Environmental and Resource Economics **19:** 173–210, 2001.
© 2001 *Kluwer Academic Publishers. Printed in the Netherlands.*

Contingent Valuation: Controversies and Evidence

RICHARD T. CARSON[1], NICHOLAS E. FLORES[2] and NORMAN F. MEADE[3]
[1]*Department of Economics, University of California, San Diego;* [2]*University of Colorado, Boulder;* [3]*National Oceanic and Atmospheric Administration, U.S. Department of Commerce*

Accepted 14 August 2000

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.

JEL classification: Q26, D61

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 1960s, several non-market valuation techniques have been developed in recognition of the importance of these services. Of these techniques, the most commonly used is contingent valuation (CV).[1] 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.[2]

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 individual project analyses.[3] Although CV is the most frequently used non-market valuation technique for environmental goods, debate persists over the reliability of CV and the overall suitability of passive use values 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.[4] 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 2 addresses welfare economic issues associated with CV and passive use. Section 3 identifies key elements in the design, implementation, and analysis of CV surveys and their relevance to CV criticisms. Section 4 discusses empirical results from CV studies and their consistency with economic theory. Section 5 discusses strategic behavior and its relevance to CV elicitation formats. Section 6 discusses issues related to the validity and reliability of CV estimates. Section 7 offers concluding remarks.

2. 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 benefit-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 given 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 form: 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.[5] 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 proposals for how to aggregate the monetized measures obtained from agents have been advanced (Mueller 1989).

2.1. TOTAL VALUE

The term *total value*, synonymous with true WTP or WTA, arose in environmental economics with the awareness that sometimes substantial 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 (Carson, Flores and Mitchell 1999), for our purposes here, it is useful to decompose total value into direct use value and passive use value. 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 directly 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 Krutilla's (1967) seminal observation that many people valued natural wonders simply for their existence. Krutilla 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.

2.2. PASSIVE USE VALUE

Three camps hold fundamentally different positions on passive use value. They are: (1) passive use values are irrelevant to decision making, (2) passive use values cannot be monetized, and thus, can only be taken account of as a political matter or by having experts decide, and (3) passive use values can be reliably measured and should explicitly be taken into account.[6]

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 options, 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 benefit-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 take those benefits into account.[7] 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.

2.3. 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, that 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 (Carson 1991), such as an actual referendum on whether to provide the public good (Deacon and Shapiro 1975), or a simulated market in which the good is actually provided (Bishop and Heberlein 1979), 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 (Borcherding and Deacon 1972).[8]

Many 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 (e.g., Arrow 1958; Schelling 1968; Mishan 1976; Blinder 1991) note the limitations that this reluctance to query people directly about their preferences imposes on the questions which economists can address.

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 environment during the last thirty years, it became 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.

2.4. 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" (Kahneman and Knestch 1992), and hence, passive use is not an economic value.[9] At some level, this argument harkens back to an age-old source of conflict between economists and non-economists: 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.[10] Neither pure nor impure altruism was seen as a non-economic motivation by these authors. Andreoni (1988, 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.[11] However, warm glow in the sense of Andreoni is largely irrelevant to a CV discussion, unless respondents get utility simply from paying more taxes to the government for the good, irrespective of whether it is actually provided.[12]

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 Research Center and the Exxon Valdez Oil Spill CV survey instrument. That test (Carson, et al. 1994) 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.

2.5. FAMILIARITY

A frequent claim (e.g., Desvousges et al. 1993a) 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.

This rationale, 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.[13]

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.

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 them 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 examples illustrate that specific personal experience is not required for making meaningful economic choices.[14]

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).

2.6. MARKET SIZE

The question of the appropriate market (i.e., the population whose values should be sought for a given amenity) 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 valued 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 (e.g., Sutherland and Walsh 1985; Bateman and Langford 1997) 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 the class of resource are likely to be a more dominant factor than distance.

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 values 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 larger cities and countries. Aggregation across agents has nothing to do with CV *per se*; it is merely part of the definition of the *value* of a *public good*.

3. Survey Design, Administration, and Analysis Issues

Contingent valuation is a basic approach to non-market valuation rather than a single specific economic valuation technique.[15] 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. This issue of elicitation formats is taken up at some length in Section 5.

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.[16] 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.[17] 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 because 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 does not face the same limitation, but researchers must still 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 valued understandable to the population of interest. The designer needs to convey enough detailed information to convince respondents that the supplier 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.[18] 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.[19] All of this must be done without overwhelming respondents with large amounts of information that they find extraneous to the choice being offered.[20]

The most commonly used modes of administration for CV surveys are in-person interviews and mail surveys. The former generally reduces the likelihood of sample selection bias because respondents who return mail surveys tend 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.[21] Many standard survey research texts contain substantive discussions of the major issues involved in the choice of the mode of survey administration (e.g., Lansing and Morgan 1971; 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).[22] Further, one should expect to see the estimates at the respondent level vary systematically in sensible ways, as discussed below.

4. 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.[23] 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, an assertion that is now routinely repeated in introductory texts on benefit-cost analysis and environmental economics (e.g., Goodstein 1995). We devote a subsection below to this topic.[24]

The price and scope tests have the advantage of being simple unidirectional hypothesis tests with very close ties to the underlying economic theory (Hanemann 1995). These tests correspond well with economic intuition. One might also make conjectures about the relationship between respondent income and WTP, on the difference between estimates 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 (Cornes 1992).

4.1. SCOPE TESTS

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 (Arrow and Leamer 1994), (2) problems in CV survey design and administration which tend to mask sensitivity to scope[25] (Carson and Mitchell 1995), 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 valued is an 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 specific survey?

A test of responsiveness to scope can be implemented either internally or externally. 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 (e.g., Brookshire, Ives and Schulze 1976). CV critics have argued strongly that respondents may simply be trying to be "internally consistent" in their answers.[26] Recent attention has focused on external tests of scope and, in particular, the evidence presented by Kahneman (1986; Kahneman and Knetsch 1992) and at the Exxon symposium (Hausman 1993), suggesting that respondents to CV surveys do not give different values to goods that differ in scope. Carson (1997) 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 purposes where two or more different levels of a good were of interest to policy makers.[27] These studies have advantages over the work of Kahneman and those reported in Hausman (1993) 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 situations where passive use considerations were thought to predominate, while the rest dealt with situations where direct use was thought to predominate.

The Carson (1997) review shows that, since 1984 (the date 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 (Carson, Wilks and Imber 1994; Carson et al. 1994), using in-person interviews and well-constructed questionnaires containing extensive visual aids depicting the good to be valued, rejected the scope insensitivity hypothesis at $p < 0.001$. Four recent meta-analyses that looked at studies valuing outdoor recreation (Rosenberger, Loomis and Shrestha 1999), air quality changes (Smith and Osborne 1996), wetland functions (Brouwer, et al. 1999), groundwater contamination (Poe, Boyle and Bergstrom 2000) also rejected the scope insensitivity hypothesis by

showing that the CV estimates from different studies vary in a systematic (and expected) fashion with differences in specific characteristics of the good.[28]

Poorly executed survey design and administration procedures appear to be a primary cause of problems in 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 surveys 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 Kahneman and Knetsch (1992) work used short telephone surveys with vaguely defined goods, provision mechanisms, and payment obligations. Desvousges et al.'s (1993b) 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. The Desvousges et al. study had three treatment programs, saving (a) 2,000, (b) 20,000, or (c) 200,000 birds, respectively, which seem quantitatively like large well-defined differences between the three treatments. Common sense suggests that one should see sizeable differences in the WTP estimates obtained for the three programs. However, the three programs were also described to survey respondents as saving (a) much less than 1%, (b) less than 1%, or (c) about 2% of a population of 8.5 million migratory waterfowl. Cast in this light, and particularly when considering that the actual number of birds saved by any of the programs and the overall population of birds must be stochastic variables, one would expect to see much smaller differences between the WTP estimates from the three programs. This example illustrates that the appropriate scale upon which to compare differences in non-marketed goods is often problematic and can influences one's perception of what results should be expected.

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. (1993), 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 ("Stated WTP to protect each of the three areas is the same") using a Kruskal-Wallis test. Carson and Flores (1996) show that their statistical test has little power to detect large differences, and instead, estimate a simple OLS regression of WTP on the number of acres in each of the three wilderness areas. They reject Diamond et al.'s (1993) null hypothesis at $p = 0.01$.[29]

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 (Rollins and Lyke 1998). 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 (Beattie et al. 1998; Jones-Lee, Loomes and Philips 1995).[30] 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 CV exercise can have a chance of working (Carson and Mitchell 2000). In a recent paper, Corso, Hammitt and Graham (2000) look at several different risk communication devices in the context of a CV survey. They find 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. (2000) who attempt to break the problem into two pieces, one involving value elicitation and the other involving standard gambles, chained together to arrive at values for small probabilities. The valuation of risk reductions is likely to remain an active research area for some time.

4.2. 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 (1994) reporting on one of Exxon's studies notes:

> An economic interpretation of the results on the relationship of income to WTP in these experiments is that preservation of Selway-Bitterroot wilderness is a "necessary" good, with a low-income elasticity. However, it seems economically plausible that preservation would be a "luxury" good that for poor households is displaced by needs for food and shelter.

The problem is that the terms *necessary* (e.g., normal but not luxury) and *luxury* are defined in terms of the income elasticities of demand; a measure based on varying quantity, not in terms of the income elasticities of WTP, a measure based upon holding the quantity fixed. Flores and Carson (1997) 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 looks 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.[31] 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 than 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 goods), 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.

4.3. Divergence between WTP and WTA estimates

If total value in an economic sense can always 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 (Willig 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 1991).[32] Changes in environmental goods tend to fall 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 (e.g., Hammack and Brown 1974). 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 settings using actual transactions (e.g., Knetsch, Thaler and Kahneman 1990).[33] Even financial assets such as junk bonds and over the counter stocks, when thinly traded, often show much larger 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 (Kahneman and Tversky 1979) and to show that within that framework (Hanemann 1991), the theory being applied failed to capture key aspects of the situation.[34] 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 (e.g., Kolstad and Guzman 1999; Zhao and Kling 1999; List 2000). The key implication of this divergence for applied policy work is that property rights 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).

4.4. 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 findings turn on the same underlying theoretical issue: substitution and income effects.

The first finding indicates that adding up what people say they 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 overestimates. However, Hoehn and Randall (1989) 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.[35]

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 (Randall, Hoehn and Tolley 1981). 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 substitution and income effects. Carson, Flores, and Hanemann (1998) show that if one assumes that the goods being valued are normal goods and (Hicksian) 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.[36] 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.[37] 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 (1995) 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 (σ_{ij}^{d}) taken from Deaton's (1974) well-known analysis of consumer demand in the United Kingdom. Good one is food and good two is clothing:

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

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 on 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 changes 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 prices for the rationed goods (food and clothing) respond to changes in the rationed levels of both of these goods. These measures of responsiveness, the virtual price substitution elasticities (σ_{ij}^{v}), are related inversely, as a system, to

the compensated price elasticities (Madden 1991; Flores 1995). For the food and clothing example, the virtual price matrix of substitution terms is:

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

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 (Carson and Jeon 2000), particularly if payment is required over a short time period. Further, income is known to be poorly measured in general population surveys (Sudman and Bradburn 1982). These sources of measurement error probably bias estimated income effects downward.

CV critics such as Kahneman and Knetsch (1992) 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. Kahneman and Knetsch's statement is applicable to economic analysis in general, which if done correctly is context specific.[38] Value in an economic sense is always a relative rather than absolute concept (Debreu 1959). Even more to the point is Flores' (1999) 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 (Mueller 1989), it would have been surprising to see how the use of CV somehow avoided it.[39]

Another context-related consistency test, termed an adding-up test, has been proposed recently by Diamond (1996). At an abstract level the test follows from satisfying duality properties that are commonly assumed (and commonly violated, e.g., Bateman et al. 1997) 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 seri-

ously. 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.

5. Strategic Behavior and CV Elicitation Formats

The possibility of strategic behavior in the form of free riding has long concerned economists dealing with public good issues (Samuelson 1954). 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.[40] 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 incentives and informational effects are equivalent between elicitation formats should one expect to see equivalent WTP estimates from them (Hoehn and Randall 1987; Carson, Groves and Machina 1999). 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 (Carson, Groves and Machina 1999). WTP estimates from binary discrete choice formats tend to 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 (IIA) condition.

It is useful to start with a (one-shot) single binary discrete choice (SBDC) question as Carson, Groves, and Machina (1999) have shown that all of the commonly used preference elicitation formats can be seen as generalizations of the SBDC question, which asks for the 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 assump-

tion 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 rankings of alternatives and, with the additional assumption that preference intensity can be measured, 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 (Gibbard 1973; Satterwaite 1975) 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.

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 (Cummings, Harrison and Rutström 1995; Cummings et al. 1997); only that other question formats cannot be.[41] Several auxiliary assumptions, which can succinctly be summarized as implying a one time 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.[42] Starting with the classic proof that a binding referendum is incentive compatible (Farquarson 1969), Carson, Groves, and Machina (1997) 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 (1978), 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.[43] 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.

Two instances where Carson, Groves, and Machina (1999) 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 others will pay for the good's provision. This predicted result has now been demonstrated repeatedly in empirical tests (Seip and Strand 1992; Champ et al. 1997). 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 indictor of CV performance in other contexts is likely to be wildly off the mark.

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 Rutström 1995; Johannesson, Liljas and Johansson 1998; Blamey, Bennett and Morrison 1999). 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.

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.[44] The most common situation here involves two different configurations of a quasi-public good; for example, an unimproved park with a low (possibly zero) entrance fee and the same park with higher quality amenities and a higher entrance fee. Since both park 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 (1999) 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 (Cameron and Quiggin 1994; Alberini, Kanninen and Carson 1997). Alberini (1995) 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, Groves, and Machina (1999) 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 the 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 also present a number of problems. In general, the optimal response 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 (Adamowicz et al. 1999). 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 but not their first. 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 (IIA) 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 reasonably 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), which is likely to be adversely impacted by the IIA violations, cancels out in calculating these tradeoffs.[45] 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.

6. Tests 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 known 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 4, 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 points in time.[46] 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.

6.1. 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 4. In Section 5, 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 variables 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 relationships 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.

6.2. 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 ratios or differences. The Carson et al. (1996) meta-analysis summarizes this evidence and concludes that CV estimates, 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. (1996) find correlations ranging between 0.78 and 0.92 depending upon the treatment of the sample comparisons.[47]

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 is 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 (1987) 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 (1996) find a similar result in their recent study comparing a prior estimate from a CV study to the actual vote on a proposition in Oregon to acquire open-space.

6.3. 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 samples at two different points in time. For instance, Carson and Mitchell (1993), 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 values were almost identical to those of the original sample of two years earlier (Carson et al. 1997). Whitehead and Hoban (1999) administered the same WTP survey involving air and water quality improvements to two separate samples 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 respondents' answers 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 (e.g., Loomis 1989; Teisl et al. 1995) have generally shown significant correlations 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 (1998) 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 of the second interview based upon the first interview.

7. 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 surveys. 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 been 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 (e.g., Niewijk 1994) 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, is 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 (2000) recently pointed out, "Contingent valuation has prompted the most serious investigation of individual preferences ever undertaken in economics."[48] 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 judgments 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 (Hanemann 1991; Ebert 1998). While many of these results were previously known, only recently has the full richness of that theoretical framework and the 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 valued 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.[49]

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 (Sagoff 1994). 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.

Acknowledgements

U.S. EPA Cooperative Agreement R-824698 provided partial support for the work reported on in this paper. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the University of California, San Diego, the University of Colorado, Boulder, the National Oceanic and Atmospheric Administration, or the U.S. Environmental Protection Agency. We received helpful comments related to this paper from Rich Bishop, Trudy Cameron, Michael Conaway, Michael Hanemann, Raymond Kopp, Robert Mitchell, Stale Navrud, Olvar Bergland, Alan Randall, Richard Ready, V. Kerry Smith, Douglas Shaw, Rob Stavins, the editor, and two anonymous referees. Earlier versions of this paper were presented at the Agricultural University of Norway, Harvard University, University College (London), the University of Nevada, Reno, the World Congress of Environmental Economists and the Western Economic Association Meeting.

Notes

1. CV is a survey-based stated preference methodology that provides respondents the opportunity to make an economic decision concerning the relevant non-market good. Values for the good are then inferred from the induced economic decision. The other two basic approaches to non-market valuation rely on observed economic behavior rather than consumer's stated preferences and are known as hedonic pricing (e.g., property value and wage models) and the household production function approach (e.g., travel cost analysis and averting behavior). These basic approaches to non-market valuation have been in use for over 30 years. See Freeman (1993) or Johansson (1993) for overviews of the theoretical and empirical issues involved. Carson, Wright, et al. (1995) provides a bibliography of over 2,000 contingent valuation papers and studies from over 40 countries.
2. In ruling on the case *Ohio* v. *U.S. Department of Interior* (880 F.2d 432 D.C. Cir., 1989), the court adopted the term "passive use value" to encompass a number of frequently used terms such as non-use value, existence value, preservation value, bequest value, stewardship value, intrinsic value, and option value. Conceptually these terms are largely synonymous except for option value, which has a differential intellectual heritage (Weisbrod, 1964). Some CV critics accept the legitimacy of option value but not other components of passive use value. Carson, Flores, and Mitchell (1999) provide a technical review of the relevant theoretical and empirical issues.
3. Much of the current debate in the U.S. is centered on the use of CV by the State of Alaska (Carson et al., 1992) to assess the natural resource damages from the Exxon Valdez oil spill.
4. Hausman (1993) contains papers from an Exxon-sponsored symposium that are highly critical of CV and the inclusion of passive use value in economic assessments. Subsequent to that symposium, Kenneth Arrow and Robert Solow chaired a panel convened by the National Oceanic and Atmospheric Administration to review the use of CV in natural resource damage assessment (Arrow et al., 1993). That panel concluded, "CV studies can produce estimates reliable enough to be the starting point for a judicial or administrative determination of natural resource damages including passive use values" and suggested a number of guidelines to help ensure the reliability of CV survey results. Two series of articles on CV and passive use value appear in a 1993 issue of *Choices* (Carson, Meade, and Smith; Desvousges et. al.; and Randall) and a 1994 issue of the *Journal of Economic Perspectives* (Diamond and Hausman; Hanemann; and Portney).
5. These monetized measures are often referred to as Hicksian consumer surplus measures. Because of its greater ease of calculation, in some instances an approximation to these measures the Marshallian consumer surplus that holds income rather than utility constant is also sometimes used in applied work. This measure is the area under the demand curve and above price and lies between WTP and WTA. Attempts to use CV to measure WTA have often been unsuccessful, although there have clearly been some studies where a WTA question has worked. The reasons lying behind the failures usually include the lack of an appropriate incentive structure and the inability to convince many respondents that they have the right to sell a collectively owned good. See Mitchell and Carson (1989) for a comprehensive discussion of this issue. There has been more success in recent years in the measurement of WTA by addressing these issues in the survey design phase (Howe and Smith, 1994; Groothuis, Van Houtven and Whitehead 1998; Smith and Mansfield (1998).
6. These three positions can be seen in papers by Rosenthal and Nelson (1992), Quiggin (1993), and Kopp (1992).
7. Starrett's concluding remark from *The Foundation of Public Economics* (1988) seems an apt counter to this view: "It is very important to reject the view that since welfare measurement is still quite imprecise, we may as well leave public decision making to the politicians. To do so invites mistakes that are costly on a scale that dwarfs any possible measurement error." With regard to experts, the quintessential question has always been: whose experts?

8. The difficulty, of course, with using votes by political representatives is that the vote on the provision of one public good may often be tied to the provision of other goods or in response to the activities of special interest groups.
9. Kahneman and Knetsch contend that "The amount that individuals are willing to pay to acquire moral satisfaction should not be mistaken for a measure of the economic value of public goods." Harrison (1992) responds succinctly by summarizing the standard neoclassical economic position: "I call my utility 'jolly'. What you chose to call your utility is, as far as I am concerned, your business. What Kahneman and Knetsch want to call it is, of course, their business. To be blunt, this hypothesis is 'meaningless' in the standard methodological sense."
10. This is also true of other instances where economists have looked at altruism as a motive, such as over-lapping generations models in macroeconomics and family relationship models in labor economics. Particular forms of altruism and its negative counterpart, envy, have long been known (Winter, 1969) to raise potential problems with aggregate welfare measures. This argument was put forth at the Exxon symposium (Hausman, 1993) as a problem with the use of CV. However, the argument is not specific either to CV or passive use value; altruism and envy influence private market transactions as well. Furthermore, as long as the agent's altruism (a) is oriented toward public goods (e.g., animals or ecosystems), or (b) is paternalistic in the sense that the agent's desire is that the environmental amenity is made available to others, or (c) is non-paternalistic in the sense that the agent only cares about other agents' utilities but that agent is informed of the plan's cost to the other agents (which is effectively done in discrete choice CV surveys set in a referendum context), then no potential conflicts with respect to including altruism-motivated value exists (McConnell, 1997).
11. It is sometimes argued that contributions to environmental groups are a relevant measure of the public's WTP for environmental goods. This is clearly a specious argument for both the well-known free-rider problem and the fact that environmental groups cannot actually provide the environmental goods they want.
12. An exception as recently noted by Chilton and Hutchinson (1999) is in examining the difference between CV estimates generated using a tax payment vehicle versus voluntary contributions to a charitable organization where warm glow (in the sense of Andreoni) does come into play.
13. This is not to say, however, that some prominent economists (e.g., McFadden, 1999) have not questioned the ability to extract reliable preference information from *any* infrequently made choice. If such difficulties do arise, they are not limited to survey choice questions but rather influence most of the more interesting choices made by economic agents in private goods and labor markets. Proponents of this view (e.g., Sugden, 1999) tend to see the standard neoclassical economic theory of choice as being incorrect.
14. For a highly informative discussion about how voters make reasonable decisions with incomplete information see Lupia and McCubbins (1998).
15. More generic terms for CV are stated preference method and constructed market. In environmental economics other terms that have been used, usually for specific variants, include bidding game, contingent choice, contingent ranking, and direct questioning. In marketing and transportation applications the terms choice experiment, conjoint analysis, consumer trade-off analysis are often used.
16. For instance, CV critics typically find it troubling that people are willing to pay more to prevent birds from being killed from human causes such as oil spills than they are from natural causes such as disease. This puzzles us because if good stewardship means minimizing human harm to wildlife, and death from natural causes is seen as part of the ebb and flow of nature, the source of the bird mortality lies at the heart of the public's WTP. A private good example may make the point even clearer. Take two otherwise identical cans of tuna and clearly label one "dolphin-safe" and now ask the simple question: is the public's WTP for the two types of cans of tuna likely to be identical?

17. While a researcher should always attempt to design a scenario where all respondents are convinced that the good can be provided, complete success in this task is generally impossible and the degree of success likely to be related to the nature of the good being valued. If the desired probability of provision is one, then failure to convince all respondents that the good can be provided will generally reduce aggregate WTP estimates.
18. While this may induce strategic behavior, we believe that it is generally better to deal directly with this issue. This issue of strategic behavior is discussed at some length in Section 5 below.
19. It should not be surprising that problems with WTP questions (e.g., Kahneman and Ritov, 1994) are concentrated in studies where the good is vaguely described, with no means of its delivery provided, no chance of having to pay for it, and no chance that the response to the question will influence any government decision. Mitchell and Carson (1995) have termed such WTP questions, pseudo CV surveys. They are not likely to have much, if any, economic content because such questions do not represent meaningful economic choices.
20. Development of the survey instrument is often a dynamic process involving focus groups and in-depth interviews to ascertain what information potential respondents want to know about the good, as well as pretest and pilot studies to help measure whether the instrument is performing adequately (Mitchell and Carson, 1995).
21. One of the advantages of doing CV surveys in developing countries is the dramatically lower cost of doing in-person interviews (Whittington, 1998).
22. Over time CV has evolved from largely asking open-ended type questions where the response were taken at face value to other response formats where the statistical assumptions play a much greater role in the estimates obtained. Beyond noting that many apparent problems with CV data are really problems of inappropriate statistical analysis, key issues related to this topic are beyond the scope of this paper. See Hanemann and Kanninen (1999) for a recent review of many of the statistical issues involved in the analysis of CV data.
23. A straightforward formulation of this hypothesis involves asking different statistically equivalent subsamples of respondents about their WTP different prices for the good and examining whether the percent that say yes declines as the amount asked increases. Usually the test is formulated in terms of finding a significant negative price coefficient.
24. While there are earlier origins, much of the recent debate is engendered by Kahneman and Knetsch (1992) and the accompanying commentary by Smith (1992). Kahneman and Knetsch used the term "embedding" to refer to multiple phenomena; some predicted by economic theory; and some not (Carson and Mitchell, 1995).
25. Design factors which tend to mask sensitivity to scope include: (1) vaguely described goods where the descriptions of the goods tend to confuse the smaller (part) and larger (whole) good, (2) questions which emphasize the symbolic nature of providing the good, (3) questions where the underlying metric on which respondents perceive the *larger* good is different from that on which respondents perceived the *smaller* good, and (4) differences in the perceived probability of the different goods actually being provided. Consider the implications of this last factor. Respondents are apt (Fischhoff et al., 1993) to perceive a likelihood of the government actually being able to provide a large good as less than the likelihood of its being able to provide a smaller good. Since a respondent's WTP for a government project should be a function of WTP for the good provided with certainty and the probability the good would actually be delivered (Mitchell and Carson, 1989), the respondent's true WTP for the larger project may actually be lower than that for the smaller project.
26. Less noted with respect to internal scope tests is that they have much greater statistical power than external scope tests because they effectively control for between-respondent variation. This can be seen in the modern variant of the original CV internal scope test which occurs naturally in contingent (multinomial) choice questions where agents are asked to pick their most preferred alternative out of $k > 2$ alternatives. While the availability of such a test is sometimes noted as

an advantage of this format (e.g., Adamowicz, et al., 1999), a test of this nature is unlikely to satisfy the critics.

27. Usually two or more versions of the questionnaire valuing different goods were conducted principally due to logistical considerations, usually the length of time available for the interview, or due to the desire to avoid the potential strategic factors which can come into play in offering multiple levels of a good to respondents.
28. These meta-analyses also provide evidence that even within a general class of goods, the claim sometimes made by CV critics (e.g., Cummings, 1989) that per agent CV estimates are always approximately $X, is wrong. Looking across a wide range of goods provides dramatic evidence against this hypothesis with CV estimates of less than $1 for some forms of outdoor recreation and environmental improvements and estimates of over $5000 for improving some health conditions. The Rosenberger, Loomis and Shrestha (1999) study is an update of the original outdoor recreation meta analysis by Walsh, Johnson, and McKean, (1992).
29. Further, statistical issues in other CV studies that claimed to show scope insensitivity lead one to question the robustness of those results. For instance, Carson (1997) reanalyzes the data from the Schkade and Payne (1993) study, which used the Desvousges et al. (1993b) questionnaire but slower cognitive interviews, and shows there is a significant difference in WTP with respect to the number of birds being killed if two very large outliers are dropped.
30. This situation serves as a good reminder that whenever problems are present in an existing market situation, they are likely to transfer over to a market constructed in a survey.
31. WTP is found by simply integrating (summing) the virtual prices for infinitesimal quantity changes over the discrete quantity change in the rationed good of interest. The relationship between the ordinary income elasticity of demand and the income elasticity of WTP for two goods can be expressed as:

$$\begin{bmatrix} \eta_1^v \\ \eta_2^v \end{bmatrix} = -\begin{bmatrix} \sigma_{11}^d & \sigma_{12}^d \\ \sigma_{21}^d & \sigma_{22}^d \end{bmatrix}^{-1} \begin{bmatrix} \eta_1^d \\ \eta_2^d \end{bmatrix} \frac{y}{e^v}.$$

where η^v are the (virtual) income elasticities of WTP, the σ are the cross price elasticities of demand, η^d are the ordinary income elasticities, and y/e^v is income divided by e^v which is equal to income plus the monetized value of all public goods consumed.

32. Indeed, the discovery of these differences first found in the context of CV studies, coupled with the initial belief that Willig's results were more widely applicable, caused some researchers to claim the observed CV results violated economic theory. Hanemann (1991), in a standard neoclassical theoretical framework, shows that for imposed quantity changes, WTP and WTA can be infinitely far apart due to interactions between income and substitution effects. Kahneman and Tversky (1979) had earlier put forth an alternative they called "prospect theory", whereby losses are valued more highly than gains. In the situation usually considered by economists, a single imposed quantity change, Kahneman and Tversky and Hanemann's models can yield the same results.
33. A problem for CV critics (e.g., Diamond and Hausman, 1994), who contend that theoretically WTP and WTA measures should be close together *and* that CV WTP estimates are gross overestimates, is the standard empirical finding that WTA estimates based on actual transactions are usually much larger than CV WTP estimates for the same good. In particular, Horowitz and McConnell's (1999) recent review of the literature suggests that the ratio of WTA to WTP estimates is roughly the same for surveys and actual transactions.
34. That a price change where the consumer is free to adjust is different from an imposed quantity change where the consumer cannot adjust seems obvious in retrospect. Indeed, it was clear to Hicks (1943) who first clearly developed the concept of utility constant welfare measures. Willig was also careful to specify that he was looking at price changes. This acknowledgement was largely left behind in the rapid incorporation of Willig's work in benefit-cost texts. His work showing that WTP and WTA were close in most likely situations involving price changes and that

the Marshallian consumer surplus measure lay between WTP and WTA, justified the common applied practice of using the Marshallian consumer surplus as adequate approximation to the desired Hicksian measure.

35. The problem of taking account of multiple changes has long been known to be troublesome in the benefit-cost literature (Just, Hueth, and Schmidt, 1982) and is in no way specific to the use of CV estimates. The problem is often ignored in many benefit-cost applications due to the time and expense associated with determining the interactions between the different goods and the belief or hope that such effects are small.
36. The reason for this is that one is destroying existing public goods in a WTA sequence. As one goes farther and farther out in the sequence, fewer substitute public goods remain, and since the agent is compensated with money at each step in the sequence, income is increasing. Carson, Flores, and Hanemann (1998) further show WTP for a good valued first in a WTP sequence is less than WTA for the good valued in any order in a WTA sequence.
37. A typical example is Kahneman and Knetsch (1992). McFadden (1994) makes the argument in a somewhat different way as a conclusion based upon his empirical analysis of a CV data set from a wilderness area study: "These results indicate that either there are extraordinarily strong diminishing returns to preserving additional wilderness areas, or that there is a context effect that makes responses inconsistent with classical economic preferences."
38. For specific examples of context effects on consumer surplus measures for price changes for private goods see Lave (1984) or Randall and Hoehn (1996). The economic intuition behind such effects is the loss in consumer surplus for a specified increase in the price of chicken that is likely to be much larger if there are also large concurrent increases in prices of beef, pork, and fish.
39. At least partially for reasons of agenda control, benefit-cost analysts are almost always told by policy makers to focus on whether a proposed program potentially improves social welfare rather than being asked how to globally maximize societal well-being.
40. Strategic behavior, in economics, is simply utility maximizing behavior. Strategic bias can occur when a respondent's maximizing answer does not represent truthful preference revelation. The possibility of random responses (Converse, 1974; Fischhoff, Slovic, Lichtenstein, 1980) has been raised by psychologists and survey researchers. This issue is related to the possibility of strategic behavior since CV choices cannot be both random and strategic simultaneously. We consider various tests of validity and reliability that relate to the possibility of random behavior in the next section.
41. The NOAA Panel (Arrow et al., 1993) recommended the use of a binary discrete choice question due to its desirable incentive properties. While the NOAA Panel did not specify the conditions necessary for this format to be incentive compatible most are present in the context under which they recommend CV's use.
42. These conditions rule out the possibility that there is (a) some other method of supplying the good, (b) that a better offer might be forthcoming if this one is voted down, or (c) that there is a direct linkage between what happens with respect to this issue and some other issue. The first difficulty can lead to a not favor vote in hopes that someone else will pay for the good's provision. The second can lead to voting against an alternative that improves the status quo in hopes that an even better offer will be made. The third leads to the response being based upon the desired outcome on both issues. These conditions can be violated in both binding referendum and CV survey questions.
43. These three mechanisms, binding referendum, advisory referendum, and advisory survey, are all incentive compatible at the perceived price and quantity. These may, however, differ from the stated price and quantity. The possibility of divergence between the perceived and stated prices and quantities is not unique to these mechanisms, but rather potentially influences almost any consumer choice. The incentive properties of these mechanisms can be adversely impacted if there is a possibility of another program being offered if the proposed project is not put into

place. In this instance, an agent who prefers this alternative program to the proposed program (which in turn is preferred to the status quo) may have an incentive to indicate that the status quo is preferred in order to increase the likelihood that the alternative program will be offered.

44. It can be shown that what a coercive payment vehicle does is to effectively convert a situation whereby an addition to the choice set (e.g., a new public good) looks like a choice between two alternatives, neither of which is a subset of the other, by ensuring the extraction of payment for the good.
45. A related implication of this type of behavior is that when revealed and stated preference data are combined, the vector of coefficients on the attributes should be statistically equivalent up to a "scale" factor. For a recent overview of the issues involved in combining stated and revealed preference data see Henser, Louviere, and Swait (1999).
46. A third type of reliability is often of interest to those without a background in survey research: comparing the results of administering the same survey instrument to two different samples at the same time under the same conditions. More than anything else, this exercise is a test of the adequacy of the sampling plan used to draw the two samples.
47. This high correlation between the valuation estimates for a good based on CV and RP techniques further suggests that if the RP estimates are sensitive to the scope of the good being valued then the CV estimates must also be.
48. The other major investigation into consumer preferences has been by experimental economists. They often reach fairly pessimistic views about violations about properties of agent utility functions typically assumed in neoclassical economics. These violations, however, do not appear to be specific to responses obtained in surveys (e.g., Cameron and Hogarth, 1999; Horowitz and McConnell 2000). They do raise substantive questions about the traditional manner in which economists consider benefits and costs.
49. In one sense the NOAA Panel report (Arrow et al., 1993) has had a negative impact on CV and that is in conveying to some that *all* CV studies need to adhere to the set of standards they proposed. Any study done according to these standards, and particularly following their recommendations with respect to type of sample, elicitation format and scope test, is extremely expensive to conduct. What should be recognized by those doing and using CV studies is that the NOAA Panel's charge was for a very specific situation; natural resource damage court cases, where the government gets reimbursed for the cost of doing the study (plus interest). Challenges to the "admissibility" of the CV study on quality grounds are possible and the government gets rebuttable presumption as to the reliability of its results if the study is admitted. In the more general case, informational costs and benefits should be compared in making choices about where to invest resources to help make better policy decisions.

References

Adamowicz, W. L., P. C. Boxall, J. J. Louviere and J. Swait (1999), 'Stated Preference Methods for Valuing Environmental Amenities', in 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. 460–479.

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 and R. T. Carson (1997), 'Modeling Response Incentives in Dichotomous Choice Contingent Valuation Data', *Land Economics* **73**, 309–324.

Andreoni, J. (1988), 'Why Free Ride? Strategies and Learning in Public Goods Experiments', *Journal of Public Economics* **37**, 291–304.

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

Arrow, K. (1958), 'Utilities, Attitudes, Choices: A Review Note', *Econometrica* **26**, 1–23.
Arrow, K. and E. E. Leamer (1997), 'Comment No. 87,' submitted to NOAA in response to Advanced Notice of Proposed Rulemaking (59 FR 1062, January 7, 1994).
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., A. Munro, B. Rhodes, C. Starmer and R. Sugden (1997), 'Does Part-Whole Bias Exist? An Experimental Investigation', *Economic Journal* **107**, 322–332.
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**, 571–582.
Beattie, J., J. Covey, P. Dolan, L. Hopkins, M. Jones-Lee, N. Pidgeon, A. Robinson and A. Spencer (1998), 'On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part I – Caveat Investigator', *Journal of Risk and Uncertainty* **17**, 5–25.
Becker, G. (1974), 'A Theory of Social Interactions', *Journal of Political Economy* **82**, 1095–1117.
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.
Blamey, R. K., J. W. Bennett and M. D. Morrison (1999), 'Yea-Saying in Contingent Valuation Surveys', *Land Economics* **75**, 126–141.
Blinder, A. S. (1991), 'Why are Prices Sticky: Preliminary Results from an Interview Study', *American Economic Review* **81**, 89–100.
Borcherding, T. and R. Deacon (1972), 'The Demand for the Services of Non-Federal Governments', *American Economic Review* **62**, 842–853.
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.
Brouwer, R., I. H. Langford, I. J. Bateman and R. K. Turner (1999), 'A Meta-Analysis of Wetland Contingent Valuation Studies', *Regional Environmental Change* **1**, 47–57.
Camerer, C. F. and R. M. Hogarth (1999), 'The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework', *Journal of Risk and Uncertainty* **19**, 7–42.
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. (1991), 'Constructed Markets', 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, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*. Boston: Kluwer, pp. 127–163.
Carson, R. T. and N. E. Flores (1996), 'Another Look at "Does Contingent Valuation Measure Preferences?: Experimental Evidence" How Compelling is the Evidence', Discussion Paper 96-31, Department of Economics, University of California, San Diego.
Carson, R. T., N. Flores and M. Hanemann (1998), 'Sequencing and Valuing Public Goods', *Journal of Environmental Economics and Management* **36**, 314–323.
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., N. E. Flores and R. C. Mitchell (1999), 'The Theory and Measurement of Passive Use Value', in 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., 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., T. Groves and M. Machina (1999), 'Incentive and Informational Properties of Preferences Questions', Plenary Address, European Association of Environmental and Resource Economists, Oslo, Norway.

Carson, R. T., W. M. Hanemann, R. J. Kopp, J. A. Krosnick, R. C. Mitchell, S. Presser, P. A. Ruud and V. K. Smith (1997), 'Temporal Reliability of Estimates from Contingent Valuation', *Land Economics* **73**, 151–161.

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: Volumes 1 & 2*. Report to the National Oceanic and Atmospheric Administration.

Carson, R. T., W. M. Hanemann and R. C. Mitchell (1987), 'The Use of Simulated Political Markets to Value Public Goods', Department of Economics, University of California, San Diego, Discussion Paper 87-7.

Carson, R. T. and Y. Jeon (2000), 'On Overcoming Informational Deficiencies in Estimating Willingness to Pay Distributions', Department of Economics, University of California, San Diego, Working Paper.

Carson, R. T., R. C. Mitchell, W. M. Hanemann, R. J. Kopp, S. Presser and Paul A. Ruud. *A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill*. Report to the Attorney General of the State of Alaska, November, 1992.

Carson, R. T., N. F. Meade and V. K. Smith (1993), 'Passive Use Values and Contingent Valuation: Introducing the Issues', *Choices* **8**(2), 4–8.

Carson, R. T. and R. C. Mitchell (1993), 'The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable, and Swimmable Quality Water', *Water Resources Research* **29**, 2445–2454.

Carson, R. T. and R. C. Mitchell (1995), 'Sequencing and Nesting in Contingent Valuation Surveys', *Journal of Environmental Economics and Management* **28**, 155–173.

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*. Northhampton, MA: Edward Elgar.

Carson, R. T., L. Wilks and D. Imber (1994), 'Valuing the Preservation of Australia's Kakadu Conservation Zone', *Oxford Economic Papers* **46**, 727–749.

Carson, R. T., J. L. Wright, N. J. Carson, A. Alberini and N. E. Flores (1995), *A Bibliography of Contingent Valuation Studies and Papers*. La Jolla, CA: NRDA, Inc.

Carthy, T., S. Chilton, S., J. Convy, L. Hopkins, M. Jones-Lee, G. Loomes, N. Pidgeon and A. Spencer (1999), 'On the Contingent Valuation of Safety and Safety of Contingent Valuation: Part 2-The CV/SG 'Chained' Approach', *Journal of Risk and Uncertainty* **17**, 187–213.

Champ, P. A., R. C. Bishop, T. C. Brown and D. W. McCollum (1997), 'Using Donation Mechanisms to Value Nonuse Benefits from Public Goods', *Journal of Environmental Economics and Management* **33**, 151–162.

Chilton, S. M. and G. W. Hutchinson (1999), 'Some Further Implications of Incorporating the Warm Glow of Giving into Welfare Measures: A Comment on the Use of Donation Mechanisms by Champ et al.', *Journal of Environmental Economics and Management* **37**, 202–209.

Converse, P. E. (1974), 'Comment: The Status of Nonattitudes', *American Political Science Review* **68**, 650–660.

Cornes, R., *Duality and Modern Economics*. New York: Cambridge University Press.

Corso, P. S., J. K. Hammitt and J. D. Graham (2000), 'Evaluating the Effects of Visual Aids on Willingness to Pay for Reductions in Mortality Risks', paper presented at the annual meeting of the Association of Environmental and Resource Economists.

Cummings, R. G. (1989), 'Comments Concerning Proposed U.S. Department of Interior Type B Natural Resource Damage Assessment Rules', submitted to U.S.D.O.I. on behalf of ASARCO, Idarado Mining Company and Newmont Mining Corporation.

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.

Deacon, R. and P. Shapiro (1975), 'Private Preference for Collective Goods Revealed Through Voting on Referenda', *American Economic Review* **65**, 943–955.

Deaton, A. S. (1974), 'The Analysis of Consumer Demand in the United Kingdom', *Econometrica* **42**, 341–367.

Debreu, G. (1959), *Theory of Value*. New York: Wiley.

Desvousges, W. H., A. R. Gable, R. W. Dunford and S. P. Hudson (1993a), 'Contingent Valuation: The Wrong Tool for Damage Assessment', *Choices* **8**(2), 9–11.

Desvousges, W. H., F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson and K. N. Wilson (1993b), 'Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability', in J. A. Hausman, ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North-Holland, pp. 91–164.

Diamond, P. A. (1996), 'Testing the Internal Consistency of Contingent Valuation Surveys', *Journal of Environmental Economics and Management* **30**, 337–347.

Diamond, P. A. and J. A. Hausman (1993), 'On Contingent Valuation Measurement of Nonuse Values', in J. A. Hausman, ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North-Holland, pp. 3–38.

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

Diamond, P. A., J. A. Hausman, G. K. Leonard and M. A. Denning (1993), 'Does Contingent Valuation Measure Preferences? Experimental Evidence', in J. A. Hausman, ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North-Holland, pp. 41–90.

Ebert, U. (1998), 'Evaluation of Nonmarket Goods: Recovering Unconditional Preferences', *American Journal of Agricultural Economics* **80**, 241–254.

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

Fischhoff, B., M. J. Quandrel, M. Kamlet, G. Lowenstein, R. Dawes, P. Fishbeck, S. Klepper, J. Leland and P. Stroh (1993), 'Embedding Effects: Stimulus Representation and Response Modes', *Journal of Risk and Uncertainty* **6**, 211–234.

Fischhoff, B., P. Slovic and S. Lichtenstein (1980), 'Knowing What You Want: Measuring Labile Values', in T. S. Wallsten, ed., *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Flores, N. E. (1995), 'The Effects of Rationing and Virtual Price Elasticities', Department of Economics, University of California, San Diego, Discussion Paper 95-20.

Flores, N. E. (1999), 'The Importance of Agenda and Willingness to Pay', University of Colorado, Department of Economics, Working Paper No. 99-30.

Flores, N. E. and R. T. Carson (1997), 'The Relationship Between the Income Elasticities of Demand and Willingness to Pay', *Journal of Environmental Economics and Management* **33**, 287–295.

Freeman, A. M., III (1993), *The Measurement of Environmental and Resource Values: Theory and Methods*. Washington, DC: Resources for the Future.

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

Goodstein, E. S. (1995), *Economics and the Environment*. Englewood Cliffs, NJ: Prentice Hall.

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

Groothuis, P. A., G. Van Houtven and J. C. Whitehead (1998), 'Using Contingent Valuation to Measure the Compensation Required to Gain Community Acceptance of a LULU: The Case of a Hazardous Waste Disposal Facility', *Public Finance Review* **26**, 231–249.

Hammack, J. and G. M. Brown Jr. (1974), *Waterfowl and Wetlands: Toward Bioeconomic Analysis*. Baltimore: Johns Hopkins University Press.

Hanemann, W. M. (1991), 'Willingness to Pay and Willingness to Accept: How Much Can They Differ?', *American Economic Review* **81**, 635–647.

Hanemann, W. M. (1994), 'Valuing the Environment Through Contingent Valuation', *Journal of Economic Perspectives* **8**(4), 19–43.

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

Hanemann, W. M. and B. Kanninen (1999), 'The Statistical Analysis of Discrete-Response CV Data', in 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. 302–441.

Hensher, D., J. Louviere and J. Swait (1999), 'Combining Sources of Preference Data', *Journal of Econometrics* **89**, 197–221.

Harrison, G. W. (1992), 'Valuing Public Goods with the Contingent Valuation Method: A Critique of Kahneman and Knetsch', *Journal of Environmental Economics and Management* **23**, 248–257.

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

Hicks, J. R. (1943), 'The Four Consumer Surpluses', *Review of Economic Studies* **11**, 31–41.

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

Hoehn, J. P. and A. Randall (1989), 'Too Many Proposals Past the Benefit Cost Test', *American Economics Review* **79**, 544–551.

Horowitz, J., and K. E. McConnell (1999), 'A Review of WTA/WTP Studies', Department of Agricultural and Resource Economics, working paper.

Horowitz, J. K. and K. E. McConnell (2000), 'Values Elicited from Open-ended Real Experiments', *Journal of Economic Behavior and Organization* **41**, 221–237.

Howe, C. W. and M. G. Smith (1994), 'The Value of Water Supply Reliability in Urban Water Systems', *Journal of Environmental Economics and Management* **26**, 19–30.

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

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

Johansson, P. O. (1993), *Cost-Benefit Analysis of Environmental Change*. Cambridge: Cambridge University Press.

Jones-Lee, M. W., G. Loomes and P. R. Philips (1995), 'Valuing the Prevention of Non-Fatal Road Injuries: Contingent Valuation vs. Standard Gambles', *Oxford Economic Papers* **47**, 676–695.

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

Kahneman, D. (1986), 'Comments', in R. G. Cummings, D. S. Brookshire and W. D. Schulze, eds., *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman and Allanheld, pp. 185–194.

Kahneman, D. and J. L. Knetsch (1992), 'Valuing Public Goods: The Purchase of Moral Satisfaction', *Journal of Environmental Economics and Management* **22**, 57–70.

Kahneman, D. and I. Ritov (1994), 'Determinants of Stated Willingness to Pay for Public Goods: A Study in the Headline Method', *Journal of Risk and Uncertainty* **9**, 5–38.

Kahneman, D. and A. Tversky (1979), 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica* **47**, 263–291.

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**, 227–237.

Knetsch, J. L., R. H. Thaler and D. Kahneman (1990), 'Experimental Tests of the Endowment Effect and the Coase Theorem', *Journal of Political Economy* **98**, 1325–1349.

Kolstad, C. D. and R. M. Guzman (1999), 'Information and the Divergence between Willingness to Accept and Willingness to Pay', *Journal of Environmental Economics and Management* **38**, 66–80.

Kopp, R. J. (1992), 'Why Existence Value Should Be Used in Cost-Benefit Analysis', *Journal of Policy Analysis and Management* **11**, 123–130.

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

Lansing, J. B. and J. N. Morgan (1971), *Economic Survey Methods*. Ann Arbor: University of Michigan Survey Research Center.

Lave, L. B. (1984), 'Controlling Contradictions Among Regulations', *American Economic Review* **74**, 471–476.

List, J. (2000), 'The Incredible Disappearing Act: The Effect of Market Experience on the WTP/WTA Disparity', paper presented at the European Association of Environmental and Resource Economists Meeting, Rythymnon, Greece.

Loomis, J. B. (1989), 'Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses', *American Journal of Agricultural Economics* **71**, 76–84.

Lupia, A. and M. D. McCubbins (1998), *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* New York: Cambridge University Press.

Madden, P. (1991), 'A Generalization of Hicksian Substitutes and Complements with Application to Demand Rationing', *Econometrica* **59**, 1497–1508.

McConnell, K. E., 'Does Altruism Undermine Existence Value?', *Journal of Environmental Economics and Management* **32**, 22–37.

McConnell, K. E., I. E. Strand and S. Valdes (1998), 'Testing Temporal Reliability and Carry-Over Effect: The Role of Correlated Responses in Test-Retest Reliability Studies', *Environmental and Resource Economics* **12**, 357–374.

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.

Mishan, E. J. (1976), *Cost-Benefit Analysis*, 2nd edn. New York: Praeger.

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

Mitchell, R. C. and R. T. Carson (1995), 'Current Issues in the Design, Administration, and Analysis of Contingent Valuation Surveys', in P. Johansson, B. Kriström and K. Mäler, eds., *Current Issues in Environmental Economics*. Manchester: Manchester University Press.

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

Niewijk, R. K. (1994), 'Misleading Quantification: The Contingent Valuation of Environmental Quality', *Regulation* **60**(1), 60–71.

Olsen, M. (1965), *The Logic of Collective Action*. Harvard: Harvard 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 annual U.S.D.A. W-133 meeting, Jekyll Island, GA.

Portney, P. R. (1994), 'The Contingent Valuation Debate: Why Economists Should Care', *Journal of Economic Perspectives* **8**(4), 3–18.

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

Quiggin, J. (1993), 'Existence Value and Benefit-Cost Analysis: A Third View', *Journal of Policy Analysis and Management* **12**, 195–199.

Randall, A. (1993), 'Passive-Use Values and Contingent Valuation: Valid for Damage Assessment', *Choices* **8**(2), 12–15.

Randall, A. (1998), 'Beyond the Crucial Experiment: Mapping the Performance Characteristics of Contingent Valuation', *Resources and Energy Economics* **20**, 197–206.

Randall, A. and J. P. Hoehn (1996), 'Embedding in Market Demand Systems', *Journal of Environmental Economics and Management* **30**, 369–380.

Randall, A., J. P. Hoehn and G. S. Tolley (1981), 'The Structure of Contingent Markets: Some Experimental Results', paper presented at American Economic Association Meeting. Washington, DC.

Rollins, K. and A. Lyke (1998), 'The Case for Diminishing Marginal Existence Values', *Journal of Environmental Economics and Management* **36**, 324–344.

Rosenberger, R., J. Loomis and R. Shrestha (1999), 'Meta-Analysis of Outdoor Recreational Use Value Estimates: Convergent Validity Tests', paper presented at the annual U.S.D.A. W-133 meeting, Tucson, Arizona.

Rosenthal, D. H. and R. Nelson (1992), 'Why Existence Value Should Not Be Used in Cost-Benefit Analysis', *Journal of Policy Analysis and Management* **11**, 116–122.

Sagoff, M. (1994), 'Should Preferences Count?', *Land Economics* **70**, 127–144.

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

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

Schelling, T. C. (1968), 'The Life You Save May be Your Own', in S. B. Chase, ed., *Problems in Public Expenditure Analysis*. Washington, DC: Brookings Institute.

Schkade, D. A. and J. W. Payne (1993), 'Where Do the Numbers Come From? How People Respond to Contingent Valuation Questions', in J. A. Hausman, ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North-Holland, pp. 271–304.

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. (1992), 'Arbitrary Values, Good Causes, and Premature Verdicts', *Journal of Environmental Economics and Management* **22**, 71–89.

Smith, V. K. (2000), 'Fifty Years of Contingent Valuation', paper presented at the Kobe Conference on the Theory and Application of Environmental Valuation.

Smith, V. K. and C. Mansfield (1998), 'Buying Time: Real and Hypothetical Offers', *Journal of Environmental Economics and Management* **36**, 209–224.

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**, 287–301.

Starrett, D. (1998), *Foundations of Public Economics*. New York: Cambridge University Press.

Sudman, S. and N. M. Bradburn (1982), *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.

Sugden, R. (1999), 'Alternatives to Neo-classical Theory of Choice', in 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. 152–180.

Sutherland, R. J. and R. G. Walsh (1985), 'Effect of Distance on the Preservation of Water Quality', *Land Economics* **61**, 281–291.

Teisl, M. F., K. J. Boyle, D. W. McCollum and S. D. Reiling (1995), 'Test-Retest Reliability of Contingent Valuation', *American Journal of Agricultural Economics* **77**, 613–619.

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

Walsh, R. G., D. M. Johnson and J.R. McKean (1992), 'Benefit Transfer of Outdoor Recreation Demand Studies: 1968–1988', *Water Resources Research* **28**, 707–713.

Willig, R. (1976), 'Consumer's Surplus Without Apology', *American Economic Review* **66**, 589–597.

Whitehead, J. C. and T. J. Hoban (1999), 'Testing for Temporal Reliability in Contingent Valuation with Time for Changes in Factors Affecting Demand', *Land Economics* **75**, 453–465.

Whittington, D. (1998), 'Administering Contingent Valuation Surveys in Developing Countries', *World Development* **26**, 21–30.

Winter, S. G. (1969), 'A Simple Remark on the Second Optimality Theorem of Welfare Economics', *Journal of Economic Theory* **1**, 1969.

Weisbrod, B. A. (1964), 'Collective-Consumption Services of Individual Consumption Goods', *Quarterly Journal of Economics* **78**, 471–477.

Zhao, J. and C. L. Kling (1999), 'Real Options and the WTP/WTA Disparity', working paper, Department of Economics, Iowa State University.