Contingent Valuation: Controversies and Evidence

Abstract

Contingent valuation (CV) has become one of the most widely used non-market valuation techniques. CV's prominence is due to its flexibility and ability to estimate total value, including passive use value. Its use and the inclusion of passive use value in benefit-cost analyses and environmental litigation are the subject of a contentious debate. This paper discusses key areas of the debate over CV and the validity of passive use value. We conclude that many of the alleged problems with CV can be resolved by careful study design and implementation. We further conclude that claims that empirical CV findings are theoretically inconsistent are not generally supported by the literature. The debate over CV, however, has clarified several key issues related to nonmarket valuation and can provide useful guidance both to CV practitioners and the users of CV results.

Key words: Contingent valuation, Passive Use Value, Willingness to pay, Welfare economic criteria.

1. Introduction

To fully assess the economic desirability of environmental policies, analysts must estimate the value of non-market commodities. Overlooking or ignoring the services provided by non-market commodities in cost-benefit analyses and other empirical economic studies severely undermine the accuracy and relevance of the results. Since the 1960's, several non-market valuation techniques have been developed in recognition of the importance of these services. Of these technique, the most commonly used is contingent valuation (CV). Its flexibility facilitates valuation of a wide variety of non-market goods, including those not currently provided. Perhaps more importantly, CV enables researchers to assess total value that includes passive use value.

In many instances, the magnitude of passive use value may be substantial. However, it has often gone unmeasured. Inclusion of passive use value potentially increases the stakes in natural resource damage assessments and may tip the scales in favor of preserving natural resources over development in environmental goods, debate persists over the reliability of CV and the overall suitability of passive use value in economic policy analysis. Consequently, several industry groups have voiced their opposition by lobbying against the use of CV and by sponsoring research aimed at investigating CV's reliability and the nature of passive use value. In the course of this debate, many theoretical and empirical issues have been raised. As a consequence, there has been a fundamental rethinking of many of the key issues related to assessing the benefits of environmental amenities. What emerges is a much richer theoretical framework for non-market valuation and a variety of approaches for assessing the quality of CV results. Within this context, the purpose of this paper is to provide a concise overview of some of the most commonly alleged weaknesses of CV and passive use value and to offer counter-arguments.

The paper is organized into seven sections. Section 11 addresses welfare economic issues associated with CV and passive use. Section 111 identifies key elements in the design, implementation, and analysis of CV surveys and their relevance to CV criticisms. Section 1V discusses empirical results from CV studies and their consistency with economic theory. Section V discusses strategic behavior and its relevance to CV elicitation formats.

Section V1 discusses issues related to the validity and reliability of CV estimates. Section V11 offers concluding remarks.

11. Welfare Economic Issues

The link between welfare economics and CV is quite direct: CV offers the potential to trace out the willingness to pay distribution for a population of economic agents for a proposed change in a good. If properly executed, CV is a useful tool for benefit-cost analysis. CV can also be used for other purposes where knowledge of the willingness to pay distribution and its relationship to other variables, e.g., income and geographic location, is of interest, Policy makers often consider distributional and political criteria in addition to welfare economic criteria.

Welfare economics, through benefits-cost analysis, seeks to reveal whether the potential change in utility resulting from a change in an economic variable, such as a change in a commodity's price or the level of provision, is positive (Just, Hueth, and Schmidt, 1982). The welfare implications are often expressed in terms of a change in an index, usually the monetary amount which would need to be taken from or give to the agent to keep the agent's overall level of utility constant. At the level of an individual economic agent, these monetary measures take a particularly simple from: for a desired increase in the good, the maximum amount the agent would be willing to pay to obtain the improvement, and for a decrease, the minimum amount the agent would be voluntarily willing to accept in compensation in exchange for accepting the decrease. Whether willingness to pay (WTP) or willingness to accept (WTA) is the appropriate measure depends upon the relevant property right to the good. A number of different proposal for how to aggregate the monetized measures obtained from agent have been advanced (Mueller, 1989).

Total Value

The term total value, synonymous with true WTP or WTA, arose in environmental economics with the awareness that sometimes portions of WTP or WTA were not accounted for in the measure of economic value obtained using market prices or revealed preference techniques. In this sense, WTP and WTA estimates derived using those approaches are defective since welfare gains or losses may be overlooked if passive use value is decisive.

While a variety of distinctions have been proposed for our purposes here, it is useful to decompose total value into direct use value and passive use value (Carson, Flores, and Mitchell, 1999). Direct use can be most easily thought of as requiring the agent to physically experience the commodity in some fashion. Since an agent need not direct use a good to get utility from it, any uses not requiring this direct contact are often deemed passive uses. Passive use values can be traced back to seminal observation that many people valued natural wonders simply for their existence (Krutilla, 1967). Their argued that these people obtain utility through vicarious enjoyment of these areas and, as a result, have a positive WTP for the government to exercise good stewardship of the land. Thus, an agent can have both direct and passive use values for a good.

Passive Use Value

Three camps hold fundamentally different positions on passive use value. They are: (1) passive use value are irrelevant to decision making, (2) passive use value cannot be

monetized, and thus, can only be taken account of as a political matter or by having experts decide, and (3) passive use value can be reliably measured and should explicitly be taken into account.

Consider a policy choice between making an area into a wildlife sanctuary for endangered species that would not be open to visitation by the public and leasing the area for coal strip mining. The benefits of the sanctuary would, therefore, derive from passive use. Those in the first camp believe that passive use value should not be considered in determining the benefits and costs of the two potions, and therefore, explicitly assign a zero value to the establishment of the wildlife sanctuary. This camp also frequently takes the position that a program should pass a benefits-cost test under which passive uses are not considered. The view held by the second camp is that the benefits of the wildlife sanctuary should be taken into account, but that those benefits cannot be monetized. Therefore, the political arena or "expert judgment" is the appropriate place to taken those benefits into account. This position is also often taken by those who argue that benefit-cost analysis should not be used as a major criteria in environmental decision making and by those who argue that benefit-cost analysis should be a simple accounting exercise whereby easily observed quantities are multiplied by correspondingly easily observed market prices and placed on the correct side of a balance sheet. The third camp believes that wildlife sanctuary benefits can be reliably measured in monetary terms and usually argues that those benefits should be taken into account in the economic assessment of the policy choice. Because CV is the most common approach for obtaining estimates of economic value when passive uses are to be included, the decision to use it often turns on beliefs about the reliability of CV as a measurement tool.

Relationship Between Contingent Valuation and Passive Use Value

CV surveys measure the total value of the described good while revealed preference techniques, which are based on observed behavior in private markets related to the environmental good, measure only direct use value. Revealed preference techniques are usually only capable of capturing the quasi-public value, which is the direct use portion of total value, because they rely on the availability of an implicit private market for a characteristic of the good in question. The availability of this market allows for potential excludability based on price. In contrast, passive use value can be seen as simply a special case of a pure public good.

Is CV the only technique capable of capturing passive use value? The answer is no. The fundamental problem in the economic valuation of environmental goods is the absence of a market for their direct provision. Any of the other members of the class of constructed markets, such as an actual referendum on whether to provide the public good, or a simulated market in which the good is actually provided, can potentially be used for this purpose. The value of a public good may also be inferred in some instances from voting decisions by political representatives (Carson, 1991).

Economists have a strong bias in favor of estimates that are inferred from observed actions, the revealed preference paradigm, as opposed to stated preferences, such as those revealed in CV surveys, although this bias is not shared by the other social sciences. Periodically, major figures in the economics profession note the limitations that this

reluctance to query people directly about their preferences imposes on the questions which economists can address (Blinder, 1991).

Unfortunately, these limitations are perhaps the most severe when dealing with environmental goods, as the government directly provides a number of the most important of these goods and provides many more indirectly by using regulation to set their levels. Observable behavior is at best often only tangentially related to the use of such goods, whereas data on their use are only sporadically recorded, if at all. With the growing concern over the environmental during the last thirty years, it become obvious that many commodities for which there exists no direct behavioral link are nonetheless valued by segments of the public. CV emerged as the major way of getting around this informational impasse, but at the expense of departing from the revealed preference paradigm that is favored by many economists.

Passive Use: Altruism and Other Motives

One of the frequent attacks on passive use value is that it is motivated by a form of altruism termed "moral satisfaction" or "warm glow", and hence, passive use is not an economic value (Kahneman and Knestch, 1992). At some level, this argument harkens back to an age-old source of conflict between economists and non-economists often try to pigeonhole economists as being irrelevant by alleging that economic theories are based on a very narrowly-drawn definition of self-interest. The answer from economic theory is very clear, it is utility whatever its source that matters for total value. Motives are essentially irrelevant from the perspective of economic theory (although policy makers may care) and acceptance of consumer sovereignty is one of the most enshrined principals of economics.

One place that economists have considered motives is in the literature on charitable giving. Motives have been examined in that context because in the simple version of the free-rider principle one would not expect to see contributions to charity in many instances where they do in fact occur. The classic motive for charitable contributions is pure altruism and arises from the simple desire on the part of an agent to increase the level of provision of a particular public good. A second motivation, impure altruism, first identified by Olsen (1965) and later termed "warm glow" by Becker (1974), recognizes that individuals also derive utility from the act of giving through the associated social approbation, prestige, or moral satisfaction. Neither pure nor impure altruism was seen as a non-economic motivation by these authors. Andreoni (1989) has used the concept to show why theoretically progressive taxation can actually lead to increases in charitable giving and why government contributions to charities do not crowd-out private giving if warm glow makes government subsidies imperfect substitutes for private contributions. However, warm glow in the sense of Andreoni is largely irrelevant to a CV discussion, unless respondents get utility simple from paying more taxes to the government for the good, irrespective of whether it is actually provided.

There is one case where warm glow in the context of CV surveys might be troublesome. This case, which can be seen as a form of interviewer bias, occurs if respondents in a CV survey attempt to please an interviewer by agreeing (or not agreeing) to pay some amount when they would not do so otherwise, except for the utility gain associated with telling the interviewer. This effect should be avoidable with well-trained, neutral interviewers.

The hypothesis was recently tested using interviewers from the University of Chicago's National Opinion Center and the Exxon Valdez Oil Spill CV survey instrument. That test used a split-sample design: half the respondents were asked the valuation question in the standard way and the other half of the respondents were asked to write down their responses on a sheet of paper, seal it in an envelope and place it in a locked ballot box so that the interviewer did not know their answer. No significant difference in the WTP estimates was found (Carson, 1997).

Familiarity

A frequent clam is that familiarity with a good is a necessary prerequisite to providing "meaningful" responses to CV valuation questions. The rationale given for needing familiarity is the assertion that respondents cannot have well-defined preferences in an economic sense for goods with which they have no direct experience (Desvousges, 1993).

This rational, however, relies upon a set of questionable assumptions concerning how people make purchase decisions, assumptions that in effect would rule out making inferences about the utility people get from making most non-routine market purchases. Personal experience or familiarity is only one factor in the decision-making process. Consumers make use of related experiences, information from reviews, advertising, and so forth. Many new products become available each year creating completely new markets in which consumers regularly make purchase decisions involving goods for which they have no prior experience. No standard microeconomic text has ever stated that prior experience is a precondition to rational decision making.

Second, commodities in most CV studies are typically valued at somewhere between $ 5 and $ 250 per household. Between 20 and 40 minutes are usually spent familiarizing the respondents with the commodity in question. It is doubtful that consumers of equivalently priced private goods, such as a meal in a new restaurant, a hardcover novel, a weekend at a nearby resort hotel, or a family evening at the movies, spend significantly more time familiarizing themselves about these purchases than respondents to a CV survey spend learning about the commodity at issue (Carson, 1997).

Third, to deny that people have meaningful preferences about new commodities, political issues, cultural questions, and the like, without having had prior personal experience with then would be tantamount to suggesting that only those individuals who had actually visited the Louvre can value the preservation of its art works and that all votes for non-incumbent politicians should be disregarded. These simple example illustrate that specific personal experience is not required for making meaningful economic choices.

For the results of a CV study to be credible to policy makers, CV survey designers need to ensure that prospective consumers understand what they are being asked to value, how it will be provided, and how it will be paid for. For the CV respondent, this means that the wording of the questionnaire must successfully convey the nature of the good and the context in which it can be purchased in a way that is plausible, understandable, and meaningful to respondents who have widely varying life experiences and educational backgrounds (Mitchell and Carson, 1989).

Market Size

The question of the appropriate market can be answered on either legal or empirical grounds. As an example of the former, assume that a state is considering raising state property taxes to buy a piece of land for a state park. Because of the payment mechanism, state policy makers would only be interested in comparing the values of state residents to the state tax payments that would be required. Thus, even if some residents of another state value the park, the state providing the park can choose not to "care" about their values. In this case, the population that should be surveyed is that of the state which is considering providing the park.

For the latter, the researcher must determine empirically the population that values the park. In principle, it is possible to define a population broadly enough so as to be assured of sampling from all agents that hold non-zero values. Studies looking at this issue tend to find that individual total value estimates decline as geographic proximity to the good decreases. For goods with primarily passive use value, population subgroups that have a particular concern for class of resource are likely to be a more dominant factor than distance (Bateman and Langford, 1997).

For some CV critics, the possibility of obtaining a very large estimate by aggregating the small WTP amounts of individuals in a very large market is itself a fundamental problem. These critics have failed to grasp that a public good's value is the summation of the value of individual agents who can enjoy it (Samuelson, 1954). It is this very characteristic which accounts for the presence of a wider array of public goods in large cities and countries. Aggregation across agents has nothing to do with CV per se; it is merely part of the definition of the value a public good.

111. Survey Design, Administration and Analysis Issues

Contingent valuation is a basic approach to non-market valuation rather than a single specific economic valuation technique. Common to most applications of CV surveys are: (1) an introductory section which helps set the general context for the decision to be made; (2) a detailed description of the good to be offered to the respondent; (3) the institutional setting in which the good will be provided; (4) the manner in which the good will be paid for; (5) a method by which the survey elicits the respondent's preferences with respect to the good; (6) debriefing questions about why respondents answered certain questions the way that they did; and (7) the collection of a set of respondent characteristics including attitudes, debriefing questions, and demographic information. To a large degree, the variations among CV surveys involve different techniques for eliciting respondents' preferences concerning the good of interest. The issue of elicitation formats is taken up some length in Section V.

The WTP estimates obtained from CV surveys are generally sensitive to other key features of the constructed market. Although this has been a source of concern to CV's critics, it should more accurately be viewed as a strength of CV because it allows one to use CV to examine the influences of key factors related to how the good is provided. People have distinct preferences over the exact manner in which they pay for goods and perceive different methods of providing a good to have different likelihoods of success. In this sense, the term "contingent" valuation is apt and one should never forget that it is only the plan to provide the good that can valued, not the good in the abstract. This issue is usually ignored in working with revealed preference data become the context in which

the observed choices are made is usually out of the researcher's control. This, of course, can create problems in trying to extrapolate the results from the situation in which the data were obtained to a different situation with different contextual elements. CV has no such limitation. Researchers must recognize the fact that economic value for a good cannot exist in the abstract independent of the terms of its provision.

The designer of a CV study must make the good to be value understandable to the population of interest. The designer needs to convey enough detailed information to convince respondents that the supplies will be able to deliver the good. The payment mechanism must be credible to respondents so that they believe that they actually could have to pay for the good. Finally, to ensure respondents provide thoughtful responses to the questions, they need to be told that the information they are providing will be used in the decision-making process. An implausible scenario in a CV survey is an invitation to respondents to treat the exercise lightly and further the survey scenario should convey to the respondent that the plan to provide the good has been well thought out. All of this must be done without overwhelming respondents with large amounts of information that they find extraneous to the choice being offered.

The most commonly used modes of administration for CV survey are in-person interviews and mail surveys. The former generally reduces the likelihood of sample selection bias because respondents who return mail surveys tent to have more interest in the good than non-respondents. In-person interviews do not exclude people with reading difficulties as do mail surveys. They also provide more control over the order and manner in which survey material, including visual aids, such as maps and photographs, is presented. Mail surveys are generally much less expensive. Many standard survey research texts contain substantive discussions of the major issues involved in the choice of the mode of survey administration (Tourangeau, Rips and Rasinski, 2000).

CV studies with questionable results usually fail along predictable lines. In such studies, the good, the provision mechanism and/or the payment obligation are vaguely described. These problems in the survey instrument are usually compounded by a survey administration mode such as a short telephone interview or self-administered questionnaire that does not encourage the respondent to take the CV survey seriously. Poor quality sampling, unlikely to be representative of the population of interest, is a frequent accompaniment. All of these problems can be observed before looking at the empirical results. With respect to the CV WTP estimates, one needs to ask whether the estimates are largely driven by a very small number of outliers and whether they are highly sensitive to any distributional assumptions made (Huang and Smith, 1998). Further, one should expect to see the estimates at the respondent level vary systematically in sensible ways, as discussed below.

1V. Consistency With Economic Theory

Do CV results conform with the predictions of economic theory? There are two obvious tests. First, the percentage of respondents willing to pay a particular price should fall as the price they are asked to pay increases. This condition, similar to a negative own-price elasticity for a marketed good, is almost universally observed in CV studies. Second, respondents should be willing to pay more for a larger amount of a desired good. This test, often referred to as a scope, test, involves observing changes in the WTP estimate as

the quantity or quality of the good is made larger or smaller. This is one of the most debated points concerning the validity of CV. Critics have argued that the apparent lack of sensitivity of CV estimates to differences in scope is the most serious empirical problem with its use, as assertion that is now routinely repeated in introductory texts on benefit-cost analysis and environmental economics (Goodstein, 1995). We devote a subsection below to this topic.

The price and scope tests have the advantage of being simple unidirectional hypothesis tests with very close ties to the underlying economic theory. These tests correspond well with economic intuition. One might also make conjectures about the relationship between respondent income and WTP, on the difference between estimate of WTP and WTA, on the effect of the order in a sequence in which a good is valued, or on the effect of aggregating independently derived WTP values for different goods. Tests of these phenomena are context specific and require judgments about relative magnitudes. Here we show that the usual economic intuition developed from observing how the quantity of a private good varies with price changes is often faulty when it comes to making inferences about what properties WTP for a public good should have. The fundamental insight is that one needs to think of a public good as a special case of a quantity rationed good (Hanemann, 1995).

A scope test looks at whether respondents are willing to pay more for a good that is larger in scope, either in a quality or quantity sense. It is important to recognize that failure to pass a scope test can be attributed to one of three factors: (1) lack of the statistical power used to detect the difference in value which would be plausible given the difference in scope, (2) problems in CV survey design and administration which tend to mask sensitivity to scope, or (3) CV survey results that violate economic theory (Hausman, 1993). The debate that has taken place in the environmental economics literature has been whether insensitivity to the scope of the good being value is a ubiquitous phenomenon or whether this phenomenon occurs only occasionally and, in such instances, is the problem traceable to a lack of statistical power or problems with the design or implementation of the special survey?

A test of responsiveness to scope can be implemented either internally. In an internal scope test, the same respondents are asked to value different levels of the good. External scope tests rely upon asking two different, but statistically equivalent, subsamples about two different levels of the good. With internal scope tests, the null hypothesis that respondents give the same WTP amount, irrespective of the level of the good they are asked about, has long been almost uniformly rejected. CV critics have argued strongly that respondents may simple be trying to be "internally consistent" in their answers. Recent attention has focused on external tests of scope and, in particular, the evidence presented by Kahneman and at the Exxon symposium, suggesting that respondents to CV surveys do not give different values to goods that differ in scope. Carson has recently conducted a comprehensive review of the empirical CV evidence from split sample tests in which one subsample was offered a good that was of larger scope than that offered another equivalent subsample. An important aspect of this review is, contrary to claims made by Kahneman and Hausman concerning the absence of studies other than the few they consider, there have been a number of studies containing an external scope test. Most of these split-sample tests were done in CV studies originally designed for policy purpose

where two or more different levels of a good were of interest to policy makers. These studies have advantages over the work of Kahneman and those reported in Hausman in that: (1) the goods being valued were usually the subject of real policy choices, (2) they generally enjoyed a more extensive survey design and pretesting effort, and (3) they tended to use more appropriate modes of survey administration and larger sample sizes. Almost two-thirds of the studies dealt with situation where passive use considerations were thought to predominate, while the rest dealt with situations where direct use was thought to predominate.

The Carson review Shows that, since 1984 (the data of Kahneman's original claim that CV results are insensitive to scope), 31 studies reject the scope insensitivity hypothesis while four do not. Two recent large CV studies for government agencies using in-person interviews and well-constructed questionnaires containing extensive visual aids depicting the good to be valued, rejected the scope insensitivity hypothesis at $p < 0.001$. For recent meta-analyses that looked at studies valuing outdoor recreation, air quality change, groundwater contamination also rejected the scope insensitivity hypothesis by showing that the CV estimate from different studies vary in a systematic fashion with differences in specific characteristics of the good (Poe, Boyle, and Bergstrom, 2000).

Poorly executed survey design and administration procedures appear to be a primary cause of problems studies not exhibiting sensitivity to scope. None of the commonly cited studies with scope insensitivity bears much resemblance to the current state-of-the-art CV survey where respondents are presented with a substantial amount of information about the good they are asked to value in a manner which facilitates their comprehension of the material. The Ksheneman and Knetsch, work used short telephone surveys with vaguely defined goods, provision mechanisms, and payment obligations. Desvousges et al.'s study of covering oil ponds to prevent birds from being killed in the Rocky Mountain area was a short self- administered survey done in a North Carolina shopping mall.

In other instances, original claims of scope insensitivity do not stand up to the use of simple but more powerful statistical tests. For instance, Diamond et al., looking at WTP for different size wilderness areas in the Rocky Mountain States, claim a p-value of 0.42 for a test of their null Hypothesis 1 using a Kruskal-Wallis test. Carson and Flores show that their statistical test has no power to detect large differences, and instead, estimate a simple OLS regression of WTP on the number of areas in each of the three wilderness areas. They reject Diamond et al.'s null hypothesis at $p=0.01$.

At this point we believe that out of sample scope tests, to the extent that they divert resources from survey design efforts and sample size, are probably not a good investment. Further, there is probably more risk to disbelieving a pair of CV results because they do not show much sensitivity to the scope of the good being valued than the opposite reaction. For many environmental goods, the public may have sharply declining marginal utility for an environmental amenity after a reasonable amount of it has been provided. This is important information to know for policy purposes. There is, however, one key area of concern with respect to scope sensitivity and the use of CV and that is in valuing changes in small probabilities of health risk. The inherent problem here is that people are known to have substantial difficulties understanding and dealing with low-level risks. As such, the risk communication problem must be solved first before the Hammitt and Graharn look at several different risk communication devices in the context of a CV

survey. They fine that almost no sensitivity to the scope of the good being valued with a simple verbal description of the risk changes. Yet with one of their visual methods of presenting the risk change, they find significant scope effects with WTP for risk reductions being almost linearly increasing in the magnitude of the risk reduction. A different approach is taken by Carthy et al. who attempt to break the problem into two pieces, one involving value elicitation and the other involving standard gambles, chained together to get arrive at values for small probabilities. The valuation of risk reductions is likely to remain an active research area for some time (Carson and Mitchell, 2000).

Income Effects

Drawing inference about economic values from intuition regarding the demand for private goods, one expects to see a positive relationship between income and WTP if the good being valued is a "normal" good. A frequently made claim, for which there is surprisingly little empirical support, is that most environmental goods are "luxury" goods. If this were the case, one would expect the income elasticity to be greater than one. The usual empirical result from CV studies is to find a positive income elasticity of WTP substantially less than one for environmental commodities.

This typical empirical result has been cited as evidence that contingent values are theoretically deficient. For instance, McFadden reporting on one of Exxon's studies notes. The problem is that the terms necessary and luxury are defined in terms of the income elasticities of demand; a measure based on varying, not in terms of the income elasticities of WTP, a measure based upon holding the quantity fixed. Flores and Carson show that the two types of income elasticities are fundamentally different. The income elasticity of demand shows how the quantity demanded increases as income increases while the income elasticity of WTP look at how WTP for a fixed quantity of the good changes as income increases. The two income elasticities can be shown to be functionally related using the concept of a shadow or virtual price that responds to changes in the levels of rationed goods. Flores and Carson's results show that for any fixed value of the income elasticity of demand, the income elasticity of WTP can differ significantly in magnitude and even sign. Thus, a good which is a luxury good in a demand sense may have a WTP income elasticity which is less than zero, between zero and one, or greater that one. If the matrix of cross-price elasticities is an identity matrix, the virtual price income elasticity is equal to the ordinary income elasticity of demand multiplied by a scale factor (the ratio of income to income plus the monetized value of all public good), which must be less than one and probably substantially less. Thus, the income elasticity of WTP is likely to be less than the corresponding income elasticity of demand.

Divergence Between WTP and WTA Estimates

If total value in an economic sense can be expressed in terms of WTP and WTA and the two measures differ substantially either theoretically or empirically, the appropriate measure for a benefit-cost analysis depends upon the property right. From a theoretical perspective, WTP and WTA should be quite close together for a price change in perfectly competitive private markets (Willing, 1976). However, for imposed quantity changes where the consumer is not free to trade to the desired quantity level, WTP and WTA may be far apart (Hanemann, 1995). Changes in environmental goods to fail into the category of imposed quantity changes. The difference between the Willig and Hanemann

theoretical results is that for a price change, an income effect alone governs the difference between WTP and WTA, and for a quantity change, both an income effect and a substitution effect together govern the difference. One of the earliest findings from CV studies was that WTP and WTA measures differed substantially. Based upon Willig, the working hypothesis was that either one or both of the CV estimates were wrong or that the theory was wrong. Work proceeded in several directions. The first direction was to show that large differences between WTP and WTA estimates were not an artifact of the survey context. Consistently large differences were found in a variety of setting using actual transactions. Even financial assets such as junk bonds and over the counter stocks, when thinly traded, often show much large bid (WTP)-ask (WTA) spreads than would be predicted by Willig's work. The second direction was to show that the WTA question format had a number of shortcomings, both from the perspective of its strategic incentives and of getting respondents to accept it as a legitimate framework for a policy choice. The third direction was to suggest new theories outside the neoclassical framework and to show that within that framework, the theory being applied failed to capture key aspects of the situation. Much of the problem with the current framework may stem from its inherent static nature. Recent models that incorporate bargaining information effects, transactions cost/experience, and uncertainty show considerable promise in being able to explain the magnitude of the divergence between WTP and WTA amounts. The key implication of this divergence for applied policy work is that property right can have a substantial influence on the magnitude of the welfare measure. Particularly when considering a reduction in an environmental service, the common practice of substituting a WTP estimate for the desired WTA measure can result in a substantial underestimate, which in turn can have substantial policy implications (Knetsch, 1990).

Sequence and Context Effects

We now turn to the relationship between CV estimates for multiple, possibly unrelated goods. Here, the context in which the CV exercise takes place is crucial. Two issues have received the most attention. The first involves the implications of adding together CV WTP estimates for different goods. The second involves the influence exerted on the estimated value of the good by the order in which it is valued as part of a sequence of goods. The two typical empirical finings turn on the same underlying theoretical issue: substitution and income effects.

The first finding indicates that adding up what people say are willing to pay for specific goods, each valued independently as the only change to the status quo (or equivalently valued first in a sequence), might easily exceed the income of some people. This strikes many non-technically oriented CV critics as conclusive proof that CV estimates, if not complete nonsense, are gross over estimates. However, Hoehn and Randall show theoretically why adding together independently derived WTP estimates for goods is likely to overstate the value of the set of goods taken as a package, and often grossly so. At an intuitive level the reason is simple: each new public good the agent obtains reduces the agent's available income to spend on private goods. Further, if the public goods are substitutes for each other, then each one added to the package looks less desirable than when valued as if it were the only new addition to the stock of public goods. The problem should not be seen as residing with the original CV estimates, but with the analyst's incorrectly aggregating them without taking into account income and substitution effects.

The second typical empirical finding is that the value of a good falls, often precipitously; the later it is valued in a sequence of goods. Consider a stylized example reminiscent of some of the early work on air pollution valuation. A subsample of respondents in Chicago are willing to pay $100 for a specified air quality change in Chicago; and, when offered an additional specified air quality improvement in the Grand Canyon, they are willing to pay $30 more. A different subsample of respondents for whom the sequence is reversed are willing to pay $60 for the Grand Canyon improvement and $70 for the Chicago improvement. Such a result may be disturbing to the policy maker who expects a good to have only one "true" value.

The standard economic explanation for this phenomenon is and income effects. Hanemann show that if one assumes that the goods being valued are normal goods and substitutes for each other, the value of a particular public good should be progressively smaller the later in a WTP sequence it is valued. An implication of this result is that the package of goods should be valued less than the sum of its independently valued constituents. The opposite effect occurs in a WTA sequence; the later in a sequence the good is valued, the more highly it is valued. Furthermore, the usual weak assumptions made concerning the curvature properties of utility functions effectively rule out the existence of a single context independent value for a particular public good.

CV critics counter that the sequence effects observed are too large because they contend the income effects should be small and goods such as those in the air quality example above are not close substitutes. However, the CV critics' arguments about the likely magnitude of income and substitution effects are faulty because they are based on intuition derived from looking at price changes for private goods. Public goods are a special case of quantity rationed goods and, as a result, the focus should be on quantity space with an inverse demand system rather than price space with an ordinary demand system where consumers are free to chose their optimal consumption levels. Flores shows the set of virtual price substitution elasticities that should lie behind the magnitude of any sequence effects is the inverse of the set of cross-price elasticities of demand upon which the CV critics' intuition appears to be based.

Consider the following set of compensated, cross-price elasticities of demand (S_{ij}^{d}) taken from Deaton's well-known analysis of consumer demand in the UK. Good one is food and good two is clothing:

$$\begin{bmatrix} \sigma^{d}_{11} & \sigma^{d}_{12} \\ \sigma^{d}_{21} & \sigma^{d}_{22} \end{bmatrix} = \begin{bmatrix} -0.28 & 0.08 \\ 0.21 & -0.17 \end{bmatrix}$$

Note that own-price (-0.28 for food and -0.17 for clothing) and cross-price elasticities (0.08 for the effect on food demand of a price increase in clothing and 0.21 for the effect in clothing demand of a price increase in food) in this example are all quite small. Thus, with respect to either good, the percentage change in demand will be small relative to the percentage change in either own price or the other good's price. Hence, particularly large context effects for price change would not be expected. However, if one restricts choices, as is the case with environmental goods where the levels are usually collectively decided, a regime of partial rationing is in effect.

Rationing requires consideration of the inverse relationship-how the shadow or virtual price for the rational goods (food and clothing) respond to change in the rationed levels of both of these goods. These measure of responsiveness, the virtual price substitution

elasticities (S_{ij}^{v}), are related inversely, ad a system, to the compensated price elasticities (Flores, 1995). For the food and clothing example, the virtual price matrix of substitution terms is:

$$\begin{bmatrix} \sigma^{v}_{11} & \sigma^{v}_{12} \\ \sigma^{v}_{21} & \sigma^{v}_{22} \end{bmatrix} = \begin{bmatrix} \sigma^{d}_{11} & \sigma^{d}_{12} \\ \sigma^{d}_{21} & \sigma^{d}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} -5.60 & -2.55 \\ -7.19 & -9.33 \end{bmatrix}$$

The same demand system cross-price elasticities which implied fairly small increases in demand of one good when the price of the other good increases (an 8% increase in food demand accompanying a 100% price increase in clothing and a 21% increase in clothing demand accompanying a 100% price increase in food), now implies very large reductions (255% and 719%, respectively) in WTP if a unit of the other good has already been provided first in the WTP sequence. This example with private goods shows that one need not resort to explanations of inconsistent preferences or goods with peculiar characteristics to predict quite large context effects with respect to public good values.

Substitution effects are sufficient to drive the sequence effects observed in CV studies. Income effects, however, are likely to play a role as well. CV critics argue that since respondent WTP is usually just a small fraction of income, income effects should be small. Much of a household's income is already committed so that available discretionary income is much smaller, particularly if payment is required over a short time period. Further, income is known to be poorly measured in general population surveys. These sources of measurement error probably bias estimated income effects downward (Carson and Jeon, 2000).

CV critics such as Kahenman and Kneetsch respond that if sequence effects are indeed large, then CV estimates are arbitrary because they can be manipulated by the choice of the sequence order. Theirs's statement is applicable to economic analysis in general, which if done correctly is context specific. Value in an economic sense is always a relative rather than absolute concept. Even more to the point is Flores' demonstration of a formal equivalence between the agenda control problem and WTP sequences for a set of public goods. As agenda control is a central issue in public choice, it would have been surprising to see how the use of CV somehow avoided it (Mueller, 1989).

Another context-related consistency test, termed an adding-up test, has been proposed recently by Diamond. At an abstract level the test follows from satisfying duality properties that are commonly assumed in other areas of applied microeconomics. The test requires that a sequence of specified changes add-up to the set of changes taken as a package. There are two practical difficulties with the test that come to light in trying to operationalize it using multiple subsamples of respondents. One approach to structuring the CV survey questions involves asking at least one of the subsamples to "pretend" that they had already received a specified good and paid a specified amount for it. It may be difficult to get respondents to take such an exercise seriously. The other involves making the assumption implicit in Diamond's illustrative example that respondents are indifferent between a program which prevents some number of existing birds from being killed and a hatchery program which produces the same number of new birds (Smith and Osborne, 1996). Substitute children for birds and the implication of this assumption becomes striking.

V. Strategic Behavior and CV Elicitation Formats

The possibility of strategic behavior in the form of free riding has long concerned economics dealing with public good issues. Economists suspicious of survey based answers made the opposite translation and believed (without theoretical justification) that survey based WTP estimates would be larger than true WTP, since they perceived no money directly changing hands. This led to early recommendations to make survey scenarios as hypothetical as possible in order to avoid strategic behavior. However, without an incentive for strategic behavior in a CV survey, any response is as good as any other and responses provided in such context cannot be given an economic interpretation. Thus, the standard CV recommendation has long been to offer respondents a real choice and take seriously the opportunities offered for strategic behavior (Mitchell and Carson, 1989).

The incentive structure for truthful preference revelation is closely related to the CV elicitation format used. Only if incentive and informational effects are equivalent between elicitation formats should one expect to see equivalent WTP estimates from them. Even the most casual examination of the literature suggests that WTP estimates are different across elicitation formats. This has often been taken as evidence by critics that survey respondents do not have "well-defined" preferences for the good they are being asked about (Diamond and Hausman, 1994).

The stylized facts concerning the comparative properties of different elicitation formats are fairly easy to develop estimates from binary discrete choice formats tend be higher than those from other formats. Responses to the two questions in the double bound dichotomous choice format are imperfectly correlated. Open-ended type questions tend to yield many zeros, few very small amounts, and a small fraction of very large amounts. Final WTP responses in iterative bidding games are correlated with the starting point used. Multinomial choice questions and sequences of paired comparisons tend to suffer from violations of the independence of irrelevant alternatives condition.

It is useful to start with a (one-shot) single binary discrete choice (SBDC) question as Carson, Groves and Machina have shown that all of the commonly used most preferred alternative out of two options. These generalizations take three directions. First, sequences of paired comparisons ask for the most preferred alternative in each pair. The key additional assumption of this format is independent responses across questions. From a strategic perspective, this mechanism includes a number of commonly used formats as special cases including double-bounded dichotomous choice (DBDC) questions, complete ranking of alternative and, with the additional assumption that preference intensity can be measured, and rated pairs. Second, open-ended type questions, including payment cards and bidding games drop the cost amount for the second alternative, and instead, asks for the amount that would make the respondent view the two alternatives as equivalent from a utility perspective. Third, multinomial choice questions asks a respondent to pick the most preferred out of $k > 2$ alternatives. It is well-known from the Gibbard- Satterwaite theorem that none of these generalizations of the SBDC question can be incentive compatible without placing restrictions on the form of agent utility. Hence, one should expect divergences between the WTP distributions implied by responses to these formats and an incentive compatible SBDC question (Gibbard, 1973; Satterwaite, 1975).

The Gibbard-Satterwaite theorem is essentially a negative result. It does not say that any SBDC question is incentive compatible, as has sometimes been asserted only that other question formats cannot be. Several auxiliary assumptions, which can succinctly be summarized as implying a onetime take-it-or-leave-it choice with the government having the power to supply the good and collect payment for it, are required to make a SBDC question incentive compatible. Starting with the classic proof that a binding referendum is incentive compatible show that the binding nature of the vote can be replaced with the more general condition that the likelihood of one of the alternatives being implemented is weakly monotonically increasing in the percent who favor that alternative without changing the original incentive properties of the binding referendum. This leads to the advisory referendum mechanism. Then, drawing upon an old result of Green and Laffont, they note that any mechanism within a broad class, including those considered here, can be implemented with an exogenously chosen sample of agents rather than the entire population. This resulting mechanism is an "advisory survey" and it has the same incentive properties of a binding referendum. This result should be useful for CV researchers working in countries where framing the choice to be offered as a "referendum" is not as natural of a setting as it is in some parts of the United States (Diamond and Hausman, 1994).

Two instances where Carson, Groves and Machine show that a SBDC is not incentive compatible are the provision of a public good via voluntary contributions and the provision of a new private or quasi-public good. In the first case, the optimal strategy of an agent who desires the public good is to say she will contribute in order to encourage an actual fund raising drive and then to free ride when the fund raising effort is mounted in hopes that other will pay for the good's provision. This predicted result has been demonstrated repeatedly in empirical tests. Results from these empirical tests are often used by CV critics (e.g. Diamond and Hausman, 1994) to suggest that CV always overestimates true WTP. Since actual contributions should be lower than true WTP due to incentives for free riding, the ratio of the two estimates as an indicator or CV performance in other contexts is likely to be wildly off the mark (Carson, Flores and Mitchell, 1999).

For provision of private or quasi-public goods, a yes response increases the likelihood that the good will be provided, however, the actual decision to purchase the good need not be made until later. Thus, a yes response increases the choice set at no expense. Hence, a SBDC question should overestimate purchase probabilities. This is the consistent empirical result (Cummings, Harrison and Rutstrom, 1995). Rather than representing the "best" case scenario for seeing how CV works as is often claimed, the private goods case is one that should (and does) perform poorly (Johannesson, Liljas and Johansson, 1998).

The desirable incentive properties of a binary discrete choice question can be restored in instances where the agent is asked to choose between two alternatives, neither of which represents a strict addition to the choice set. The most common situation here involves two different configurations of a quasi-public good; for example, an unimproved park with a low 9possibly zero) entrance fee and the same park with higher quality amenities and a higher entrance fee. Since both quality/price configurations do not exist simultaneously, the optimal response involves choosing the most preferred alternative.

Note, though, that the response to such a question does not convey any information about changes in the visitation rate to the park.

Turning now to other elicitation formats, Carson, groves and Machina show that asking the second question in the DBDC format should signal agents that the cost is uncertain, that the quality of the good has changed or that the government is willing to bargain over costs. Any of these interpretations suggest that the latent WTP distributions implied by the first and second distributions should be different (i.e., a correlation coefficient less than one), with the WTP distribution estimated from the second question being generally left-shifted relative to that from the first question. These predictions are consistent with the empirical evidence. Alberini shows that even though the double-bounded format may induce a downward bias in the estimate that this format may still be preferred in a mean-squared error sense due to the large confidence intervals associated with the binary discrete choice format.

Optimal response strategies with respect to open-ended type questions are highly dependent upon beliefs about how the responses are going to be used. Carson and Machina have shown, however, that the optimal response under most plausible beliefs "pivot" on expected cost in the following sense: the optimal response for an agent whose WTP is less than the expected cost is to give a zero response (and protest in other ways if possible) while the optimal response for an agent whose WTP is greater than expected cost is to give a response ranging from expected cost to the highest "acceptable" amount. This pattern of responses is sufficient to provide WTP responses characterized by all three of the stylized facts about this elicitation format: many zero responses, few small amounts, and a correlation between the WTP response and any variable (e.g., the starting point used in a bidding game) that is perceived correlated with expected cost.

Sequences of paired comparisons involving multiple goods and multinomial choice questions present a number of problems. In general, the optimal responses depends upon beliefs about how other agents will respond and about the rule being used to aggregate responses. When only one good will be provided, the general finding from the voting literature is that optimal strategies collapse toward picking one of the two alternatives perceived as being the most preferred independent of the agent's response.

When multiple goods can be provided it is necessary to first ask whether the agent's utility is defined only on one good (e.g., the recreation site visited) or on the bundle of goods. In the latter case, interpretation of the alternative chosen is difficult because it may be optimal, in some instances, to pick a good that would be least preferred if only one good was provided. The inherent problem here is that since utility is defined on the bundle of goods to be provided, the researcher is asking an ill-posed question by requesting that the agent pick the single most preferred alternative.

In the case where utility is defined by only one of the goods (a situation that characterizes many consumer product and quasi-public good choice contexts) more optimistic results have been obtained (Carson, Groves and Machina, 1999). There are many examples of the multinomial choice format being successful in this case. If all but one of the goods will be provided, the multinomial choice question can be shown to be incentive compatible because such a question is formally equivalent to a binary discrete choice which matches the agent's most preferred alternative against another (unknown)

alternative competing for the one good that will not be provided. The worst case for agents is having their second most preferred alternative provided. As one reduces the number of goods to be provided from k-1 to k-2 of the goods, the optimal response strategy will be to provide the most preferred alternative or the second most preferred alternative. For provision of k-3 of the alternatives, the indicated choice should be one of the agent's top three and so on. Response behavior of this sort results in violations of the independence of irrelevant alternatives (11A) assumption; something that is frequently observed in empirical studies using this elicitation format. Further, optimal non-truthful preference revelation should usually involve trying to induce provision of the desired set of attributes while altering the true price sensitivity. An implication of this sort of behavior is that it should be possible to obtain reasonable correct estimates of marginal tradeoffs between attributes since the scale factor (i.e., the marginal value of money, usually taken to be the negative inverse of the coefficient on price in simple models), likely to be adversely impacted by the 11A violations, cancels out in calculating these tradeoffs. Total value estimates, however, will be problematic because the scale factor plays a large role in their determination.

To summarize our discussion on strategic behavior, theoretically, one should expect different answers from different elicitation formats. The empirical evidence from a directional perspective is in accord with those theoretical predictions, although the magnitudes of the differences are often not as extreme as simple theoretical models would suggest. At present, the choice a CV researcher typically faces is between using an elicitation format that is unbiased but with a large confidence interval and using one that is potentially biased but with a much tighter confidence interval.

V1. Test of Validity and Reliability

Random survey responses represent the antithesis of survey responses that have passed various validity and reliability tests. Validity refers to the correspondence between what one wished to measure and what was actually measured. Reliability refers to the measurement's replicability. Both terms can be operationalized in a variety of ways. The ideal way of determining validity is by comparing the measurement made to some criterion measurement know to be correct, e. g., the kilogram weight kept at the National Bureau of Standards. Unfortunately, such a criterion to which CV WTP measurements can be compared does not exist. Furthermore, no such criterion exists to which any other consumer surplus estimate can be compared, irrespective of the econometric technique used or whether the good is private or public. Consumer surplus represents the difference between what the agent had to pay and the maximum the agent would have been willing to pay, a quantity that is inherently unobservable. In such instances, researchers look at different approaches to determining validity; two common ones are construct validity and convergent validity. Construct validity refers to how well the measurement is predicted by factors that one would expect to be predictive a priori. Section 1V, on the consistency of CV results with the predictions of economic theory, examines one of the standard ways of examining construct validity. Below we explore the concept more generally. The second approach we consider is convergent validity. This approach can be taken only when measurements of the phenomena of interest are available using two different techniques. Two types of reliability have interested CV researchers. One is the temporal stability of the estimate if two different samples of the sample population are interviewed

with the same survey instrument at two different point in time. The other is the classic test-retest reliability where an initial sample of respondents is later re-interviewed using the same survey instrument. We consider specific construct and convergent validity, and reliability issues below.

Construct Validity

Most CV studies provide an equation that relates some indicator of the respondent's WTP to the respondent's characteristics and to characteristics of the good. As already discussed, economic theory suggests that the probability of wanting the good should go down as the price of the good increases. This effect is almost universally found. The issue of the sensitivity of CV results to the scope of the good being valued was discussed in Section 1V. In Section V, we noted a number of situations where theory predicted differences between WTP estimates based upon the elicitation format used and that the empirical evidence generally supported the direction of these divergences.

Other variable are more specific to the particular good being valued. Usually one would anticipate direct users of the good to be willing to pay more than those who do not use the good. Therefore, one would expect to find that environmentalists are willing to pay more for environmental goods than non-environmentalists. This too is almost always found. Often, it is possible to ask about environmental attitudes specifically tied to the good. As one would expect, these are generally better predictors of WTP than self-identification as an environmentalist. Income usually has a positive and significant effect on WTP. Age usually has a negative effect while geographic proximity usually has a positive effect. Perception variables related to the provision of the good tend to be predictive of respondent WTP in the expected manner. In particular, perceptions that the program to provide the good will not be successful or that the payment vehicle is not appropriate tend to be very negatively associated with WTP.

These general findings are a synthesis drawing upon the specific findings of a large number of CV studies. More important in practice is to look at the relationship in each particular study. If a valuation function does not have substantial explanatory power, the results of the study should be viewed with skepticism. In that case, one of two problems exist with the study; either the responses to the WTP question are insensitive to the examined characteristics of the respondent or of the good, or the researcher has failed to collect sufficient relevant economic covariates to explain much of the variation in respondent WTP. In both instances, the result suggests that the researcher does not understand the factors that drive the population's WTP.

Convergent Validity

For three decades, CV estimates for quasi-public goods have been compared to those from other non-market valuation techniques based on observed behavior such as travel cost analysis and hedonic pricing. These comparisons can be made in two ways. First, one can compare the actual estimates as rations or differences. The Carson et al. meta-analysis summarizes this evidence and concludes that CV estimate, on average, are somewhat, but not greatly, smaller than those based on revealed preference techniques. Second, one can look at the correlation between estimates based on different techniques. Here Carson et. al. find correlations ranging between 0.78 and 0.92 depending upon the treatment of the sample comparisons.

Another source of comparisons are referenda. A market for goods with substantial passive use considerations can be constructed using a referendum model such that, if the good provided, the government will extract payment. Surveys taken immediately before an election have an enviable record of accurately predicting subsequent election outcomes. In a CV context, Carson, Hanemann, and Mitchell asked a large sample of California households a question patterned after a legislatively initiated California bond issue on water quality which was to appear several months later on the election ballot. There was little information disseminated on the bond issue during the election campaign beyond that provided in the California voter's pamphlet. They found that their survey results predicted the actual vote quite closely. Polasky, Gainutdinova and Kerkvliet find a similar result in their study on 1996 comparing a prior estimate from a CV study to the actual vote on a proposition in Oregon to acquire open-space.

Reliability

Reliability, as opposed to validity, is an index of the reproducibility and stability of a measure. For CV studies, the index that is relevant for policy purposes is the stability of WTP measures over time. Several studies have replicated results with similar questionnaires administered to independent sample at two different points in time. For instance, Carson and Mitchell, after adjusting for inflation, report finding values within $1 for a national water quality improvement in two national surveys three years apart. Recently, the original Alaska Exxon Valdez questionnaire was administered to a new sample; the values per household and the coefficients on the two regression equations predicting those value were almost identical to those of the original sample of two years earlier. Whitehead and Hoban administered the same WTP survey involving air and water quality improvements to two separate sample of the same population five years apart and found the estimated valuation function unchanged, although WTP estimates were different because values of some of the main predictor variables had changed.

A number of CV studies have looked at the issue of correlation between the same respondent' answer at two different points in time. Respondents may not give the same answer for many reasons, such as changes in the respondent's financial situation, changes in expenditure opportunities, and perhaps most importantly, a retesting effect. These studies Loomis and Teisl studies have generally shown significant correlation in the range of 0.5 to 0.9, between respondent answers at different points in time. In a more ambitious variant of this type of test, McConnell, Strand and Valdes interviewed respondents at two different points in the fishing season, and found that the valuation function obtained was similar in both instances. After accounting for the differences in the nature of the fishing opportunities in the second time period, they were able to predict the results or the second interview based upon the first interview (Loomis and Teisl, 1995).

V11. Concluding Remarks

The recent debate surrounding the use of CV is, to some degree, simply a reflection of the large sums at stake in major environmental decisions involving passive use and the general distrust that many economists have for information collected from survey. In an academic context, that debate has often been healthy. CV research has matured as a result of the spotlight that has been placed upon it. The theoretical foundation underlying CV

has elaborated and many problems of empirical measurement usually ignored or avoided by economists are highlighted by its use.

Outside, of academic journals, though, criticism of CV takes a largely anecdotal form, ridiculing the results of particular CV studies. Many of these studies use techniques known to be problematic or are low budget graduate student exercises labeled as state-of-the-art. The implication drawn, however, it that all CV surveys produce nonsense results upon which no reasonable person would rely. Contrary evidence is almost completely ignored. Unfortunately, such an approach has more potential to confuse rather than enlighten.

The CV critics' attack of theoretical inconsistency has forced environmental economists to think much more deeply about what the underlying theory says about the provision of environmental amenities As Smith recently pointed out, "Contingent valuation has prompted the most serious investigation of individual preferences ever undertaken in economics" In this regard, it is now clear that benefit-cost analysts have for too long relied upon the much more well-developed theoretical framework for price changes, the same framework critics took as their intuition on which to base judgment of the theoretical consistency of empirical CV results. The imposed quantity changes that characterize most environmental amenities have a number of fundamentally different welfare economic properties than do price changes for marketed goods. While many of these results were previously known, only recently has the full richness of that theoretical framework and general consistency of empirical CV results with it become apparent.

A long-standing issue with CV is that it seems to many like an easy even trivial task to ask people what they are willing to pay for a good. Many CV critics fail to appreciate the difficulty of asking such a question. If preferences can be measured at all by asking people survey questions, then the CV critics effectively argue that it should not matter how implausible the questions are to respondents or how many counterfactuals the respondent is told to "suppose". Given that premise, if the responses to such questions are deemed implausible, or violate economic theory in some fashion, CV, as an approach, is deemed to be flawed. However, we believe the results of a survey question should not be given a direct economic interpretation unless the good to be value is clearly explained, its delivery to the public made plausible, and a realistic expectation of payment created. A reliable CV survey is neither simple nor inexpensive to implement. Indeed, we believe that at this point in the development of CV, the key objective in terms of methodological development should shift to trying to determine how to reduce the cost of conducting CV studies while still maintaining most of the quality of the very best studies now being conducted. Development and research along these lines will be crucial in effectively incorporating the public's preferences into the environmental decision making arena.

A perhaps justifiable fear of the CV critics is that the estimates from any CV survey done will be automatically accepted in policy forums. Poor quality and self-serving benefit-cost studies of all kinds, including those based on CV surveys, abound in most contentious policy debates. Like other economic methodologies, however, objective and readily identifiable criteria are available by which the quality of CV studies can be judged. Furthermore, there are many social scientists with substantial knowledge of CV capable of making such judgments.

Even if all of the survey related issues to valuing a public good can be overcome, CV is not without its limitations. CV shares, with other neo-classical preference-based approaches to economic value, two principal limitations to which some object. First, WTP measures are inherently limited by wealth. This limitation is offensive to many who believe that government decision making should not be based to any extent on ability to pay. Second, only the preferences of the current generation for themselves and for future generations are taken into account. The actual preferences of future generations are not explicitly considered and, from a neoclassical economic perspective,

Are inherently unknowable. However relevant these limitations are from a policy perspective, they are not issues per se of the measurement of economic value. Without stated preference survey methods, though, economists have to admit that they are not measuring the passive use aspects of environmental and other non-market goods, and that these are the aspects about which people may care about most. A benefit-cost analysis that omits these considerations will at best be incomplete and at worst completely misleading.

References

Andreoni J., (1989), '*Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence*', Journal of Political Economy 97, pp. 1447-1458.

Bateman I.J. and I.H. Langford, (1997), '*Non-users' Willingness to Pay for a National park: An Application and Critique of the Contingent Valuation Method*', Regional Studies, 31, pp. 571-582.

Blinder A.S., (1991), '*Why are Prices Sticky: Preliminary Results from an Interview study*', American Economic Review 81, pp. 89-100.

Carson R.T., (1991), 'Constructed Market', in J.B. Braden and C.D. Kolstad, eds., *Measuring the Demand for Environmental Commodities*, Amsterdam: North-Holland, pp. 121-163.

Carson R.T., (1997), 'Contingent Valuation Surveys and Tests of Insensitivity to Scope', in R.J. Kopp, W. Pommerhene and N. Schwartz, eds., Determining the Value of Non-Marketed Goods: Economic, Psychogical and Policy Relevant Aspects of Contingent Valuation Methods. Boston, Kluwer, pp. 127-163.

Carson R.T., N.E. Flores and R.C. Mitchell, (1999), 'The Theory and Measurement of Passive Use Value, in by I.J. Bateman and K.G. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the USA, EC*, and Developing Countries, New York: Oxford University Press, pp. 97-130.

Carson R.T. and R.C. Mitchell, (2000), '*Public Preferences Toward Risk: The Case of Trihalomethanes*', in A. Alberini, D. Bjornstad and J.R. Kahn eds., Handbook of Contingent Valuation, Northampton, Edward Elger.

Cummings, R.G., G.W. Harrison and E.E. Rutstrom, (1995), '*Homegrown Values and Hypothetical Survey: Is the Dichotomous Choice Approach Incentive-compatible*?', American Economic Review 85, pp. 260-266.

Desvousges W.H., (1993), 'Contingent Valuation: The Wrong Tool for Damage Assessment, Choices 8, No. 2, pp. 9-11.

Diamond P.A. and J.A. Hausman, (1994), '*Contingent Valuation*: Is Some Number Better than No Number?, Journal of Economic Perspective 8, No.4, pp. 45-64.

Flores N.E., (1995), '*The Effects of Rationing and Virtual Price Elasticities*', Department of Economics, University of California, San Diego, Discussion Paper, pp. 95-120.

Gibbard A., (1973), '*Manipulation of Voting Schemes: A General Result*', Econometica 41, pp. 587-601.

Goodstein E.S., (1995), *Economics and the Environment, Englewood Cliffs*, NJ: Prentice Hall, Oxford University Press, pp. 47-58.

Hanemann W.M., (1995), '*Contingent Valuation and Economics*', in K.G.Willis and J.T. Corkindale, eds., Environmental Valuation New Perspectives, Oxon: CAB International.

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

Huang J.C., and V.K. Smith, (1998), '*Monte Benchmarks for Discrete Response Valuation Methods*', Land Economics 74, 186-203.

Just R.E., D.L. Hueth, and A. Schmidt, (1982), *Applied Welfare Economics and Public Policy*, Englewood Cliffs, NJ: Prentice-Hall.

Kahneman D. and I. Knestch, (1992), '*Valuing Public Goods: The Purchase of Moral Satisfaction*', Journal of Environmental Economics and Management 22, pp. 57-70

Knetsch J.L., (1990), '*Environmental Policy Implications of Disparities Between Willingness to Pay and Compensation Demanded Measures of Values*', Journal of Environmental Economics and Management 18, pp. 227-237.

Krutilla J.V., (1967), '*Conservation Reconsidered*', American Economic Review 57, 777-786.

Loomis J.M. and A.W. Teisl, (1995), '*Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses*', American Journal of Agricultural Economics 71, pp. 76-84.

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

Mueller D.C.,(1989), *Public Choice 11*. Cambridge: Cambridge University Press.

Poe G.L., K.J. Boyle and J.C. Bergstrom (2000), '*A Meta-Analysis of Contingent Values for Groundwater Contamination in the United States*', paper presented at the European Association of Environmental and Resource Economics, Rythymnon, Greece.

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

Satterwaite M., (1975), '*The Pure Theory of Public Expenditure*', Review of Economics and Statistics 36, pp. 387-389.

Smith V.K. and L. Osborne, (1996), '*Do Contingent Valuation Estimates Pass a Scope Test? A Meta-Analysis*', Journal of Environmental Economics and Management 31, pp. 287-301.

Tourangeau R., L.J. Rips and K. Rasinski, (2000), '*The Psychology of Survey Response*', New York: Cambridge University Press.