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

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).

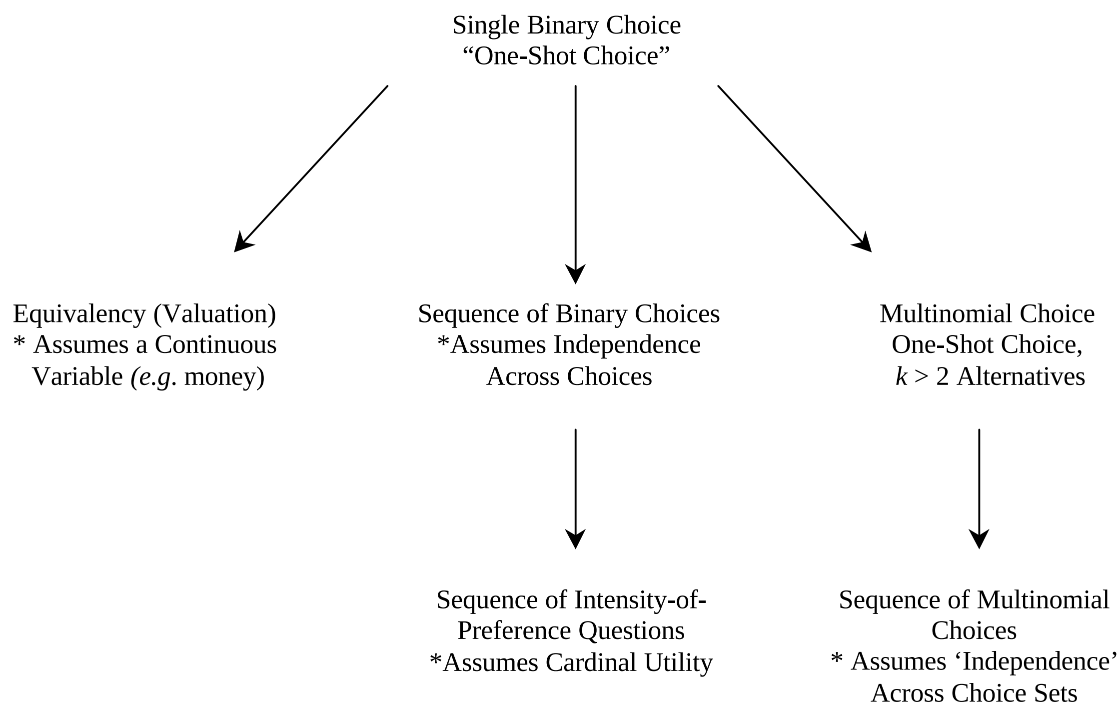
[7] For example, if a subset of agents does not believe that the good can be provided in any amount, these agents should be insensitive to the quantity (scope) of the good to be provided even though they may place a positive value on it. Divergences between the intended and answered question will always 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 public 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, $b_1 z_{1i}^q + ... + b_K z_{Ki}^q$,[2] and an additive random component, e_i^q, that follows an extreme value distribution, the probability that consumer *i* prefers choice 1 over alternative 0 is:

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

where:

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

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:

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

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:

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

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:

$$M\hat{R}S_{ml} = \frac{\hat{b}_l^q}{\hat{b}_m^q} \tag{4}$$

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, θ_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:

$$W\hat{T}P = \frac{-\hat{a}}{\hat{b}} \tag{5}$$

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:

$$y_i = a + bx_i + \boldsymbol{e}_i \tag{6}$$

with $\boldsymbol{e}_i$ 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:

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

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 ... 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:

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

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:

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

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 $w_i = P_i(1-P_i)$. 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:

$$|I| = \sum_{i=1}^{N}\sum_{j=1}^{N}\cdots\sum_{z=1}^{N} p_i p_j \ldots p_z (1-p\ \ )(1-p_j)\ldots(1-p_z)\left|X^*_{i,j,\ldots,z}\right|^2 \tag{18}$$

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 $q_{1i} = b_1 x_{1i} + b_2 x_{2i} + \ldots + b_K x_{Ki}$, $q_{ki} = x_{ki}$ for $k = \{2,\ldots,K\}$ 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 $P_i(1\text{-}P_i)$ 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:

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

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:

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

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:

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

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 $g(B) = b_2 / b_1$, the first order conditions reduce further to:

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

or,

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

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:

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

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:

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

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

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:

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

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

Observation	Alternative	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?