmitted to edge-of-field)

- Environmental models that translate the outputs from economic behavior (e.g., quantity of nutrients transmitted to edge-of-field) into quality attributes of natural resources, (such as inland, estuarine water quality) that can feed into:

- Economic valuation models of the natural resources (based on either value of use of resources, or direct valuation of resources) — these use as inputs the changes in resource quality resulting from policy changes and provide the final link between, say, water quality policy and the value it may provide to the public in improved water quality

Accomplishing these linkages takes long-term investments in inter-disciplinary research, which is not consistent with the standard reward structures in most academic and other research organizations. Promoting this work will take require sustained commitments by funding agencies and creative organizational responses by research organizations.

C. We have to be able to conduct valuation in a cost-effective way, in many cases with a minimum of data collection and a maximum range of scenarios covered.

Basic strategies for meeting this goal include:

1) Designing valuation studies to be as flexible as possible for evaluating projects within the specified policy context, potentially long after the survey has been completed.

- Public decision-making processes may become very extenuated, though beginning before an SP survey is started, concluding with a final decision long after the survey has been concluded. Circumstances change, options are eliminated, and a 1 or 2 fixed scenario CV survey may not provide values for what the ultimate options turn out to be.

- Consequently, stated choice elicitations that make it possible to develop valuation functions could serve a very important role in providing the needed flexibility.

2) Developing strategies for conducting benefits transfer from analysis sites to policy sites in the most effective and efficient way.

- Practical reality is that it is not feasible to collect data in each policy context

- At the same time, it has been documented that there are limits to the reliability of benefits transfer approaches to different contexts with different populations — so more work needs to be done there.

Question 3: How has your agency used stated preference research in the past and what type of stated preference research does it need for the future?

ERS is an economic research unit in the USDA, providing economic research for all agencies within USDA (except for the Forest Service, which conducts its own economic analysis). Its formal mission is to conduct economic analysis on efficiency, efficacy and equity issues related to agriculture, food, the environment and rural development to improve public and private decision-making. (See the ERS website, www.ers.usda.gov.)

Two broad policy areas in which we have conducted and/or are currently conducting SP analysis are:

1) Food and drinking water safety

2) Environmental policies or policies with environmental implications (along with income support goals), including policies promoting:

- Use of environmentally sound practices

- Set-aside of environmentally sensitive land — wetlands, or lands that can generate environmental damage is cropped — e.g., highly erodible land

- Agricultural lands preservation

Most ERS analysis is conducted in-house. A major exception is the Food Assistance and Nutrition Research Program, which provided approximately $10.0 million in external funding in the area during FY 2000.

A. Valuation of reduction in morbidity and mortality risks from consumption of food and drinking water consumption (private market goods)

USDA policy role:

- Each year, there are approximately 76 million cases of food-borne illnesses, including about 5000 deaths. USDA, EPA and FDA are responsible for regulating health risks in food.

- For drinking water, USDA has a role in promoting farmer behavior that may reduce drinking water contamination, for example nitrates in well water.

ERS research accomplishments:

- Drinking water

To assess consumers' WTP for safe drinking water, ERS included multiple-bound discrete choice SP questions in the National Survey of Recreation and the Environment and has analyzed the data in several reports.

References

Stephen R. Crutchfield, Joseph C. Cooper, and Daniel Hellerstein. 1997. "The Benefits of Safer Drinking Water: The Value of Nitrate Reduction." USDA/ERS Agricultural Economics Report No. 752, 15p.

Stephen R. Crutchfield and Joseph C. Cooper. 1997. "Valuing Risk Reduction: The example of Nitrates in Drinking Water." Food Review 20(1):38-41.

"Outdoor Recreation in American Life: Participation Trends, Final Report.

Results from the National Survey on Recreation and the Environment." 1998. In Cordell, H. Ken et. al, 1998. Outdoor Recreation in American Life: A National Assessment of Demand and Supply Trends. Champagne, IL. Sagamore Publishing

- Food safety

ERS has two ongoing multi-year cooperative agreements in this area.

ERS current research goals:

The emphasis in current research is to provide an empirical foundation for moving beyond regulatory impact analyses (RIAs) using cost-of-illness and revealed preference value of life literature, based primarily on wage market studies. The goal is to be able to value risk reduction in a specific context, including different populations (children, elderly, etc) and to value specific impacts, including morbidity.

B. Environmental risks

USDA policy role:

USDA has a number of programs to promote the use of more environmentally beneficial agricultural practices and land use. Examples include rental or easement payment programs for temporary or permanent land set-asides or cost/share programs to induce farmers to use environmentally beneficial practices.

ERS research accomplishments:

The research has focused on the two sides of the coin in policy design:

- The supply-side, in order to assess the minimum amount of compensation that is needed to change farmers' behavior, and

- The benefit side, in order to target better the programs to producers.

1) **Supply side**: willingness of producers to adopt new production practices (or change land use) — in order to better design compensation policies or insurance policies to promote economically efficient adoption.

   a) **Best Management Practices**. Environmentally beneficial production practices, often referred to as "best management practices" (BMPs), are encouraged by the USDA. Adoption of these practices in concept is profitable to producers in many cases, though they may involve incurring investment costs. Despite this, many producers have chosen not to adopt them, and little was known about determinants of adoption. In the 1992 Area Studies Survey, we elicited information from farmers with which we estimated an adoption schedule used SP techniques drawn from the CVM literature. The results of this research influenced the Agricultural Conservation Innovation Center in the development of BMP-Plus, an insurance program designed to encourage farmers to adopt BMPs.

References

Joseph Cooper. 1997. "Combining Actual and Contingent Behavior data to Model Farmer Adoption of Water Quality Protection Practices," Journal of Agricultural and Resource Economics, Vol. 22(July):30-43.

Joseph Cooper and Russ Keim. 1996. "Incentive Payments to Encourage Farmer Adoption of Water Quality Protection Practices," American Journal of Agricultural Economics, Vol. 78(February):54-64.

Peter Feather and Joseph Cooper. 1995. "Strategies for Curbing Water Pollution." Agricultural Outlook, Vol. AO-224 (November).

Peter Feather and Joseph Cooper. 1995. "Voluntary Incentives for Reducing Agricultural Nonpoint Source Pollution." USDA/ERS Agricultural Information Bulletin No. 716. 11p.

Margriet Caswell, Keith Fuglie, Cassandra Ingram, Sharon Jans and Catherine Kascak, 2001. "Adoption of Agricultural Production Practices: Lessons Learned from the U.S. Department of Agriculture Area Studies Project." USDA Economic Research Service Agricultural Economic Report No. 792, forthcoming.

Analogous issues arise with conservation tillage, which are considered to be profit-maximizing investments for farmers in many circumstances.

b) **Conservation Reserve Program** (CRP): In the CRP, rental payments may be larger than necessary, especially for farmers who enrolled in earlier rounds. In 1993, trichotomous choice SP survey questions addressing farmer re-enrollment in the CRP were elicited from over 8,000 CRP contract holders. With this data, ERS estimated acreage re-enrollment as a function of the rental rate.

Reference

Joseph Cooper and Tim Osborn. 1998. "The Effect of Rental Rates on the Extension of Conservation Reserve Program Contracts," American Journal of Agricultural Economics, Vol. 80(February):184-194.

2) **Benefit side:** valuation of non-market environmental benefits of policies

a) **Conservation Reserve Program, Wetlands Reserve Program, etc:** To complement benefits estimations based on benefits transfer of revealed preference studies of hunting, wildlife viewing and freshwater based recreation, ERS conducted a CV survey to measure changes in **total value** of grassland bird populations due to CRP. Two papers are currently in draft form.

b) **Valuation of rural amenities from agricultural land use:** are the benefits primarily realized by local residents and therefore captured in market land valuations, or is there a substantial component of value from others in the region or beyond?

In states with active expansion of urban/suburban land use (particularly the west and east coasts), there is a tremendous amount of policy activity in farmland preservation. A series of policy tools have been employed. The newest ones establishing markets for easements in which property owners sell their rights, apparently in perpetuity, to develop their land. Purchases may either be the public sector (public development rights) or other land developers seeking dispensation for development elsewhere in the region (tradeable development rights). USDA has a small program in farmland preservation. This work is currently ongoing.

Question 4: What are the three biggest stated preference research priorities?

A. Incentive compatibility (i.e., incentives for truth-telling) of alternative formats for stated preference methods

Many of the "biases" that have been identified in the literature can be traced to the incentive properties of the survey instrument. The line of work begun by Carson, Grove and Machina is very important, in that it *differentiates* incentive properties among different

elicitation formats and provides many testable hypotheses about differing results on validity across the literature. This line of work is extremely promising for high returns: it provides an important organizing principle for a meta-analysis of the extensive but fragmented stated preference literature. It should provide guideposts for understanding the bias and reliability of survey data, and the most appropriate approaches for analyzing and interpreting the data.

B. Methodological development of choice experiments

Among the stated preference portfolio of valuation methods, substantial investments have been made in the development of the single (or 2) scenario approach of "contingent valuation (CV)." The CV framework for eliciting stated preferences has limited flexibility for use in policy analysis — the policy outcomes to be valued have to be well-specified ahead of time.

In contrast are **choice experiments**, a less well-developed stated preference approach, in which survey respondents are given repeated opportunities to choose among alternative policy outcomes in which several variables (attributes) are allowed to vary. Because choice experiments allow the analyst to estimate valuation functions for multiple attributes, it is possible to value a wide range of scenarios with changing levels of attributes, rather than simply 1 or 2 pre-defined scenarios.

- The approach has the potential for several major advantages over the CV framework, including:
    - it facilitates a broader evaluation of the efficient scale of programs, rather than evaluating simple yes/no choices of 1-2 pre-defined scales;
    - it facilitates valuing provision of multiple public goods, which is critical when valuing the providing of alternative bundles of public goods (as opposed to measuring damages to natural resources from an accident, where the appropriate approach is to value the damage holding all else constant).
- At the same time, a variety of methodological issues arise in implementing any stated preference approach, including the nature of the incentives for truth-telling, which remain to be evaluated for this approach.

C. Understanding and identifying whether respondents scenario rejection, or scenario redefinition

We need to be able to diagnose when respondents are either rejecting the scenarios, or when they are redefining the question to terms that they think are more plausible.

Panel Discussion

Stated Preferences: An Outsider's View

by Randall Lutter, American Enterprise Institute-Brookings Joint Center for Regulatory Studies

I. Why Me?

- Never a producer of stated preference estimates, I am a sometime user of CV.

- Formerly close to decision-making

    - At OMB and CEA I presented estimates of costs and benefits for a variety of environmental initiatives to senior officials.

- Will offer a pragmatic outsider's views about

    - the overstated importance of CV
    - its validity and defensibility for use in policy and regulatory decisions
    - the underrated importance of context.

II. Why CV?

CV can usefully complement revealed preference studies for categories of benefits that can be analyzed with both. For example, estimates of the value of reduced mortality risk are made more reliable by complementary estimates based on both revealed preference and stated preference methods. My remarks focus on applications of stated preference to areas where revealed preference estimates do not exist and can provide no benchmark for comparison.

Pure science: Understanding what people say/mean when asked questions about payment for natural resources.

Policy: Environmental economists justify CV on pragmatic grounds: the environment will be more highly valued in decision-making than would otherwise be the case and no better decision aids are available. (Pearce et al. 1989, Pearce 1998)

A. Let's Consider This Notion Carefully

In U.S. most environmental policy is federal.

In federal policy-making, benefit-cost info has 2 uses:

Management tool

For decision-making within EPA, Administration

Public accounting device

Informing the public about the merit of different policies

B. As A Management Tool Does Benefit-Cost Analysis Suffer From The Non-Monetization of Passive-Use Values?

Non-quantification and non-monetization are routine. Hahn et al. (2000)[1] surveyed 46 economically significant rules and found:

Only 70 % had any quantitative benefits estimates.

< 50 % had any benefits estimates in $.

EPA data were similar. See Figure 2 from page 212 of Hahn et al. (2000)

Non-monetization is common outside of categories of benefits related to passive use.

Non-quantification occurs because:

No data on exposure to toxic substances

Epidemiological evidence is too crude

Non-monetization occurs because:

Some health effects are not monetized in the literature:

birth defects
sterility
neurotoxic developmental effects

How many of the decisions would be improved by better or new CV estimates of previously un-monetized categories of benefits? My subjective answer is very few. As a management tool improved economics has had little value in environmental policy making.

Thus as a practical matter, non-monetization of passive use values, the forte of CV, may be a small part of a superficial effort to rationalize administrative decision-making.

C. Public Accounting: Would greater and better use of CV improve the public's understanding of the effects of environmental policy? Perhaps. But this is a

[1] See Hahn et al. (2000) at http://www.aei.brookings.org/publications/working/working_00_01.pdf

hard question, so I answer an easier one. Has the past use of CV contributed to improving the public's understanding of environmental policies?

Most CV studies value benefits well before action is taken. Thus they inform people about the reason to take action, but not about the merit of specific actions taken.

Solutions to the benefits transfer problem are so poor as to hinder credibility.

Species extinction and preservation of wildlife areas are understood to be important even without expressing this importance in terms of dollars.

D. Thus CV, if well done, may solve less of the problem of inefficient resource allocation than its proponents suggest.

III. Validity and Defensibility

A. EPA already uses CV in regulatory decision-making

1. Examples:

- Great Lakes Water Quality Initiative
  - WTP to fish in waters "free from" contaminants
- Coastal Zone Oil and Gas, WTP for wetlands
- Pesticide Management Plans, WTP for clean ground water
- Regional Haze, WTP for visibility

2. Did such CVs help decision-makers? — Not really.

- No credible evidence of scope or valuation on the margin.
- Difficult problems of benefits transfer. Waters are never "free from" contaminants.
- Little internal consistency/ validity
  - Protest bids, inability to understand or believe scenario.
- Little external validity: non-random survey, very low response rates.
- No treatment of uncertainty

These difficulties are echoed in those reported at this conference where WTP estimates vary sharply with estimates of the confidence that respondents had that their answers were right.

B. Could CV serve as basis for greater consideration of BCA than allowed by all environmental statutes other than TSCA, old FIFRA, and perhaps the new SDWA?

CV as practiced to date is not up to judicial scrutiny.

Substantially greater validity and relevance would be needed.

IV. Decision-Making Context Affects Estimated Values More Than We Acknowledge

A. "I struggled with this money business."

Many respondents dislike and resist answering WTP questions. See Clark, Burgess and Harrison (2000) who analyze respondents to a survey that sought to adopt NOAA recommendations wherever possible. "[Respondents] unequivocally rejected CV as an acceptable means of representing their values or views to decision makers."

Many difficulties arose:

- "[Respondents] felt it was impossible for them to make a meaningful judgment about the worth of the scheme in relation to the large number of probably equally worthy schemes around the country." p. 55.

- Of 31 visitors "asked directly if they felt that the amount that they agreed to pay was a good measure of what conserving wildlife on the Levels was worth to them, 19 answered no, six answered yes, and the rest were unsure or avoided the question." p. 55.

- "There was consensus in all three groups that decisions about such things should be made by government, advised by experts who had an understanding of relative claims of different places and different nature conservation schemes, and based on national standards." p. 56.

- Post-survey discussions increased doubts about the use of the WTP figures and feelings that participants had been duped. p. 56.

Some related literature reaches similar conclusions. Thus there may not be much internal validity.

B. The Process May Affect WTP Estimates.

- When groups were told how WTP figures are analyzed and what the results might mean to economists or decision-makers, a number of individuals expressed anger and distress, feeling that they had been manipulated. ("Don't sort of hoodwink us, you know.") p. 57.

- This suggests that the use of WTP estimates in any decision-making process may affect stated WTP. Consider an extreme and heuristic example that is timely but not perhaps legally sound:

- EPA bans mercury emissions under Toxic Substances Control Act. CV estimates of the value of loons and panthers figures as part of benefits analysis that is subject to judicial review under TSCA's unreasonable risk standard. Inadequate or unreliable monetization may leave EPA's rule fatally vulnerable to legal challenges.

- EPA mandates such stringent technology based (MACT) standards under the Clean Air Act that coal consumption is infeasible and electric generating plants switch to natural gas and oil. The legal standard for MACT prohibits any consideration of benefits. Thus if EPA were to conduct a benefits analysis it would be only for use as a minor managerial tool, in that the law precludes the consideration of benefits in setting the standard, and for public accounting purposes. The benefits analysis would be exempt from judicial review.

- I conjecture that stated WTP would vary according to the prospective use of the WTP estimates.

- But such variation would have very troubling implications for the interpretation of stated WTP. How could it reflect exogenously given preferences, if indeed these varied with the context in which stated preferences would inform policy-makers?

- Stated preference methods have a long ways to go to have the validity necessary to be a respected contribution to informed decision-making.

Panel Discussion

Discussion of Questions 3 & 4

by Al McGartland, Director, US EPA National Center for Environmental Economics

QUESTION 3. How has your agency used SP research in the past and what type of SP research does it need for the future?

EPA's use of stated preference research tends to be in a benefit transfer context.

Economists tend to represent a small minority in the Agency. The Agency employs many engineers, risk assessors, toxicologists, lawyers, etc.; but there is only a small community of economists. At EPA, economists really are the "tail of the dog." That is, we take what hard scientists provide us and attempt to estimate the benefits of changes in environmental conditions using this information. Being able to employ the "damage function approach" makes us feel more confident about our benefit-cost analyses. In this approach changes in emissions or concentrations of pollutants are translated into changes in health endpoints.

The Office of Air is one office for which there are a relatively large number of studies allowing EPA to provide estimates of changes in health endpoints. In the case of mortality, from the valuation context, however, economists must still use estimates provided by the hedonic wage literature. But is this the right value to use for estimating the value of risk reductions for environmental pollutants? It is unlikely because in the environmental context we're often dealing with long-term illnesses such as cancer and/or diseases with a latency period – that is with deaths that have different attributes than those dealt with in the hedonic wage literature. Further, in many cases health scientists don't have a good understanding of some of the attributes of these illnesses, particularly latency. In many cases economists are only as good as risk assessors can make us.

Dealing with water may be even more of a problem because scientists aren't always able to translate changes in environmental pollutants into effects economists are able to value.

Economists may be further hampered by other developments and shortcomings in other fields. For example, the Agency is developing cancer risk assessment guidelines that move away from the provision of continuous dose-response functions for many contaminants. Risk assessors are much more comfortable providing a contaminant level above which is considered "safe" and below which is considered "unsafe." This movement actually makes it more difficult for economists to provide benefit estimates. Recently we convened a meeting of economists and toxicologists at which I made the case that for benefit-cost analysis we need a shift in thinking.

In a few cases, the Agency has tried to conduct/fund a stated preference study in anticipation of a regulation. However, we haven't had much success with these studies. For example, EPA funded research on visibility that hasn't fared well in the literature. EPA also

funded a study on the value of protecting groundwater. It was believed that because groundwater is relevant to a few programs (pesticides, solid waste, water), the payoff to having reliable values would be large. However, this study was also not well received.

So, we remain in the benefit transfer game. As it stands there aren't a lot of standards that govern benefit transfer. There are examples of both good and bad transfers both within and outside of the Agency. I believe that Kerry's *(V. Kerry Smith, North Carolina State University*) idea of preference calibration and benefit transfer will allow us to do a better job.

I'd also like to make a plug for more "replication" studies. While these may not be as publishable as those dealing with new methods, theory, or even a new commodity, there is great value for them at EPA.

QUESTION 4. What do you see as the three biggest stated preference research priorities?

My office conducted an intranet survey of economists that asked them to identify where EPA should spend it's economic research dollars. Those that involve the use of stated preference methods are noted in bold in the table below.

Top 8 Research Areas Identified by EPA Staff Economists	
1	**Estimation of ecosystem services benefits**
2	**Estimation of morbidity risks**
3	**Estimation of other welfare benefits**
4	Uncertainty and economic analysis
5	**Estimation of mortality risks**
6	**Estimation of non-use benefits**
7	Equity and Distribution
8	**Estimation of benefits to vulnerable populations**
Source: Report on the Results of the Agency-Wide Economic Research Agenda Questionnaire (May 1998)	

My three research priorities are:

1. Value of groundwater improvements or protection.
   Again, this affects many offices and the lack of a core study accepted in the literature means economists can't provide monitized benefit estimates.

2. Value of improvements to coastal and estuarine waters.
   Research valuing national improvements in these areas does not exist.

3. Value of reductions in mortality risks related to environmental causes.
   The current transfer of $5.8 million (1997) to all risks is too simplistic and doesn't take into account how the nature of the risk and the death differs from those considered in the hedonic wage literature. Additionally, this category tends to be the major benefit of many of EPA's regulations.

Panel Discussion

Responses to the Itch

by V. Kerry Smith, Center for Environmental and Resource Economic Policy, Department of Agricultural and Resource Economics, North Carolina State University and Resources for the Future

Outline

- Historical Perspective
- NOAA/CV Performance Standards
- Issues Posed to the Panel
- My Answers

What Did the NOAA Panel Say About Reliability?

"If a CV survey suffered from any of the following maladies, we would judge its findings 'unreliable':

- a high nonresponse rate to the entire survey or to the valuation question
- inadequate responsiveness to the scope of the environmental insult
- lack of understanding of the task by the respondents
- lack of belief in the full restoration scenario
- 'yes' or 'no' votes on the hypothetical referendums that are not followed up or explained by making reference to the cost and/or the value of the program." (Arrow *et al.*, 1993 p. 4609)

Presenting an Object of Choice

- Alaska Survey
- Montrose Survey
- CV/SP Studies and Reliability

What was Learned?

- Scope Test Satisfying the NOAA Panel Guidelines

CV / SP Studies and Reliability

Construct Validity (Mitchell and Carson, NOAA Panel)

CV responses related to:

- cost or financial consequence
- measure of availability if relevant to access to what is offered
- income
- factors related to quality of object of choice
- availability of substitutes
- taste-related demographics and attitudes

Consistent with Adding-Up Property

Headlines Conditions

Issues Posed

I. **SP Reliability —** current status and research to enhance it
II. **SP and Policy** — design and evaluation of policy
III. **SP Research** — methods or applications

CV Research Since the NOAA Panel

- Prompted the most serious investigation of individual preferences ever undertaken in economics; types of research include:
  - refinement in econometric methods (new parametric, semi-parametric and non-parametric methods)
  - application of repeated choice, preference scaling, ranking and matching questions with focus on attributes of commodities
  - investigation of incentive properties of different elicitation modes using theoretical, experimental and survey methods
  - integration of revealed and stated preference data in joint estimation of preferences

- Transformed framework used in experimental economics:
  - conventional experimental economics — evaluates performance of institutions using *induced preferences*
  - new environmental economics uses *known incentive properties* of institutions to estimate preferences and evaluate ways of eliciting them
- Supplement to revealed preference methods at a very general level in that methods argued we can learn about individual preferences for goods whose consumption is rationed by prices.

Path to Reliability

- There is no crucial experiment (or set of experiments) that once conducted will allow a decision up or down with the method. This strategy will never succeed.
- Reliability will not be realized by focusing on estimating values for well-defined changes in an environmental objective of choice. Instead must estimate economic value as part of larger set of preferences.

Policy and CV

- Focus on measuring Hurwicz-Uzawa income compensation functions; policy is never about point estimate of single object of choice
- Research — complementary sample analysis; companion samples linked to large on-going sources of data

Three CV Research Priorities

- Characterizing CV object of choice
- Treating CV/SP information as economic data linked to preferences
- Characterizing individual heterogeneity

Question and Answer Period for Session IV

Edna Loehman, Purdue University, asked for comment on the payment card method, which got a black mark from the NOAA panel. She believed the method could be useful and informative if researchers took appropriate care to deal with scaling. Incentive problems are perhaps not as important as problems with communicating the nature of the good being valued. She did a study of common morbidity effects with payment cards and got results with surprisingly good scaling and proportionality. However, she recently got involved in a study valuing highway safety and found use of the payment card method difficult. The difference was that the highway study was asking people about unfamiliar and uncertain risks of accident injuries instead of familiar and certain morbidity from headaches and colds. The psychology of such choices is little understood and needs exploration.

Richard Carson remarked that the payment card method in theory is not incentive-compatible, but its biases are well known. It gives too few small answers, and on the high end respondents tend to shift down towards where respondents think the costs are. The result is shrinkage near zero and at the high numbers. If you are prepared to accept that shrinkage, the format will give you a lot of information without much loss. Among open-ended-type formats, this format is probably the best one.

Why? Because the incentives that underlie an open-ended question pivot on costs. The old psychological literature that said willingness to pay (WTP) should be independent of cost is completely wrong. When you look at the optimal response strategy, if your WTP is below cost you should go towards zero, and if your WTP is above cost you should go towards the cost. So a bidding game question conveys a fair amount of information about cost, and an open-ended question forces the respondent to think about the costs. A payment card format actually diffuses whatever the original prior on costs was. As the prior on costs gets diffused so the person gets risk-averse, you converge from below to the true WTP number. It works reasonably well, as long as you don't get hung up on the downward bias.

Michael Hanemann was a bit less positive in his assessment. As long as there is no controversy in the results, this is an acceptable method. But if someone wanted to attack your results, he could devise a different payment card survey that would give different results. Something like this happened in Great Britain two years ago.

Glenn Harrison, University of South Carolina, followed up on Kerry Smith's earlier concerns about what valuation methods are going to be considered reliable in policymaking situations. There seem to be two settings in which reliability is going to be judged. One is in open adversarial questioning by the interested publics and the other is in litigation. Harrison asked the panel about their experience since the Exxon Valdez on the acceptance of contingent valuation method (CVM) studies by the courts and by the academic community.

Richard Carson said he was not very familiar with what has happened in litigation. However, he believed that many attacks on methods there were simply convenient ways to couch what were really attacks aimed at the bottom line. As a result, some attacks against stated preference (SP) were simply thinly veiled fights over money.

Michael Hanemann related an anecdote concerning a beach closing in Los Angeles. At trial, he testified on the lost value due to closure, based on a study of travel costs, a revealed preference (RP) method. On the witness stand, the other side's attorney said, "You claim the value of a visit to the beach is about $15. Have you ever asked anyone whether he had a consumer surplus of $15 to go to the beach?" In other words, the attorney attacked Hanemann for not having SP data. The point is, any method that relies on analysis and modeling assumptions will be vulnerable in any public policy debate. If your analysis is not transparent to the lay audience, you are open to disbelief.

David Chapman noted that attorneys are paid to attack during litigation. But to his knowledge in the damage assessment area, no CVM studies have gone to trial. That does not mean the studies have not been useful.

Richard Carson noted that in the Exxon Valdez case, the studies happened to value an actual future incident. The Coast Guard adopted a spill prevention plan almost equivalent to one favored in a study, including the use of escort ships. About three years later, a tanker out of Valdez lost power, and the escort ships prevented the tanker from running aground and towed it safely out to sea.

The top journals have been sporadic about survey use. But a 1995 review found about 2000 CV surveys in the literature, and now there are about 3600 surveys from 90 countries reported. So survey use continues to explode. A large number of the studies were in developing countries, on practical policy issues like provision of water systems and sewers or eco-tourism use of parks. Often, regardless of the quality of the survey work, the underlying engineering estimates of costs are poor.

Kerry Smith said that he was not directly involved in any litigation and so got to observe it from both sides. He noted a change in the structure of analysis that goes into litigation, with a move towards stated choice. In many cases, the models used were unable to come up with WTP.

In fact, the attorneys do not care about the correct WTP. They only want numbers to start the bargaining, to put a position in play to bound the negotiation.

On another point, he would argue that the study that Hanemann mentioned involved not primary analysis of travel cost, but benefit transfer. Most litigation does not involve primary research. If it is cheaper to get a transfer number than a primary number, the attorneys will go the cheaper route if they think the number will hold up in court.

Regarding the status of contingent valuation in academia, many economists treat contingent valuation lightly, almost condescendingly. They would not allow their graduate students to do it.

Glenn Harrison noted that economists who were critical of CV often could offer no practical alternatives to its use in a given situation. Many academic economists do not pay attention to whether their work is specifically relevant to pending policies.

Richard Carson said once you exclude macroeconomists, the international trade specialists, and most of the econometricians who are doing time series, you are left with a

much smaller group of people who might care. His department is satisfied to have just one environmental economist.

John Halstead, University of New Hampshire, said he has met economists whose attitude towards empirical research is that everyone should do it – once. Some branches of economics do not take the sub-profession of environmental economics seriously. He wondered if anyone would ever win a Nobel Prize for environmental economic work.

Kerry Smith thought that the general perception of environmental economics was not as negative as the perception of CV. Look at the composition of EPA's environmental economics advisory committee. It includes people working on auctions and other issues relevant to environmental economics who are happy to be recognized for their work. Another example is the NBER summer workshops on public economics and environmental economics, where you will find little work on SP but much on other relevant topics.

Michael Hanemann said the gap between valuation and policy analysis and design is not that large. He personally is interested in what it takes to shift behavior, which is very relevant to practical policy design. Most economists think abstractly, but what they do is still relevant to policy.

Kerry Smith noted that the Nobel citation for Dan McFadden mentioned that his work helped in the valuation of the Exxon Valdez damage.

Glenn Harrison raised the role of an agency's research incentives and funding efforts in resolving outstanding issues. He noted that some issues, such as equity issues or varying the value of statistical lives (VSLs) for children of different ages might be too controversial for an agency to take on in-house. Could such work be usefully done outside the agency? To what extent can academic work complement agency work and tackle issues the agency would like to but cannot?

Al McGartland said EPA's environmental economics advisory committee did grapple with the issue of valuing children's lives and agreed it was too sensitive an issue for the agency to specify values.

He thought that it would be good to engage the next EPA Administrator on the issue of research. Few political leaders in EPA have actively sought to be briefed on economic tools.

Maureen Cropper asked, wasn't EPA funding research on valuation of children's health? Al McGartland said yes.

Matthew Clark said EPA was probably going to redo the evaluation of children's health solicitation. He encouraged people to watch EPA's web page for details.

Richard Carson found interesting that people could understand changes in life expectancy better than they understood risks stated in terms of increased numbers of deaths per year. Couple that with some notion of how people discount risk over different time periods, and you might have one function explaining risk perception that politically might be acceptable.

Maureen Cropper noted that the whole QALY (*quality adjusted life years*) literature asks people to trade off two things: quality of life and length of life. The concept of the survival curve and its relation to life expectancy is not easy for people to grasp. To communicate the true meaning of the risk is hard.

Richard Carson said that work had progressed from studies that showed people are insensitive to increased risk to studies that showed it is possible to communicate about risk to make people's responses more proportional. This issue will not be solved soon, but if it is a priority for EPA, a concentrated, coordinated effort could yield results.

James Hammitt, Harvard University, noted that on valuation, trade-offs between an individual's own money and own risk turn on a combination of age and other things. Empirically, these life-cycle models can lead to a wide range of results. We have not nailed things down yet. People have been working on risk communication for a long time – the first risk ladder dates to the early 1980s.

With regard to using academic economists to develop work and break new ground, EPA played a major role in stimulating CV research in the first place. That initiative has been a success.

Matthew Clark asked how much agency-sponsored research should be on basic methods and how much should be on practical applications?

Richard Carson encouraged the agency to welcome speculative proposals.

Also, he believed that the agency has been too passive in filling in the gaps in the benefit transfer grid. The agency should have a systematic plan to fill in the gaps. EPA should not expect outside researchers to initiate these studies without encouragement. An academic researcher's inclination is to design work that can be published, not necessarily novel work that satisfies practical needs of policymakers.

He concurred with the idea of setting up a standing agency peer review panel to encourage a consistent high level of research.

Kerry Smith proposed the EPA assemble a panel of four to six senior economists to operate under a model pioneered by the Russell Sage Foundation. The panel would meet annually to award, say, one million dollars for SP research. They could not fund their own work. They would critique the funded research once a year for three years. One person on the panel would be from EPA, and one could be from another agency, but the majority would be from outside the government.

Kelly Brown, EPA, commenting on Richard Carson's suggestion to fill in the gaps, noted that it can be hard to get academics to pursue work they cannot easily publish. And while EPA might like to do the work within the agency, it is difficult for the agency to get the legally required approvals to do surveys. Making SP more respectable among academics would help.

Glenn Harrison suggested institutional support would be a good step in that direction. Perhaps EPA could fund two or three centers in the United States to focus on SP research and teaching.

Richard Carson agreed, noting that students could support project contracts. An institution could use graduate students to pursue projects with a high level of peer review in a cost-effective manner.

Carol Jones noted that contracts were not grants and wondered if academics would be interested in working under contractual constraints. She also noted that graduate student work would have to be carefully supervised to be credible in policymaking or litigation.

Richard Carson thought that since the contracts would be with the professors, not the students, there would be little problem with quality control or continuity of research.

Kerry Smith said his proposal would use grants, not contracts. The topics would be under control of a group with built-in peer review, not EPA.

On the issue of how EPA can get the work done that it needs, Smith noted that EPA and USDA Economic Research Service (ERS) are already collecting data such as the 1994 national recreation survey of water-based sites. How much work has actually been done with that survey? Another survey is about to go into the field. Why not link small-scale CV studies to these large RP surveys?

Daniel Hellerstein, USDA ERS, said it was hard enough to get an RP survey approved through the federal bureaucracy. To try to get a linked CV study approved adds an additional hurdle. Kelly Brown concurred. Kerry Smith asked if the linked studies could be done by outside investigators through grants. Glenn Harrison said there is a vehicle for doing that. David Chapman agreed that if EPA does not have certain control over the process of data collection, he understood that the survey would not require OMB approval. But if EPA is actively involved in the study or is planning to use the results in a particular way, the survey needs approval.

Glenn Harrison saw potential for piggybacking studies on the large government data collection project, taking advantage of the large data set to magnify the usefulness of the small studies.

David Chapman said the National Marine Fisheries Service piggybacks small surveys on the broader data collection they do annually or every few months.

Michael Hanemann said he was doing a large RP survey in Los Angeles, hanging small CV modules on the RP work. We need basic research on survey modes. This can be hard to fund, but it would be a public good if it leads to money-saving techniques. Perhaps a consortium of funders, including the National Science Foundation, could fund such basic CV work.

Hanemann also seconded Smith's idea of a program of research funded through a peer review panel.

Richard Carson said his idea of filling in the benefit transfer grid also needs a systematic agency funding mechanism and is not a substitute for Smith's proposal. Smith's proposal would address deeper, more fundamental issues; Carson's would address more practical issues. The agency needs to address both, and to do so well, in a coordinated way that creates an increasing, integrated base of knowledge.

Carol Jones said EPA should use grants to support research on fundamental issues and contracts to direct work on specific practical areas.

Matthew Clark said NOAA, EPA, and USDA could be looking at jointly supporting research on fundamental issues.

Richard Carson urged agency people to get together to identify what their common needs are and where they have run into common problems, limiting information, or bad studies. If they define those problems, that may identify areas of research that could generate publishable studies.

Matthew Clark concluded the session with special thanks to the panelists, the researchers, the organizers at EPA's National Center for Environmental Economics, and the cooperating agencies including USDA ERS.