

The Discrete Choice Experiment Approach to Environmental Contingent Valuation

Richard T. Carson

Mikołaj Czajkowski

1. Introduction¹

Assessing the economic desirability of environmental policies requires estimating the value of non-market commodities. In response, several valuation techniques have been developed since the 1960's. They utilize two general data sources—revealed and stated preference data. The former refers to situations where people's choices are observed in actual market situations. Conversely, stated preference data refers to situations in which choices are observed in a survey context.

Stated preference (SP) methods allow collection of information about respondent preferences for the environmental amenities of interest by observing choices in hypothetical situations presented in a survey. Observed choices are contingent on scenarios posed in the survey and the environmental economics literature commonly uses the term *contingent valuation* (CV) to describe the process of utilizing stated preference data for valuation. Additionally, there are many different ways to elicit preference information in a CV study and the one most commonly used are *discrete choice experiments* (DCE).

CV is an inherently more flexible tool than revealed preference (RP) techniques such as hedonic pricing, averting behavior, and the travel cost method, because in principle, it is possible to use it to examine preferences for provision levels of goods that are substantially different from what is currently observed or has been observed in the past. RP data from market behavior tied to the environmental good is frequently unavailable or of limited usefulness due to a limited range of observed variation in the key variables of interest. As such, CV is often the only approach available for providing the economic valuation inputs needed for cost-benefit analysis.²

CV is also usually the only approach to obtain another distinctive property of many environmental goods—the passive use component of their economic value (Krutilla, 1967; Carson, Flores, and Mitchell, 1999).³ Passive use value is sometimes referred to as non-use, existence value or stewardship value and often is defined to include bequest value and

¹ This work draws heavily upon our earlier work, and in particular: Carson, Flores and Meade (2001), Carson and Hanemann (2005), Carson and Groves (2007), Carson and Louviere (2011), Carson (2011) and Carson (2012).

² The public character and free access aspect of many environmental goods often requires government funding or mandated provision. This in turn creates the problem of finding the socially optimal level of their provision—a problem that requires estimates of the value consumers place on the goods. As a result, cost-benefit analysis (and hence CV) has become an essential tool for policy makers concerned with environmental goods.

³ Environmental goods are not unique with respect to having a passive use component. For instance, passive use is thought to be an important component of many cultural resources (Navrud and Ready, 2002) and a public good like national defense is a classic case of a passive value.

option/quasi-option value when uncertainty comes into play. For many environmental goods, consumers may hold positive value for them even though they are not directly or indirectly using them. Passive use is thought to lie behind much of the total economic value for some environmental goods; therefore, its measurement is crucial for policymaking.

It is hard to overestimate the central importance of CV to modern welfare economics. This can be illustrated by over 7,500 papers and studies referenced in Carson (2011) utilizing the method with the largest group focused on environmental valuation. Estimates derived from SP data constitute almost 60% of the estimates in Environment Canada's very large Environmental Resource Inventory (EVRI) database maintained in conjunction with environmental protection agencies in several other countries.⁴ Discussion of CV is now standard in almost all textbooks on environmental economics. Estimates from CV studies are used for assessing very large policies such as the U.S. Clean Water Act (U.S. EPA, 1994) and for a wide range of policy decisions (Griffiths, *et al.*, 2012). CV studies are often done explicitly to look at specific policy issues and the results from older studies are frequently used to evaluate new policy issues in the form of benefits-transfer exercises including serving as meta-analyses inputs.

It is impossible to "review" the CV literature *per se* or even cover all of the major papers in the area in any detail. Instead, given the scope of this handbook, we provide an overview of selected CV and DCE issues which appear the most relevant, and hence, receive relatively much more attention in environmental economics than in other fields. Readers are referred to the other sections of this Handbook for the treatment of issues such as experimental design theory, econometric treatment of discrete choice data, survey administration, or methodological frontiers in the use of choice modeling. We first provide a brief history of CV with an eye toward the increasing dominance of DCE as the preferred elicitation format. We then take up the relationship between different elicitation formats used in CV with an emphasis on the distinction between different DCE formats. From there we move on to a closely related issue—the incentive structure of different elicitation formats. This issue has been examined in some detail in environmental economics, but is only now starting to receive attention in other applied fields. After this we look at neoclassical welfare theory with an emphasis on quantity changes which characterize many

⁴ <https://www.evri.ca/>.

environmental goods rather than price changes which apply in other fields such as marketing. In this section, we address issues related to decomposing maximum willingness to pay (WTP) into different components such as direct and passive use and the role of motives. After our brief theoretical tour, we turn to its predictions on quantities like the difference WTP and minimum willingness to accept (WTA) compensation, the size of the income elasticity of WTP and the magnitude of sequence effects where it was first thought that empirical CV results violated neoclassical theory.⁵ Many of these results are now thought to be generally consistent with neoclassical economic theory, while others have been shown to be behavioral effects that also characterize RP data. From there we turn to a discussion of external and internal tests of scope, an issue which has been at the heart of the debate over the validity of CV. Issues of criterion and convergent validity are then briefly considered. In the last section, we provide some thoughts on where we think contingent valuation using DCE is headed in the future.

2. A short history of contingent valuation

Economists have long realized that much of an individual's utility might be provided by goods for which market prices did not exist. Bowen (1943) and Ciriacy-Wantrup (1947) were the first to propose the use of specially structured public opinion surveys to value what Bowen called "social goods" and Ciriacy-Wantrup called "collective, extra-market goods". The first area where the lack of monetary units for a key output of government projects was considered to be a major problem was outdoor recreation. Once policy makers recognized the need to know what people wanted and how much they were willing to pay for it, they considered surveying the public to find out what they wanted with respect to its national parks. Building the major water projects of the 1950's and 1960's brought into sharp focus the need to value an array of different project inputs and outputs at different points in time which eventually led to modern cost-benefit analysis. The tradeoffs between water-based recreation, electricity and flood control could not be tackled without taking the value of recreation into account. This led to considerable interest in developing both RP and SP methods to reliably place a monetary value on different types of outdoor recreation.

The first application of CV in the academic literature was Davis's 1963 Harvard dissertation which used surveys to estimate the value hunters and tourists placed on a

⁵ Obtaining reliable estimates of WTA tends to be more difficult than for WTP. Throughout we will usually refer to obtaining WTP estimates unless there is need to conceptually distinguish between WTP and WTA.

particular wilderness area. He reasoned that it may be possible to “approximate a market” in a survey by describing alternative kinds of areas and facilities that could be made available to the public, and then simulate market bidding behavior. Davis joined Resources for the Future and a comparison (Knetsch and Davis, 1966) between CV and the new travel cost method showed both approaches produced similar estimates for an outdoor recreation example.

Over the next several years, other economists started to follow Davis’s lead and used CV to value an ever increasing array of public goods.⁶ There were two major developments that are of interest here. The first was the initial applications of CV to value goods that were thought to comprise mostly existence value as defined by Krutilla’s seminal 1967 *American Economic Review* paper. The key paper here is Randall, Ives and Eastman (1974) who looked at changes in visibility related to air pollution in the southwestern United States. Soon researchers were valuing the protection of endangered species (e.g., Samples, Dixon and Gower, 1986) and preserving wild and scenic rivers (Walsh, Sanders and Loomis, 1985). From the policymaking perspective, the U.S. Water Resources Council’s (1979) “Principles and Standards for Water and Related Land Resources Planning” published in the Federal Register set forth the guidelines for determining project benefits. The inclusion of CV as one of the three recommended methods (the other two were the travel cost and the unit day value method) was a sign of CV’s growing respectability. The U.S. Environmental Protection Agency had a major research program focused on developing CV as an approach to non-market valuation in the 1970’s and 1980’s (e.g., Cummings, Brookshire and Schulze, 1986). Mitchell and Carson (1989) provided a comprehensive treatment of a wide range of issues in the design and analysis of CV surveys that was highly influential.

The second major development was the exploration of different ways of eliciting information on preferences from survey respondents. Davis’s work, as well as those of many early CV studies, used a “bidding game”. In a bidding game, a respondent is asked if he is willing to pay a specific amount for the program to supply the good; if he said yes, a higher amount, often \$1 more, is asked and, if no, a lower amount is asked. This is repeated until an

⁶ CV also started to spread outside of environmental economics. See for instance Acton’s (1973) pioneering work on valuing health effects and work by Throsby (1984) on government support for the arts. Unrelated work in transportation would start to have a synergistic effect with environmental economics at a later date with recognition that valuing travel time (e.g., Hensher and Truong, 1985) had similarities to valuing some types of environmental amenities and that transportation policies such as those involving alternative fuel vehicles (e.g., Brownstone, Bunch and Train, 2000) had large environmental implications. One of Louviere’s early papers (1974) looked at how distance (travel time) and stream quality (trout per mile) influenced stream preference.

initial yes changed to a no or vice versa. This format is a binary discrete choice question that is iterated until it effectively obtains a continuous estimate of the respondent's WTP. A difficulty though was soon discovered with the bidding game, that the initial amount asked influenced the final amount. This phenomenon was called starting point bias (Thayer, 1981). Some researchers tried simply asking for WTP using an open-ended question, which became known as a direct question. To overcome the high non-response rate often associated with the direction question, Mitchell and Carson (1981) proposed a payment card whereby respondents were shown an array of numbers on a card and asked to pick the amount on the card or any amount in-between that best represented their WTP. This elicitation format, while not without some of its own problems, performs well in a variety of settings and is the only non-DCE format currently receiving widespread use.⁷

Bishop and Heberlein (1979) put forward an application using goose hunting permits where they compared WTP and WTA with the WTA comparison further featuring a comparison between a survey context and actual re-purchase of the permits. They offered to buy (or sell) goose permits at several prices which were randomly assigned and fit a logit model to the data. The application's single binary discrete choice question immediately caught the attention of other researchers because of its simplicity and close connection with choice behavior in markets with posted prices. Discrete choice elicitation formats for CV got a major boost when Hanemann (1984a; 1984b), starting from McFadden's (1974) random utility model, put forward a comprehensive statistical framework for estimating standard neoclassical Hicksian welfare measures based on the indirect utility function. Cameron and James (1987) and Cameron (1988) for the probit and logit model, respectively, put forward an alternative framework based on expenditure functions that exploits the fact that cost is randomly assigned.⁸ McConnell (1990) showed the two approaches to be duals of the same utility maximization problem when there is no random component but differ in how a random component enters. Hanemann and Kanninen (1999), Haab and McConnell (2002), and Hensher, Rose and Greene (2005) provided comprehensive overviews of econometric issues related to estimating WTP measures.

⁷ Even this format can be cast in a choice context as it can be modeled as interval censored data with the two amounts from the card defining a lower and upper bound on WTP (Cameron and Huppert, 1989). Responses to a binary discrete choice question can also be easily cast as interval censored data. Metcalfe, *et al.* (2012) show that it is possible to combine data from the two elicitation formats in a common statistical model.

⁸ Efforts to estimate models for DCE in WTP space (e.g., Train and Weeks, 2005) exploit Cameron's approach.

The CV study done for the damage assessment for the Exxon Valdez oil spill (Carson, *et al.*, 1992; 2003) used a discrete choice elicitation format with randomly assigned cost. It quickly defined the state of the art.⁹ In response to a critique of CV put forward at an Exxon sponsored symposia contained in the Hausman (1993) volume, the U.S. government convened a panel co-chaired by Nobel Laureates Kenneth Arrow and Robert Solow to assess the CV method and its use in measuring passive use values. The Panel (Arrow, *et al.*, 1993) concluded that “CVM studies can produce estimates reliable enough to be the starting point of a judicial or administrative determination of natural resource damages, including lost passive values.” The Panel made a number of recommendations to enhance the reliability of CV studies including using a single binary discrete choice question, noting its desirable incentive properties when casted in a voting context.

CV researchers had always wanted to extract as much information about the range of policy options with early studies (e.g., Randall, Ives and Eastman, 1974) asking about multiple programs using bidding games, and later studies (e.g., Carson and Mitchell, 1993) using payment cards. With WTP estimates from multiple programs it was possible to estimate valuation functions where one or more attributes of the environmental good had been varied, even going as far as tracing out the response surface with an experimental design in Carson, Mitchell and Ruud (1990).

CV researchers started to become aware of incentive problems with bidding games and open-ended type CV survey questions (Hoehn and Randall, 1987; Mitchell and Carson, 1989) and with offering multiple levels of a pure public good when only one level of a public good can be provided. This growing awareness of incentive issues though did not initially discourage researchers starting with Carson (1985) and Hanemann, Loomis and Kanninen (1991) from exploring how to obtain more information from each respondent about one level of a good. This was done by asking a double-bounded binary choice question which asked respondents who said yes to the amount in an initial binary choice question about whether they would pay a higher amount and those said no about a lower amount. This

⁹ The study estimated that the U.S. population was willing to pay approximately three billion dollars to prevent a spill similar to the Exxon Valdez in the future. Natural resource damage assessment rules require compensation to come in the form of restoration and on expenditures to provide similar resources to those injured where restoration is not possible. Exxon spent about two billion dollars on response and restoration and provided the government with about one billion dollars for a fund to buy additional natural resources. Subsequent to the Exxon Valdez CV study, the government implemented a prevention plan similar to that described in the study and key components of that plan have been deployed several times to avert a spill.

format only involved one good and binary choice questions. The response, if taken at face value, reduces the length of the interval in which the respondent's WTP lay, and hence, decreases the confidence interval around WTP estimates given a fixed sample size.

To get even more information from respondents, researchers began implementing more complex DCE formats involving either multiple choice sets, multinomial choice questions, or both. The first environmental economics paper using such an approach was Carson, Hanemann and Steinberg (1990) which asked about options involving Kenai King salmon fishing and fit a nested logit model that included a *status quo* no-purchase option.¹⁰

The paper though that generated attention in the environmental economics community was Adamowicz, Louviere and Williams (1994) that gather what was then thought to be an impossible amount of data, 16 choice sets each with 3 alternatives. The paper also showed how to combine choice data from a DCE with choice data involving recreational trip destinations that allowed for differences in the variance of the error component in the utility function estimated for the two types of data. A later paper by Adamowicz, *et al.* (1998) demonstrated more complex DCE formats could also potentially be applied to situations involving passive use. An influential paper by Hanley, Wright, and Adamowicz (1998) summarized the quickly growing number of studies using DCE to explicitly value program attributes. Adamowicz wrote a chapter on using DCE for environmental valuation that appeared in the seminal Louviere, Hensher and Swait (2000) book.

At this point confusion in the environmental economics literature starts to set in as to what was CV and what was a DCE (often then referred to as a choice experiment [CE] or choice modeling [CM]). Adamowicz, *et al.* (1998) one of the first papers to draw a distinction between CV and DCE was careful to note "the most common elicitation approach is the two-alternative (referendum) contingent valuation method (CVM) (see Mitchell and Carson [1989]). Other stated preference presentations are also possible. In this article we examine

¹⁰ An early effort (Rae, 1982) involving air pollution had respondents completely rank order a set of programs but was shown to have substantial econometric problems given the techniques of the time. Other periodic efforts to implement conjoint related techniques including ratings of alternatives (e.g., McKenzie, 1993) have not been widely copied in the environmental economics literature due to the lack of a well-accepted way of exploiting any extra preference information beyond that contained in choice data. In spite of frequent reference to the term "conjoint" to refer to DCE, Louviere, Flynn and Carson (2010) recommend not using this term. Conjoint measurement as proposed in the psychological literature is substantively different than the random utility theory and Lancasterian attribute framework that underlies DCE. Choice-based conjoint was a term used to try to move marketers accustomed to using ad hoc approaches based on conjoint measurement theory over to using DCE. It would appear to have long outgrown its usefulness even in marketing.

an extension or variant of this traditional approach, the choice experiment, which employs a series of questions with more than two alternatives to elicit responses that allow estimation of preferences over attributes of an environmental state.” Thus, to some researchers CV was defined to mean a single binary discrete choice question while DCE were defined to mean a sequence of multinomial choice questions. In contrast, Hanley, Wright and Adamowicz (1998) use a series of open-ended direct questions as the definition for CV and compared this to a set of multinomial choice questions they labeled as “a choice experiment”. A sequence of binary choice questions was sometimes called CV and but also sometimes referred to as being a choice experiment rather than CV.

In part, an ill-defined distinction between CV and DCE took hold in the environmental economics literature because arguments were put forward: (a) that DCE with more than two alternatives or multiple choice sets overcame some of the problems critics (e.g., Diamond and Hausman, 1994) asserted existed with CV, (b) that DCE represented a new approach taken from marketing (which fostered journal publication), and (c) that there was a large demand from policy makers for valuing changes in attributes. DCE as distinct from CV were ill-defined because CV was never a particular elicitation method as opposed to the use of SP data to estimate well-defined economic welfare measures for public goods (Mitchell and Carson, 1989).¹¹ Early CV studies had focused on valuing multiple levels of environmental quality and some DCE simply expanded that to looking simultaneously at varying a substantial number of non-cost attributes rather than one or two. Over time it also became clear that DCE more complex than a single binary choice question could never overcome the problems critics perceived with that format. That is because with enough statistically equivalent samples a single binary discrete choice format can be used to estimate anything that a more complex DCE can but without order effects. The advantage of more complex DCE was that it gathered considerably more preference information per respondent and, as such, had considerable cost advantages. The ability to easily vary multiple attributes made them popular with policy makers who needed to consider a wide range of options.¹² One currently sees a variety of different DCE elicitation formats in use

¹¹ The term “contingent” refers to the estimated values obtained using the data collected being contingent on the features of the survey’s scenario, or constructed market as it is sometimes referred to , for providing the good(s) including the specific description of the good(s) and their manner of their provision.

¹² Overviews include Louviere, Hensher and Swait (2000), Bennett and Blamey (2001), Hanley, Mourato and Wright (2001), Bateman, *et al.* (2002), Champ, Boyle and Brown (2003), Kanninen (2007), and Hoyos (2010).

with choice of the particular elicitation format being driven by the specific characteristics of the good being valued, available funding, and the range of policy options that needed to be evaluated.

3. A typology of stated preference elicitation formats used in CV

Carson and Louviere (2011) provide a nomenclature for organizing different elicitation formats used in CV. The first aspect of this nomenclature is to note that, while there are many ways to ask stated preference questions, the only two approaches that result in data consistent with neoclassical welfare economics are those that involve either matching formats or choice formats.¹³

A matching question essentially solves the problem of what quantity makes the consumer indifferent between two situations, typically the *status quo* and an alternative. The quantity that makes the agent indifferent effectively sets utility in the two situations. When the quantity is some standard numeraire like money, the truthful response to a matching question corresponds to one of the Hicksian welfare measures. This, of course, would have great advantages because it would short circuit much of the need for making assumptions about the structure of utility functions and error components. Obtaining a reliable answer to a matching question has proven difficult for two reasons. First, a direct question (DQ) for WTP tends to yield high non-response rates, most likely because agents in many countries are used to making decisions in markets with posted-price. Second, the incentive structure of a DQ is not conducive to truthful preference revelation.

Contemporary neoclassical consumer theory is built on the ability to make choices between options. Through a series of steps, this eventually leads to the indifference point being arrived at even though agents did not originally know their WTP or WTA.¹⁴ If one assumes truthful preference revelation and an absence of anchoring effects, this naturally leads to Davis's bidding game (BGAME) as the way to obtain an estimate arbitrarily close to the desired matching response. Recognition that it was costly for respondents to exert the effort to formulate the matching response and that this could induce anchoring effects, leads

¹³ For some commonly used formats such as ratings questions, it is sometimes possible to throw away "extra" information on the intensity of preferences and assume that they can be translated into the choices that consumers would make. Likewise, rankings data can typically be translated into choice data.

¹⁴ Some work in psychology and behavioral economics implicitly assumes that neoclassical economic theory requires agents to carry around a complete vector of WTP and WTA for all possible situations and that they can effortlessly retrieve those numbers but this has never been required by neoclassical economic theory.

to the payment card (PCARD) as a way to get reasonably close to the desired matching measure. BGAME and PCARD formats move away from a DQ toward having respondents answering one or more simpler choice questions. DQ, BGAME, and PCARD elicitation formats all been used valuing both a single (SINGLE) good and a sequence (SEQ) of goods.

After Bishop and Heberlein's seminal work, a single binary choice (SBC) question seemed to be a more natural way for environmental economists to have respondents provide preference information. They were happy to replace matching elicitation formats with the SBC as long as it was possible to efficiently convert that information into the standard Hicksian welfare measures they needed. The SBC is the canonical form of a DCE and it may be useful to step back and formally define the two essential elements to any DCE. The first is that the agent is asked to make a discrete choice between two or more alternatives in a choice set. The second is that the alternatives presented for choice are constructed by means of an experimental design that varies one or more attributes to be able to estimate economic quantities tied to preference parameters. The SBC format randomly assigns the cost of the second choice, with the cost of first good (typically the *status quo*) held constant (often at zero), which allows for the estimation of the relevant Hicksian welfare measure.

Policy makers, however, often want to value more than one possible change from the *status quo*. This is straightforward to do with the SBC format. If there are j possible variants of the goods that policy makers are interested in, ask j statistically equivalent subsamples a SBC question involving the first/*status quo* good versus one of the j alternatives. The data from each subsample can be stacked and alternative specific constants (ASC) of each of the j goods can be estimated along with different slope parameters for the variants of the good. This will produce estimates for each of the j goods. If the j goods differ from each other in some systematic way, this can be accommodated by replacing or augmenting the ASC's with a set of attribute variables describing those differences. While this approach can have desirable properties from the perspective of both survey design and providing desirable incentive properties for truthful preference revelation, it can also be quite expensive, and prohibitively so if j is sizeable because SBC responses reveal little preference information.

There are two ways to enhance the amount of preference information obtained from a single respondent within a discrete choice framework asking a sequence of question or enlarging the choice set. A variant of the first has already been noted – ask about a second cost amount that conditions on the first response. This elicitation format, known as the

double bounded binary choice (DBBC), is sometimes called a double bounded dichotomous choice question. It yields interval censored data and reduces the sample size needed for a given confidence interval. This reduction can be quite large with a good experimental design if the response to the second question is consistent with the WTP distribution implied by the first question and substantial research has revolved around the veracity of this assumption (Alberini, Kanninen and Carson, 1997). This is the simplest of the sequential elicitation approaches as it introduces a second choice set without changing any attribute of the good other than cost.

The second way to obtain more preference information is to ask for the choice to be made from a larger set of alternatives. A single multinomial choice (SMC) with $j \geq 3$ effectively generates $j-1$ binary comparisons between the alternative chosen and all of the j possible alternatives. In principle, an enormous amount of preference information could be obtained from a single agent by using a choice set with an extremely large number of alternatives. The belief that there are severe limits to the number of choice alternatives that agents will seriously entertain leads one in the direction of quite strict limits on the j . This in turn implies asking the same agent to choose their most preferred alternative from more than one choice set or asking different agents to make choices in SMC questions.

The belief that different agents may have heterogeneous preferences and that these may play a role in policy decisions lead researchers to sample multiple agents. Efficiency concerns coupled with cost constraints tend to lead researchers to ask respondents a sequence of choice sets. The two most common elicitation techniques using a sequence of choice sets are a sequence of binary choice questions (BC-SEQ) and a sequence of multinomial choice questions (MC-SEQ). The key issue with all sequential formats is whether respondents answer the choice sets independently.

Other sequential formats that divide alternatives into more and less preferred subsets are also possible. An early example is a complete ranking (RANK) of a set of j alternatives. A more recent example known as best-worse choice (BWCHOICE) asks respondents to indicate their best alternative and their worst alternative among a set of $j > 2$ alternatives (Marley, Flynn and Louviere, 2008) and can be seen as a less burdensome version of ranking that extracts more information than a SMC.

4. Elicitation formats and incentive compatibility

Environmental economists have paid considerable attention to the incentive properties of CV elicitation formats. This appears to be due to two factors. First, the use of multiple elicitation formats in CV studies lead researchers to explore the properties of responses to individual formats (e.g., Thayer, 1981) and to compare responses from different formats (e.g., Smith and Desvousges, 1986). Finding that responses to specific formats did not behave as expected and that different formats yielded different estimates of WTP dismayed researchers. This suggested to some that respondents had problems valuing unfamiliar public goods and the divergence between these estimates is a major reason why some CV critics argue estimates from SP questions should be rejected (Hausman 1993; McFadden 1994).

The second was recognition that underlying the belief that different elicitation formats should yield statistically equivalent estimates of WTP was the assumption that respondents always truthfully reveal their preferences irrespective of the economic incentives to do so. This, of course, is an unusual assumption for an economist to make. Indeed, Paul Samuelson (1954) in his classic article defining the properties of pure public goods saw strategic behavior, in the form of a downward bias, was the main problem in using surveys to determine public goods provision and taxation.

Carson and Groves (2007) have put forth a comprehensive neoclassical framework use mechanism design theory to examine the incentive and informational properties of stated preference elicitation formats. The first result is that if respondents view the stated preference question(s) as consequential, then their outcomes can be analyzed as revealed economic behavior. To be consequential, a survey question needs to meet two properties. Agents need to view responses as having (a) a positive probability of influencing agency's actions and (b) agents must care about the outcomes. SP questions that do not meet these two conditions are inconsequential. The difficulty with inconsequential questions is that a random response is as good as any other response because the response cannot influence the agent's utility.

For consequential survey questions, mechanism design theory can be used to analyze the optimal responses. Carson and Grove's (2007) starting point is the well-known Gibbard-Satterwaite theorem that says that no mechanism with larger than a binary message space can be incentive compatible without restricting the space of allowable preference functions.

This means that all elicitation formats other than a SBC can present situations where it is optimal for some respondents to reveal preference information inconsistent with true WTP.

Further, conditions must be imposed for an SBC question to be incentive compatible. The basic conditions are well known from the voting literature on a single binary vote (Farquharson, 1969). To be incentive compatible, a referendum on a pure public good needs to be a take-it-or-leave-it offer, where the vote doesn't influence any other offers that may be made to agents and where the payment mechanism is coercive in the sense that each agent can be required to pay independently of how the individual agent voted. These conditions are quite restrictive. For instance, they rule out private goods because the government cannot force individual agents to buy private goods but the government can force agents to pay taxes. For a quasi-public good like a recreational fishing lake, it is possible to structure a SBC between the current *status quo* quality level and price and an alternative quality level and price that is incentive compatible with respect to revealing which of the two configurations the agent prefers. However, like private goods, this choice will not reveal information about the number of trips that would be made under either configuration. It is straightforward to show that a question that can influence multiple outcomes cannot be incentive compatible with respect to one outcome because agents should take into account the influence of their response on all possible outcomes when formulating it.

If a binding referendum vote on a single issue, meeting the auxiliary conditions noted above, is incentive compatible, what about an SBC question meeting the same auxiliary conditions? Carson and Groves (2007) show that the binding nature of a referendum vote is not needed. All that is necessary is that the government be more likely to undertake the action if the vote meets a specific plurality (typically a simple majority or two thirds in favor). Such advisory referenda are common in many places. Is the plurality voting rule necessary? The answer is no. Carson and Groves (2007) show that a weaker condition, that the probability of the alternative being implemented is weakly monotonically increasing in the percent in favor, works. This condition includes plurality voting as a special case. Green and Laffont (1978) have shown that any economic mechanism of the type being considered here can be implemented using a statistical sample. Putting these together yields a SBC in an advisory survey that has the same property as a binding referendum vote.

Carson and Groves (2007; 2011) show that it is possible to examine the incentive structure of other CV elicitation formats. Starting with the BGAME, if the starting point used

is thought by a respondent to convey information about the possible cost of the program, then it is optimal to condition his response on it. Thus, WTP estimates from bidding games should be correlated with the starting points used, which is what is empirically observed. DQ and PCARD responses should be consistent with observing a concentration of responses at zero coupled with observing relatively few very small WTP amounts. This should happen because the optimal response for respondents whose WTP is less than the expected cost is zero.¹⁵ Another theoretical prediction is that two WTP distributions revealed by a DBDC should not be perfectly correlated as the standard interval censored estimator assumes.

It is possible to generate a more specific prediction about the properties of particular elicitation formats with stronger assumptions about respondent beliefs. In more complex DCE like SMC, BC-SEQ and MC-SEQ, it is often harder to predict the impacts of incentive and information effects beyond noting that they should generate violations of the independence of irrelevant alternatives (IIA) assumption without more structure, but some flavor the nature of problems likely to be encountered can be illustrated.¹⁶ In an SMC (and by extension both the BC-SEQ and MC-SEQ formats), the fundamental problem with pure public goods is that only one level can be supplied and it is the same for all agents. From the theoretical literature on voting, it has long been known that truthful preference revelation is often not optimal. If the agent has non-uniform priors over the two alternatives most likely to be chosen by other people, then it is generally optimal for the agent to choose the most preferred of these two alternatives, even though their true preferences may be for another alternative.¹⁷ In contrast, the SBC question has a dominant strategy that does not depend on beliefs about the preferences of other agents. There may be cases where respondents have flat priors over their beliefs about the preferences of other agents for all of the available alternatives in which case truthful preference revelation is optimal, but this is a very strong assumption. What is true in all of these cases is that in a consequential survey, the agent should not be picking alternatives randomly rather they should be picking a “good” choice,

¹⁵ An interesting aside is that the popular Becker, DeGroot and Marschak (1964) mechanism, and other similar mechanisms that in principle can elicit an incentive compatible matching response with weak restrictions on preferences, cannot be implemented in a survey context. This is because they rely on the agency not being able to exploit the extra information beyond a binary response, which cannot be guaranteed in a survey context.

¹⁶ There are, of course, other reasons why IIA violations may occur.

¹⁷ In a three candidate race where the candidate with the largest number of votes wins, the candidate perceived to be in last place tries to convince voters that their chance of winning is higher and more uncertain than thought and that there is no effective difference in the positions of the two leading candidates. Success in this strategy makes voters who truly favor the third place candidate more likely to vote for that candidate.

indeed the optimal choice given the elicitation format and belief structures. The difficulty is that this choice may not reflect the agent's unconditionally most preferred alternative, as is typically assumed in most analyses. This sort of behavior can under very strong conditions yield marginal estimates of WTP for changes in attributes which are consistent even though estimates of total WTP for a good may be biased. Agents want the agency to supply their most desired mix of attribute levels (not a random mix) with most of the strategic behavior focused on influencing the agency's pricing or overall provision decision which tends to show up in the estimates for the cost parameter or the ASC on the *status quo* option. More generally strategic effects can mimic or reinforce learning and behavioral effects as well as influence scale in sequential DCE.

5. Economic theory and contingent valuation

5.1 Economic welfare measures

The goal of a CV study is to measure an individual's monetary value for a change in one or more goods of interest. No good can be valued in an economic context independent of how it is delivered and paid for.¹⁸ As such, the first key concept is that it is the policy (plan) to deliver the good that the agent is valuing; and, we will continue to talk about providing a good, except when it is useful to explore how the policy perspective might be important. Sometimes there is only one policy alternative to the *status quo* under consideration in which case, standard neoclassical economic theory looking at the utility gained from the addition of a single additional good to the utility function is applicable. In other instances, policy makers are interested in exploring options that differ from each other along one or more well-defined dimensions or attributes. When this is the case, Lancaster's theory of consumer choice (Lancaster, 1966) is often brought to bear with the standard goods of neoclassical economic theory replaced with bundles of attributes. Policy discussions are often interested in how value for the good changes as one or more of a good's attributes are varied in either a continuous or discrete manner. The monetary value of the marginal change in only one of the good's attributes is often referred to as the 'implicit price' of the attribute. In this sense, attribute-based DCE models bear some resemblance to hedonic pricing models.

¹⁸ In a marketing context, imagine the difference in demand for the same product in a store with and without a sign that said the product had been made by exploiting workers under terrible conditions.

Let us begin by denoting the item being valued (be it the change in the amount of a composite good or the change in one of its attributes) by q , and assume the individual has a utility function defined over the quantities of various market commodities denoted by the vector x , and q , $u(x,q)$. Corresponding to this direct utility function, we can write an indirect utility function, $v(p,q,y)$, where p is the vector of the prices of the market commodities and y is the person's income. We make the conventional assumption that $u(x,q)$ is increasing and quasi-concave in x , which implies that $v(p,q,y)$ satisfies the standard properties with respect to p and y .¹⁹ We make no assumptions regarding q . If the agent regards q as a "good", $u(x,q)$ and $v(p,q,y)$ will both be increasing in q ; if he regards it as a "bad," $u(x,q)$ and $v(p,q,y)$ will both be decreasing in q ; and if he is indifferent to q , $u(x,q)$ and $v(p,q,y)$ will both be independent of q . We also make no assumption regarding quasi-concavity with respect to q .

The act of valuation implies a contrast between two situations—a situation with the item, and one without it. We interpret what is being valued as a change in q . Specifically, suppose that q changes from q^0 to q^1 ; the person's utility thus changes from $u^0 \equiv v(p,q^0,y)$ to $u^1 \equiv v(p,q^1,y)$. If he regards this change as an improvement, $u^1 > u^0$; if he regards it as a change for the worse, $u^1 < u^0$; and if he is indifferent, $u^1 = u^0$. The value of the change to him in monetary terms is represented by the two Hicksian measures, the compensating measure, C , which satisfies

$$v(p,q^1,y-C) = v(p,q^0,y), \quad (1)$$

and the equivalent measure, E , which satisfies

$$v(p,q^1,y) = v(p,q^0,y+E). \quad (2)$$

Observe that

$$\text{sign}(C) = \text{sign}(E) = \text{sign}(u^1 - u^0). \quad (3)$$

If the change is an improvement, $C > 0$ and $E > 0$; and C measures the individuals' WTP to secure the change, while E measures his WTA to forego it. If the change is regarded as being for the worse, $C < 0$ and $E < 0$; in this case, C measures the individuals' WTA to endure the change, while E measures his WTP to avoid it. If he is indifferent to the change, $C = E = 0$.

To emphasize the dependence of the compensating and equivalent measures on (i) the starting value of q , (ii) the terminal value of q , and (iii) the value of (p,y) at which the

¹⁹ $v(p,q,y)$ is homogeneous of degree zero in p and y , increasing in y , non-increasing in p , and quasi-convex in p .

change in q occurs, we sometimes write them as functions: $C = C(q^0, q^1, p, y)$ and $E = E(q^0, q^1, p, y)$. To simplify things, we can define the WTP function,

$$\text{WTP}(q^0, q^1, p, y) = \begin{cases} C(q^0, q^1, p, y) & \text{if } C \geq 0 \\ -E(q^0, q^1, p, y) & \text{if } C \leq 0 \end{cases} \quad (4)$$

The WTA function, $\text{WTA}(q^0, q^1, p, y)$, is defined analogously. The goal of a CV survey is to measure one or another of these valuation functions – either the entire function, or one or more particular points on the function.

5.2 Random utility framework

If one could reliably observe WTP and WTA in matching questions, analysis in CV surveys would be straightforward in the case of a single policy change of interest and involve a standard response surface regression approach in cases where it was of interest to trace out the implicit price curves for individual attributes. For reasons noted earlier, DCE are used to elicit preference information. Because choices rather than the Hicksian welfare measures themselves are observed, one must turn to a model that translates choices into preference parameters related to underlying utility. Typically this is done in a RUM framework.

The standard economic version of the RUM assumes that while an individual knows his preferences with certainty and does not consider them stochastic or otherwise behaving in a random manner, his preferences contain some components that are unobservable to the econometric investigator, and thus, are treated by the investigator as random (McFadden, 1974; Manski, 1977). In contrast, the version of the RUM model from psychology, starting with Thurstone's original 1927 presentation in the context of a probit model, envisions a true stochastic component related to choice, couched in the inability to perfectly discriminate between alternatives. The economic version of the RUM does not rule out respondent uncertainty. Indeed, in most decision contexts including market purchases, respondents face some uncertainty about various aspects of the good. The difference between these two perspectives can best be seen by considering data generated by the DBBC elicitation format in the absence of any information or incentive effects. In the Manski formulation, because there is no true random component, the WTP distributions implied by the two binary choice questions should have the same mean, variance, and be perfectly correlated. In the Thurstone formulation, the means and the variances of the two WTP distributions should be the same but the correlation between the responses will be less than

one.²⁰ Empirical evidence from fitting bivariate probit models to DBBC data tends to reject the hypothesis that the correlation coefficient is one, but there is considerable debate over the cause of the rejection ranging from anchoring to strategic behavior to being artifact of fitting the wrong distribution for WTP to the data (Versoni, Alberini and Cooper, 2011).

Representing the stochastic component of utility function by ε , we write the indirect utility function as $v(p, q, y; \varepsilon)$. It is usually assumed that the stochastic component enters the utility function additively, so the utility function can be decomposed to the part which is deterministic and the part which is random: $v(p, q, y; \varepsilon) = \mathcal{G}(p, q, y) + \varepsilon$. Empirical implementation of random utility theory requires making an assumption about statistical properties of the random term, typically that the random component of the utility function is independently and identically (IID) distributed across individuals and alternatives and, for computational reasons, the Extreme Value Type 1 distribution is commonly used. The resulting multinomial (conditional) logit model (MNL), has the convenient closed-form expression of the probability of an individual i choosing alternative j from a set of all available alternatives J :

$$P(j | J) = \frac{\exp(\mathcal{G}(p_{ij}, q_{ij}, y_{ij}))}{\sum_{k=1}^J \exp(\mathcal{G}(p_{ik}, q_{ik}, y_{ik}))} \quad (5)$$

The structure of the conditional logit model leads to restrictive IIA property, which while having intuitively desirable features at the individual level, frequently does not hold empirically at the sample level. There are many ways to relax this restriction by allowing agents to differ either with respect to their taste parameters and/or their random component. We refer interested readers to other chapters of this Handbook that are focused on statistical modeling issues for relevant discussions.

5.3 Individuals' motivation and disaggregation of total economic value

So far we have made no assumption about the individual's motive for valuing q , or the nature of this value. This is because motives are essentially irrelevant for the neoclassical economic theory of value—economists simply accept consumer sovereignty. Economists do not judge if consumers' preferences are right or wrong. Our objective here is to observe

²⁰ In the Thurstone framework, different variances could be generated if different parts of the stimulus space being explored involved differential ability to discriminate between alternatives.

choices and from them infer consumers' relative preferences for different goods in monetary units.

Whatever the reason why an individual cares for q , if he does care, this is reflected in his direct and indirect utility functions $u(x,q)$ and $v(p,q,y)$. However, this is not to say that motives are irrelevant at all. In fact, understanding the motives may help policy makers to interpret the estimated economic value correctly. There is a large literature on motives and their utility representation (Mitchell and Carson, 1989; Carson, Flores and Mitchell, 1999).

Originally researchers defined total value as one of the four Hicksian welfare measures and then distinguished between use and non-use (Mitchell and Carson, 1989), but the tendency now is to distinguish between use and passive use (Carson, Flores and Mitchell, 1999). The difference involves how to treat uncertainty which invokes option value. The DC Court of Appeals in the 1989 *Ohio v. U.S. Department of Interior* ruled that the government should be measuring passive use values in natural resource damage assessments, where it effectively defined passive use values to include Krutilla's existence value and other types of value that were unlikely to leave a trace in market behavior. While it is potentially possible for option value to manifest itself in market transactions, such markets rarely, if ever, exist. Passive use value then is comprised of existence value including variants such as stewardship value, bequest value, option value and quasi-option value.²¹ Each of these can be explained in terms of a conceptualization of what is being valued and a specific formulation of an individual's preference structure.

5.3.1 Use value

The key characteristic of use value is some type of technical relationship with some marketed good. One classic case is where consumption of some marketed good is required to enjoy the environmental good of interest. Travel cost analysis is one technique that exploits such a relationship. The other classic case is where the environmental amenity of interest is not directly sold but bundled into some marketed good where hedonic pricing can be used. Some environmental goods examined by environmental economists are comprised of mostly, if not exclusively to use value. Analysis of DCE in this instance more closely resembles their counterparts in other areas of applied microeconomics.

²¹ Option value and quasi-option value involve the introduction of uncertainty over future state of the world and the role of potential learning. They are beyond the scope of this chapter. Carson and Hanemann (2005) and Zhao and Kling (2009) provide detailed discussions.

5.3.2 Existence and stewardship values

Existence value was a term coined by Krutilla (1967) who articulated the current conceptual framework and, in doing so, fundamentally changed environmental economics. Krutilla observed that people might be willing to pay for an environmental resource—say a remote national park, even though they knew they would never visit it because, as he famously put it, they “obtain satisfaction for the mere knowledge that part of the wilderness in North American remains, even though they would be appalled by the prospect of being exposed to it.” While earlier economists had grappled with the notion of public goods that were truly pure public goods in the sense that people could not be excluded from using them, nor did their use create congestion externalities, Krutilla’s key insight was that gaining utility from such goods would not leave any behavioral traces in market activity. This means that existence value cannot be measured with techniques like the household production function or hedonic pricing that rely on some type of technical relationship like weak complementarity between the non-marketed good and one or more marketed goods. It is possible an agent may value a public good for both direct and passive use considerations as they are not mutually exclusive, and indeed, may be positively correlated.

Krutilla’s emphasis on potentially gaining utility from the simple existence of a good starkly set out why relying on inference from market data might substantially undervalue some environmental amenities. Other motivations lead to the same utility specification. Stewardship is one of these and occurs if people believe the government should be a “good steward” of a resource and are willing to pay to see the government undertake costly actions consistent with being a good steward. Expressed another way, some people perceive a duty to protect some entity from harm imposed by human activity.²² Good stewardship can, among other things, involve preservation of endangered species, setting aside land rather than developing it, or ensuring that water quality in rivers is maintained at a high level.²³

²² Being a good steward can mean being willing to spend money to keep animals from being killed by human related causes but not natural causes. This conceptually causes no problem from a theoretical perspective, but utility is not defined on the number of animals but rather on the animals and the source of the threat to them.

²³ Pure altruism and an impure form known as warm glow (Andreoni, 1989) are sometimes advanced as other motives for environmental values. From the perspective of economic theory, the source of the motivation at the individual level is irrelevant. From the perspective of aggregating welfare measures, interdependence of utility across agents of which altruism is a special case, can be issue but not in most cases considered in environmental valuation. Carson and Hanemann (2005) provide a detailed discussion of issues related to altruism.

Several ways have been proposed to represent existence value in utility theoretic terms. We use the most common representations here and start by assuming the direct utility function takes the specific form of:

$$u = u(x, q) = T[\bar{\phi}(x), q], \quad (6)$$

where $T[.,.]$ is a bivariate function, and $\bar{\phi}(\cdot)$ is a sub-function that aggregates the x 's. (6) involves weak separability between the x 's and q since the marginal rate of substitution between the consumption of any pair of market commodities, x_n and x_m , is independent of the level of q . A consequence of (6) is that the ordinary demand functions for the x 's are each independent of q ; they take the form:

$$x_n = h^n(p, q, y) = \bar{h}^n(p, y) \quad n = 1, \dots, N, \quad (7)$$

where the functions $\bar{h}^n(\cdot)$ are the ordinary demand functions associated with maximization of the sub-utility function $\bar{\phi}(\cdot)$ alone: maximizing $u(x, q)$ and $\bar{\phi}(x)$ subject to a budget constraint on the x 's leads to exactly the same solution. The corresponding indirect utility function is:

$$u = v(p, q, y) = T[\bar{\psi}(p, y), q], \quad (8)$$

where $\bar{\psi}(p, y)$ is the indirect utility function that corresponds to maximization of the sub-utility function $\bar{\phi}(\cdot)$ alone. While the person cares for q , it enters his direct and indirect utility functions via the $T[.,.]$ function and the presence of q in his utility function has *no* effect on his utility maximizing choice of the x 's. It is in this sense that one could say that this individual values q for reasons that are *unconnected* with his valuation of the market goods x . In this case, q would be said to have a pure existence value.

Now consider a modified version of the above utility function:

$$u = u(x, q) = T[\bar{\phi}(x, q), q], \quad (9)$$

where, as before, $T[.,.]$ is a bivariate function and $\bar{\phi}(\cdot)$ is a sub-function. In this case, q enters the utility twice, once through its appearance in $\bar{\phi}(\cdot)$ and the other as the second argument in $T[.,.]$. Here, the individual values q for two reasons: the first is connected with his consumption of the x 's and is represented by the interaction of x and q in $\bar{\phi}(x, q)$; the second is unconnected with his consumption of the x 's and is represented by the appearance of q as the second argument of $T[.,.]$. In this case, the ordinary demand functions *do* depend on q ; they take the form:

$$x_n = h^n(p, q, y) = \bar{h}^n(p, q, y) \quad n = 1, \dots, N \quad (10)$$

where the functions $\bar{h}^n(p, q, y)$ are in fact the ordinary demand functions associated with the maximization of the sub-utility function $\bar{\phi}(x, q)$ alone. The crucial implication of this fact is that revealed preferences based purely on estimation of the demand functions for market goods $\bar{h}^n(p, q, y)$, $n=1, \dots, N$, will recover only the sub-utility function $\bar{\phi}(x, q)$, but *not* the function $T[.,.]$. The indirect utility function associated with (9) is:

$$u = v(p, q, y) = T[\bar{\psi}(p, q, y), q], \quad (11)$$

where $\bar{\psi}(p, q, y)$ is the indirect utility function that corresponds to maximization of the sub-utility function $\bar{\phi}(x, q)$ alone. The total value that the individual places on a change in q , denoted C^{TOT} , is given by:

$$T[\bar{\psi}(p, q^1, y - C^{\text{TOT}}), q^1] = T[\bar{\psi}(p, q^0, y), q^0]. \quad (12)$$

This has both a use value component, associated with $\bar{\psi}(p, q, y)$, and a passive use component, associated with $T[., q]$; the use component, C^{U} , satisfies:

$$\bar{\psi}(p, q^1, y - C^{\text{U}}) = \bar{\psi}(p, q^0, y), \quad (13)$$

while the passive component C^{PU} would be defined as the difference:

$$C^{\text{PU}} \equiv C^{\text{TOT}} - C^{\text{U}}. \quad (14)$$

From a CV perspective, the distinction between use and passive components of the individual's total value for a change in q could, in principle, be captured by formulating a model based on a specification of a utility function that conforms to the structure of (11), where the $\bar{\psi}(p, q, y)$ and $T[.,.]$ functions could be separately identified. The crucial feature of this structure is that prices and income interact in a manner that is partially separable from q . This has not usually been done, as commodity prices are typically not used as explicit variables in utility specifications and identification would rest on the functional form assumption. As such, *only* information elicited in CV surveys relates to C^{TOT} .

Some early CV researchers (e.g., Walsh, Loomis and Gillman, 1984) approached the problem of trying to separate C^{TOT} into components by first asking respondents for their total value and then requesting that they allocate this total among several specific motives. The difficult is that decomposition questions are difficult for respondents to answer if they think holistically about goods and that decompositions are sequence dependent and not generally unique (Mitchell and Carson, 1989, Cummings and Harrison, 1995, Carson, Flores, and

Mitchell 1999). Given the structure of (11) which generates the decomposition of the total value for the change in q into a use value component based on $\bar{\psi}(p, q, y)$, and a passive use value component based on $T[.,q]$; there are two more fruitful approaches to identifying the two separate components of value. The first is to accept a different type of decomposition where total value is defined in the usual way using a question where the respondent is able to use the resource and another question is asked where the resource will not be available for the respondent to use. The difference between WTP from these two questions is use value, under the assumption that the resource must first exist and then the ability to use it is added. This second question may not be plausible. The second is to collect two sets of information, one being total value from a CV survey and the other RP data on the demand functions for one or more of the x 's. Suppose, for simplicity, there is sufficient RP data to estimate a complete demand system for the x 's. This would be combined with SP data on total economic value, using an assumed specification of the indirect utility function given above to estimate a system consisting of:

$$\begin{cases} x_1 = h^1(p, q, y) = -[\partial \bar{\psi}(p, q, y) / \partial p_1] / [\partial \bar{\psi}(p, q, y) / \partial y] \\ \dots \\ x_N = h^N(p, q, y) = -[\partial \bar{\psi}(p, q, y) / \partial p_N] / [\partial \bar{\psi}(p, q, y) / \partial y] \\ T[\bar{\psi}(p, q^1, y - C^{TOT}), q^1] = T[\bar{\psi}(p, q^0, y), q^0] \end{cases} \quad (14)$$

The advantage of this approach is that the RP data enrich the SP data, the two sets of data are analyzed, conditional on the functional form assumption, in a mutually consistent manner, and they permit separate estimation of use and existence value as well as total value. Pioneering studies following this approach include Cameron (1992) and Adamowicz, Louviere, and Williams (1994) with Azevedo, Herriges and Kling (2003) providing a recent review.

5.3.3 Bequest value

Krutilla (1967) also introduced the concept of "bequest value" where some agents are willing to pay to protect a wilderness area or national park in order to preserve it for their children or grandchildren. Bequest value represents concern for a future generation where the current generation wants the future generation to have access to a particular resource. A utility function can be written with this motive entering in a similar fashion to q , impacting welfare but exogenous to current consumption. As such, we could think of q as a vector with

two separate components, $q=(q_1, q_2)$, where q_1 is the current period with respect to the existence of the wilderness area, which affects them either through being a current period park visitor (use value) and/or through the park's current period existence value. q_2 represents an increase in the agent's well-being due to future generations having the wilderness area available. Because of the bequest motive, the protection of the wilderness now involves a shift in *both* elements of q . With this, the formalism in q_1 and q_2 carries over as the definition of the individual's WTP and WTA to protect the wilderness area. As such, a bequest motive results in a potential re-specification of the agent's utility function, but it does not otherwise change the formal theory of the CV response probability function.

6. Theoretical validity of CV results

Almost from its inception, CV has turned up behavior thought to potentially be at odds with standard neoclassical economic theory. This is the main thrust of attack by some economic critics of CV (e.g., Hausman, 1993). With fifty years of experience conducting CV studies, these potential anomalies fall into three categories. The first are instances where our understanding of what neoclassical theory predicted turned out to be faulty. Much of the problem here stems from using intuition concerning demand and welfare theory for price changes where agents are free to adjust quantities to infer what sort of behavior should be observed for imposed quantity changes, the situation that characterizes many environmental goods. The second are the presence of various behavioral effects. Here it is useful to point out that some of the best known behavioral effects were first seen in CV surveys. These were later shown to be quite robust across a wide range of economic behavior. CV surveys cannot eliminate behavioral effects but they can provide respondents with choice questions that facilitate careful decision making. The third are anomalous behavior induced by poor survey design and administration. CV studies that have invested considerable time and effort into understanding what people believe, into presenting a credible choice scenario with a well-defined good and a coercive payment mechanism, and into a survey administration that enhances the survey's consequentiality tend to produce results that are well-behaved. To keep this section focused, we concentrate on the main anomalies that have been put forward in the literature.

6.1 Different elicitation formats yield different WTP estimates

Early on (e.g., Smith and Desvousges, 1986) CV researchers had clearly documented that different elicitation formats yielded different estimates of WTP.²⁴ This was troubling if one thought respondents truthfully revealed their preferences; and, these differences were often used by CV critics (McFadden, 1994) as an indication that people did not have well-formed preferences for the environmental goods. There are three difficulties though with this view. The first is that these differences appear to be systematic not random. Second, their existence was seen by psychologists as a function of framing, not as a survey artifact. Indeed, Tversky, Slovic and Kahneman (1990) argued that “matching” and “choice” frames being consistent with different behavior, as perhaps the major problem with neoclassical theory. The third was that Carson and Groves (2007) showed that if respondents are standard rational maximizing economic agents facing a consequential survey, then they should exploit all of the incentive and information characteristics of the particular elicitation format used. Their neoclassical model yield the strong prediction that different elicitation formats should generally produce different estimates of WTP. Further, they showed that their framework predicts the fairly complex pattern of typically observed differences between elicitation formats. In recent years, there has been a dramatic change in the literature, from difference in WTP estimates from different elicitation formats suggesting major problems with CV to an acknowledgement that the direction of the observed divergences are predicted by neoclassical theory.

6.2 Consistency of CV results with theoretical predictions

Results from CV surveys have raised many interesting questions about neoclassical economic theory which have prompted extensions or clarifications of that theory. The key to understanding these economic questions is recognizing there is a considerable difference in terms of how monetized utility differences are defined between a situation where an agent faces a price and determines the quantity to purchase to maximize utility which characterizes private goods and the situation where the agent faces a choice between having or not having access to a public good. Most textbook discussions of welfare economics are based on the first situation, as is the intuition of most economists. Hicks (1943), however, in his classic formulation of current welfare economics, showed that there are four consumer

²⁴ If different elicitation formats produced statistically similar WTP estimates, the choice of format to use would likely have stayed focused on issues of respondent burden and statistical efficiency.

surplus measures. Two of these were routinely ignored because they involved “rationed” goods, a World War II phenomenon of interest to Hicks.²⁵ Cornes and Sandler (1996) were to later show that pure public goods are just a special case of rationed (i.e., imposed quantity change) goods.

Much of the inherited wisdom about the price change case comes from Willig’s (1976) seminal paper that showed the difference between WTP and WTA was sufficiently small to be ignorable for small price changes. Willig, though, was careful not to include quantity changes in his analysis. The unraveling of structure of demand for imposed quantity changes that was set in motion by Hanemann’s (1991) showing that, for this case, divergences between WTP and WTA could be arbitrarily large.

6.2.1 WTP versus WTA

Total value from an economic perspective can always be cast in terms of the maximum amount that the agent would pay for the good or the minimum amount the agent would take in exchange for selling the good. Property rights determine whether WTP or WTA is the right welfare measure. The appropriate property is generally well-defined for private goods and it would not matter for environmental goods if WTP and WTA were close. If WTP and WTA are not close from a theoretical perspective, then it matters how property rights are assigned. This is often a complicated legal and political question for environmental goods.

Willig’s (1976) result provided considerable comfort to early researchers working on environmental valuation since it meant that a WTP measure could be substituted for WTA where needed with little loss of precision. Initial efforts, though, to measure WTA (e.g., Hammack and Brown, 1974; Bishop and Heberlein, 1979) suggested that WTA was considerably larger than WTP. This was initially taken as a sign of problems with CV or, at least, using CV to measure WTA. Subsequent developments suggested that the problem was with using Willig’s results to make inference about the imposed quantity change case.

Hanemann (1991) showed that the key difference between the price and quantity cases is that the difference between WTP and WTA is governed by one parameter, an income effect, while in the case of an imposed quantity change; the difference is driven by the ratio of the income effect to a substitution effect. As the substitutability of the public

²⁵ Hick’s four consumer surplus measures, compensating surpluses and variations and equivalence surpluses and variations, might be better seen today as WTP and WTA crossed with price and imposed quantity changes.

good with the available private goods becomes small which is likely to be the case for many environmental goods, the difference between WTP and WTA grows large. A competing explanation was prospect theory (Kahneman and Tversky, 1979) which predicted a sizeable divergence between WTP and WTA for behavioral reasons that clashed with neoclassical economic theory and there are richer variants of neoclassical economic theory that including dynamic effects, uncertainty, strategic behavior and bargaining that predict sizeable divergences (Kling, Phaneuf and Zhao, 2012) . Experiments using actual transactions in a variety of settings showed large divergences between WTP and WTA (e.g., Knetsch, Thaler and Kahneman, 1990). It soon became clear that the divergence between WTP and WTA first identified in CV studies was not a survey artifact, and indeed, routinely manifested itself in markets. Even traded financial assets did not seem to be immune, with junk bonds and thinly traded stocks showing much larger bid (WTP)-ask (WTA) spreads than predicted using Willig's result. Horowitz and McConnell (2002) in a widely cited review of the literature show that the ratio of WTA to WTP estimates is roughly the same for surveys and actual transactions. The key implications for CV are twofold—first should expect to see divergences between WTP and WTA. This is not a sign that CV is invalid and, second, this divergence can be important in some policy contexts (Knetsch, 1990).

6.2.2 Income elasticity of WTP

A claim often made by CV critics (Hausman, 1993) is that environmental goods should be luxury goods, and as such, one would expect the income elasticity from a CV study to be above one. As before, this claim also has problems. First, there is no good reason to believe that the environment is a luxury good. The typical result from CV studies is that most goods are necessary (e.g., normal but not luxury) goods. While income elasticities are typically significant, they do strike many researchers as smaller than might be expected, and the occasional, but not infrequent lack of significance, troubling.

Two very distinct reasons lie behind the empirical results. The first is that the income elasticity of WTP, the quantity estimated in CV studies is a very different economic quantity than the income elasticity of demand upon which the usual economic definition of luxury goods is based. Flores and Carson (1997) show the two elasticities are functionally related. In a world with two goods, the relationship being given by:

$$\begin{bmatrix} \eta_1^v \\ \eta_2^v \end{bmatrix} = - \begin{bmatrix} \sigma_{11}^d & \sigma_{12}^d \\ \sigma_{21}^d & \sigma_{22}^d \end{bmatrix}^{-1} \begin{bmatrix} \eta_1^d \\ \eta_2^d \end{bmatrix} \frac{y}{e^v} \quad (15)$$

where η_i^v are the (virtual) income elasticities of WTP, the σ are the cross-price elasticities of demand, η_i^d are the ordinary income elasticities, and y/e^v is income divided by e^v which is equal to income plus the value of *all* public goods consumed expressed in monetary terms. This result shows the income elasticity of demand and the income elasticity of WTP can differ significantly in magnitude and even sign. Because the term y/e^v is less than one, a researcher should usually expect to see the income elasticity of WTP being less than the corresponding income elasticity of demand.

The second reason for expecting the income elasticity of WTP to be smaller than what intuition might suggest is measurement error, as income is generally very difficult to measure. Indeed, the relevant income measure that should be used is open to question, ranging from wealth, to some definition of permanent income, to mental accounts, to some measure of discretionary income. These measurement problems will tend to attenuate the statistical estimate of the income elasticity of WTP toward zero.

6.2.3 Sequence effects and the adding-up test

An early CV finding was that the order in which two goods were valued matters, and could matter a lot. Consider a stylized example taken from some of the early work on air pollution valuation (Randall, Hoehn and Tolley, 1981). Respondents from Chicago have an average WTP of about \$100 for a specified air quality change in Chicago. When offered an additional specified air quality improvement in the Grand Canyon, they are willing to pay \$30 more. A different sample of respondents is given the reversed sequence. Their average WTP is \$60 for the Grand Canyon improvement and \$70 for the Chicago improvement. Policy makers are troubled to see such a result, believing there should be only one “true” monetary value.

By training, an economist does not believe that there is only one true monetary value. There should be substitution and income effects that come into play in different ways depending on the order of the sequence. The magnitude of the difference in this stylized example is larger than what many economists would expect. The question is what does neoclassical economic theory predict?

Carson, Flores and Hanemann (1998) show that if the goods being valued are normal goods and Hicksian substitutes for each other, which is the typical case, then the value of a particular public good should get progressively smaller the later it is valued in a WTP sequence. This finding says that a package of goods should be valued less than the sum of its independently valued constituents, comporting with the argument put forth in Hoehn and Randall (1989) that the benefit estimates used for policy purposes might be considerably too large if all goods were valued individually and the government provides more than one. Carson, Flores and Hanemann (1998) show the opposite effect occurs in a WTA sequence. With fewer substitute goods and more income as the sequence progresses the later in a sequence the good is valued, the larger its value. Further, they show that the WTP for a good valued first in a sequence is always less than or equal to WTA for the good valued in any order in the sequence.

Turning now to what determines the magnitude of the sequence effect, consider the set of compensated, cross-price elasticities of demand (σ_{ij}^d) from Deaton's (1974) well-known analysis of UK consumer demand for two goods: food and clothing. Own-price elasticities are -0.28 for food and -0.17 for clothing and cross-price elasticities are 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. All of these elasticities are reasonably small and they yield the following well behaved demand system:

$$\begin{bmatrix} \sigma_{11}^d & \sigma_{12}^d \\ \sigma_{21}^d & \sigma_{22}^d \end{bmatrix} = \begin{bmatrix} -0.28 & 0.08 \\ 0.21 & -0.17 \end{bmatrix}. \quad (16)$$

If, however, one restricts quantities in the choice that agents face, then the situation involves partial rationing and hence the inverse demand system with quantity restrictions. To do this we need to look at how the virtual (shadow) prices for the rationed goods (food and clothing) respond to changes in the rationed levels of both of these goods. The virtual price substitution elasticities (σ_{ij}^v), measuring responsiveness, are related inversely to the compensated price elasticities (Madden, 1991). For our food and clothing example, the virtual price matrix of the substitution terms is:

$$\begin{bmatrix} \sigma_{11}^v & \sigma_{12}^v \\ \sigma_{21}^v & \sigma_{22}^v \end{bmatrix} = \begin{bmatrix} \sigma_{11}^d & \sigma_{12}^d \\ \sigma_{21}^d & \sigma_{22}^d \end{bmatrix}^{-1} = \begin{bmatrix} -5.60 & -2.55 \\ -7.19 & -9.33 \end{bmatrix}. \quad (17)$$

The same demand system whose cross-price elasticities imply relatively small increases in demand of one good when the price of the other good increases (an 8% increase in food demand in response to a 100% price increase in clothing and a 21% increase in clothing demand in response to a 100% price increase in food), 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. Deaton's demand system involves only two common private goods so there is no need to resort to explanations involving inconsistent preferences or even peculiar characteristics of public goods to get large sequence effects.

While substitution effects are clearly sufficient to drive the sequence effects observed in many CV studies, income effects may also be important. While budget constraints are often thought relatively unimportant unless the cost is fairly sizeable, some households may not have much discretionary income, particularly if payment is required over a short time period.

CV critics have argued that if sequence effects are large then the value of goods can be manipulated by choosing the sequence order. There is some truth to this statement. While goods that people do not care about do not magically become valuable by virtue of placing them first in a sequence, it is possible to drive down the value of a particular public good by embedding it far out in a sequence of other desirable public goods. It is not surprising that supporters of a project want it considered first and rivals for available funds want it considered after their projects. This is the agenda control problem in politics and it would be peculiar if CV, or benefit-cost analysis, more generally, was able to circumvent this issue.

Another context-related consistency test, termed an adding-up test, was proposed in the Hausman (1993) volume and expanded upon in Diamond (1996). The test requires that the estimated monetary value of a bundle of two (or more) goods $[WTP(A, B)]$ be equal to a properly constructed sequence of WTP for the individual goods that takes account of payment and provision $[WTP(A) + WTP(B | A, -c)]$, where A and B are the two goods and c is the payment for A. At an abstract level, the test follows from satisfying duality properties commonly assumed in microeconomics. There are several problems with the test though. It is difficult to implement in a survey because it asks respondents to pretend they have received a specific good and paid a specific amount for it, when they have not. Closer examination of the test's theoretical underpinnings shows it relies on a strong function form

assumption that can best be seen by noting that it requires people are indifferent between a program that prevents some number of existing birds from being killed and a hatchery program producing the same number of new birds (Smith and Osborne, 1996). The last problem with the test is that people generally do not conform to the adding-up test predictions in real markets. Bateman, *et al.* (1997) run an experiment with students and two of their favorite commodities, pizza and coffee, and show they violate it. This should come as no surprise. Many marketing activities such as upselling exploit violations of the test. When the bundle is relevant, a good CV study should value it.

6.3 Scope tests

An issue involving the validity of CV that has received considerable attention after being put forward as a major concern by Kahneman and Knetsch (1992) and Hausman (1993), is whether CV estimates are sensitive to the scope of the good being valued.²⁶ The Arrow, *et al.* (1993) panel largely adapted the methodology used in the Carson, *et al.* (1992) Exxon Valdez study as its recommended practice except that it diverged on the scope issue by recommending that CV surveys being done for litigation purposes pass a scope test. The proximate cause for this recommendation was the striking results of Desvousges, *et al.* (1993) in which respondents gave roughly the same WTP for preventing 2000, 20,000, and 200,000 birds from being killed by oil, and the contention of the CV critics that CV researchers had not previously examined the issue of scope insensitivity.

Mitchell and Carson (1989) had earlier raised the possibility of scope insensitivity calling it “part-whole” bias and argued that the problem was caused by a failure of the survey designer to clearly communicate the characteristics of good(s) in the CV survey. This alternative helped to frame the debate. Was scope insensitivity a generic characteristic of CV surveys that happened because respondents were giving generalized answers not tied to the specific characteristics of the good being valued or was the phenomenon linked to specific poorly designed CV surveys that offered vaguely described goods or payment obligations?

To examine the issue of scope sensitivity it was necessary to sort the relationships between sequencing, nesting, and scope (Carson and Mitchell, 1995).²⁷ Nesting occurs when

²⁶ A good that is larger in scope than another good can have more of one or more desirable attributes. The attribute(s) in question can be quantitative or qualitative.

²⁷ Kahneman and Knetsch (1992) originally used the term “embedding” in their well-known paper to refer to two distinct phenomena. The first involved sequencing of goods where standard theory predicts that WTP for the same

one good is a proper subset of another, which can happen in one of two ways. One is where different goods can be valued in different combinations while the other is where different numbers of units of the same attribute are provided. An example of the first way is where a respondent values a particular wildlife refuge, a particular forest, and a particular beach. This composite good, in turn, nests the same forest and beach, which in turn is a composite good that nests the same beach. The second way has two goods differing along one quantitative attribute so that a larger good nests a smaller one. When one good nests the other, the theoretical prediction (if the nested good and its complement are both desirable) is that WTP should be the same or greater for the larger good. Thus, the scope sensitivity hypothesis can be tested by having respondents value two goods that differ in scope. This test can be external using two independent statistically equivalent subsamples each of which values one of the two goods or internal by having the same respondents value both goods.

The answer as to whether CV surveys suffered from generic insensitivity to scope was soon answered in the negative. Carson (1997) reviews the large body of empirical evidence on split sample external scope tests that quickly developed after the Arrow, *et al.* (1993) panel report. This evidence came from four sources. First, there were a number of existing tests of the scope insensitivity hypothesis. They had not been labeled as such because the hypothesis had prior to Kahneman and Knetsch (1992), not been of much interest and because these tests occurred in studies that had been done for policy purposes where the reason for using a split sample was to avoid giving respondents contradictory information about the goods being valued. These tests tended to almost uniformly reject the scope insensitivity hypothesis. The second source was new studies that contained explicit tests of scope insensitivity. These included two large CV studies (Carson, *et al.*, 1994; Carson, Wilks and Imber, 1994) which involved goods with primarily passive use considerations and in-person interviews and they too overwhelmingly rejected scope insensitivity. The third source were meta-analyses (e.g., Smith and Osborne, 1996) done across CV studies valuing the same class of good that showed estimates systematically varied with the characteristics of the good being valued. The fourth source was re-analyses of some of the key studies CV

good should differ depending upon the order in which it is valued. The second involves the relationship between the values of two goods, where one of the goods nests another. Carson and Mitchell (1995) recommend that the term "embedding" not be used because of ambiguity in its meaning and relationship to theoretical predictions.

critics put forward suggests claimed results are not robust and shows much the surveys used far from state-of-the-art CV surveys (Carson, 1997).

The conclusion of the Carson (1997) review has held up well over time. There have been a number of additional tests of scope insensitivity with most rejecting the hypothesis. When problems are found they tend to fall into a few areas. The first involves the use of small risk probabilities. Many people do not process small probabilities well. This issue is now well-known to not be isolated to surveys, but rather, is widely reflected in financial and insurance decisions. The second involves endangered species when the protection plan involves land. Here respondents often appear to correctly perceive that protecting the land is likely to protect everything living on it. The third involves the use of voluntary payments where theory analysis of the incentive structure suggests no scope effect may be seen.

There are other important aspects of conducting scope tests that are likely to be useful in thinking about evaluating CV results more generally. Subsequent research has shown information about how the overall choice task is conveyed (e.g., Bateman, *et al.*, 2004) or how a particular attribute like low-level risk (e.g., Corso, Hammitt and Graham, 2001) can influence sensitivity to scope. This suggests researchers need to be careful about how goods are presented when sensitivity to scope is likely to be an issue. With public goods, respondents are often skeptical that the government can deliver the good described and adjust their WTP accordingly. With respect to scope tests, there is evidence that suggests that respondents may be more skeptical about the government's ability to provide the larger version of the good than the smaller version. This can create the seemingly perverse result that respondents are willing to pay more for less.

It is also possible to conduct internal scope tests. The most common way to do this is with a DCE that offers respondents more than one alternative to the *status quo*.²⁸ When these differ by one or more attributes, it is possible to test whether changes in the attribute influence WTP in a statistically significant manner. Such tests are routinely passed though, and are thought to be weaker tests than the external version of a scope test because respondents typically can easily see the difference between goods.²⁹ However, it should be

²⁸ Other variants of internal scope tests are possible where there are covariates such as distance to a recreation site on which a difference in WTP is expected if respondents are paying careful attention to a good's details.

²⁹ It is possible to fail such a test with respect to a particular attribute, which is sometimes referred to as attribute non-attendance (e.g., Scarpa, *et al.*, 2009).

noted that side-by-side comparisons are a typical characteristic of markets.

6.4 Criterion and convergent validity³⁰

Economists suspicious of survey responses often ask the question: “How do CV results correspond with actual behavior?” This question, to some degree, is ill-posed in two ways. First, it assumes actual behavior represents a criterion validity comparison in the sense that the criterion contains no systematic error. This is in contrast to a convergent validity test which assumes that both measures may be influenced by various factors that may cause divergences with respect to the particular theoretical construct of interest. What has been shown consistently is that a criterion standard is rare if not non-existent. Almost all comparisons of CV estimates to some other type of measurement should be treated as tests of convergent validity (Mitchell and Carson, 1989). Second, it typically suggests an underlying belief that if there is a divergence then there is a problem with CV. However, a more sophisticated view is that CV estimates should be consistent with actual behavior only when theory and the conditions under which it is implemented suggest that there should be consistency. Divergence should be seen under conditions that predict divergence.

The Carson and Groves (2007) framework has been particularly useful in thinking about this situation. It is instructive to take up the three cases most often invoked by CV critics. The first involves comparing what people actually give in response to requests for voluntary contributions for an environmental good to an SP based estimate of what respondents say they will contribute.³¹ The survey estimates tend to be considerably higher than actual contributions, but this is to be expected. Economic theory predicts that people should free-ride with respect to letting others provide public goods and there is a large body of empirical evidence supporting this prediction, although free-riding behavior does not seem to be as complete as theory suggests. What was less obvious until Carson and Groves (2007) paper is that if respondents took a survey asking about voluntary contributions as consequential, then indicating a willingness to contribute would increase the likelihood that an actual voluntary fundraising drive would be mounted, which in turn can provide the good

³⁰ There are other forms of validity and reliability that can be examined. One of these is construct validity which asks if factors that might reasonably be expected to predict higher WTP do so (Mitchell and Carson, 1989). This is part of most CV studies and the results here strongly suggest that responses to CV questions are not random. Another is to ask about temporal reliability. This has been examined less often, but major comparisons such as the replication of the Exxon Valdez study (Carson, *et al.*, 1997) are encouraging.

³¹ For examples see Duffield and Patterson (1991), Seip and Strand (1992) and Champ, *et al.* (1997).

to a free-riding agent. The only prediction that arises is the survey should over-estimate WTP for the good and that the voluntary contributions should under-estimate WTP for it. Even this bound though is subject to caveats (Chilton and Hutchinson, 1999) because this often involves a shift in who will provide the good which can influence perceptions about the nature of the good including its likelihood of being provided.

The second case involves private goods where it is possible to observe actions in actual markets (including economic experiments) as well as asking survey respondents about them. Sometimes the “revealed” behavior is taken from actual markets and sometimes from economic experiments. It is often argued (e.g., Neil, et al. 1994) that private markets represent the best situation for SP estimates to perform well because the goods are more familiar to people, but familiarity is not a relevant issue. What is relevant is the survey’s incentive structure (Carson and Groves, 2007). If a respondent has a positive probability of wanting to buy a new good at the price stated, he should say “yes” because it increases the likelihood that the good will be produced and made available for sale. Not surprisingly, marketing surveys tend to over-estimate the percent of the public who actually buy. This is the expected result though, because the survey which may still be useful if calibrated correctly, traces out potential demand, not actual demand.³² In some environmental contexts, being able to measure potential demand for one-time actions, like hooking up to a new water system in developing countries has proven quite useful because the survey-based estimates tend to be reasonable predictions of how many households hooked up over a longer time horizon (World Bank, 1993).

The opposite situation to introducing a new good often happens when the survey asks respondents about existing private goods. Respondents appear to be less likely to buy at a given price in the survey than they are in an actual market context if they believe the survey is being used for pricing purposes.³³ These surveys can still be quite useful, as responses are often very informative about the relative attractiveness of different attribute levels since the incentives are to get the firm to supply the good’s preferred configuration at

³² There are obviously other factors at work in comparing survey results with market behavior including informational differences and marketing efforts.

³³ This is typically reflected in being more price sensitive in the survey than actual markets, the good’s ASC or both. Brownstone and Small (2005) provide a discussion in the context of variable toll road pricing. Louviere and Hensher (2001) suggest the broader attribute range in DCE, including price, may be a factor in decreasing marginal WTP estimates. Hensher (2010) discusses factors that may influence the nature of the divergence.

a low price. When both availability and price are at issue, the incentive structure can become quite complex and respondents with different beliefs are likely to rely on different heuristics.

The marketing and transportation literatures have often seen surveys as producing useful but biased estimates that need to be calibrated with revealed preference behavior. Without a coherent theory of how people respond to the surveys, however, over-estimates for new goods and a lower proclivity to purchase existing goods in pricing exercises have coexisted in the form of “forecast” problems from hypothetical bias using surveys (Hensher, 2010). Looking at the incentive structure provides the missing link. What should be clear from this brief discussion is that contrary to the initial belief that private goods should represent the best case for survey estimates, private goods represent difficult situations where one should expect to see systematic deviations between survey and market behavior.

The third place that researchers have looked at (and there are a number of these tests) is how well estimates from SP surveys perform relative to actual transactions in lab experiments. The typical experiment asks respondents about WTP for a good in a “real” treatment where payment is required and a “hypothetical” treatment where it is explicitly made clear that the response given will have no effect on either provision or payment. Murphy, *et al.* (2005) provide a meta-analysis of both treatments. They find a median ratio of estimated WTP in purely hypothetical treatments to estimated WTP in the actual payment treatments of 1.35.³⁴ While this upward bias is troubling, it is well within the range of uncertainty that characterizes much economic analysis. Perhaps more importantly, though, it is not clear what relevance these studies have with respect to judging how well CV studies are likely to perform. Most are done with students who are paid to participate. The survey instruments used often bear little resemblance to real CV studies and it is likely those receiving the “hypothetical” treatment are suspicious as to why they are being paid. The most important difference, however, is that a good CV survey is not purely hypothetical, it is consequential in the sense of potentially having a policy impact. Carson and Groves (2007) show that under a purely hypothetical treatment, a random response has the same impact on the agent’s utility as any other response. As such, it is not clear that anything useful from the perspective of a consequential CV survey can be learned from these tests.

³⁴ Murphy, *et al.* (2005) also find that there are a small number of very large outliers that can have a dramatic effect on the mean since one is dealing with a ratio. It is not surprising that these studies are prominently featured in claims that CV studies grossly over-estimate.

Let's now turn to the first case that can provide some information on the performance of CV surveys. With quasi-public goods such as outdoor recreation that require some type of complementary purchase or which are bundled into private goods, it is possible to compare WTP from both SP and RP (e.g., averting behavior, hedonic pricing, and travel cost analysis) approaches. Carson, *et al.* (1996) provide a meta-analysis of the ratio of CV to RP estimates based on 616 comparisons from 83 studies. They estimate the mean CV/RP ratio to be .89 with a confidence interval of [0.81-0.96]. This suggests that CVM estimates in the case of quasi-public goods are on average a bit lower than RP estimates and highly correlated (0.78).³⁵ Carson, *et al.* (1996) do identify a publication bias—studies are most likely to get published if the CV/RP ratio is close to one or quite large.

The literature has expanded dramatically so recent comparisons have focused on particular classes of goods. Rosenberger and Loomis (2000) perform a meta-analysis using 682 estimates from 131 CV and travel cost studies. They find the CV-based estimates are on average significantly lower than RP-based estimates controlling for the characteristics of the type of outdoor recreation being valued. Kochi, Hubbell and Kramer (2006) find that CV studies valuing statistical lives tend to produce smaller estimates than those obtained from hedonic wage studies.

Carson and Groves (2007) show that consequential SBC questions featuring a take-it-or-leave-it offer not tied to any other decision with a coercive payment should result in truth preference revelation so it is useful to look for comparisons to votes on binding referendum. Carson, Hanemann and Mitchell (1987) performed the first such test with a ballot proposition involving a water quality bond and a CV survey that mimicked the ballot proposition administered as part of the Field Institute's California Poll. Close correspondence was found between the percent in favor in the survey and the actual vote.³⁶ Four subsequent U.S. comparisons of estimates involving CV surveys have been done. Champ and Brown (1997) look at a referendum involving road maintenance in Colorado, Vossler, *et al.* (2003) look at an open-space bond issue in Oregon, Vossler and Kerkvliet (2003) look at a different

³⁵ Some studies valued multiple goods or used a different statistical approach to value the same good. Giving equal weight to each study rather than each comparison results in the estimate of average CV/RP ratio being 0.92 with the confidence interval [0.81-1.03].

³⁶ This measure was put on the ballot by the state legislature and did not have any organized group who spent substantial amounts supporting or opposing the measure. Surveys of support for a ballot measure are incentive compatible if respondents believe that there is any type of bandwagon effect such that other voters are more likely to favor a measure the higher the percent in favor when the survey results are released to the public.

Oregon open space bond issue, Johnston (2006) looks at a water supply bond issue in a small Rhode Island town. All find that the survey results and referenda votes are statistically indistinguishable and conservative if all don't know responses are treated as no votes which is common CV practice. This is not surprising since surveys on two candidate races and referenda taken close to elections have long been known to be quite good predictors of election outcomes.

7. Concluding remarks

There are now a vast number of CV studies and papers focused on various aspects of environmental valuation and fifty years of empirical experience implementing variants of the approach. A great deal has been learned over the course of this process and CV has substantively influenced how environmental economists view many issues. Nowhere is this clearer than with respect to the potential importance of passive use considerations and the development of a deeper understanding of the pure public goods nature of many environmental goods. At best, this chapter provides a broad overview of some of the main CV results and points the interested reader toward relevant literature for future exploration.

One of the main areas where CV has matured is in understanding the implications of the underlying neoclassical economic theory. Particularly for pure public goods and passive use values, these implications are often quite different from predictions concerning welfare measures based on price changes upon which most economists' intuition is based. Once these are taken into account, CV results by and large tend to be consistent with theoretical theory. Results from CV studies can be used to show agents violate some of the standard tenets of neoclassical economics, but none of these violations are specific to surveys. Agents routinely violate them in markets but it is possible to design CV surveys to help facilitate careful decision making by respondents.

CV moved toward the SBC elicitation format as a way of helping to ensure incentive compatibility when there was a pure public good for which only one level could be provided. On the other hand, CV often utilizes more complex DCE as a way of obtaining more information from each respondent than an SBC question provides and as a way of being able to value a much larger range of policy options at reasonable cost. There is some tension here. Carson and Groves (2007) show that while a SBC question can be incentive compatible if a set of auxiliary conditions are met, more complex DCE are not. Once one knows where to look for the violations, they were readily apparent. But these can often be dealt with without

substantially comprising results. In this sense, it is another factor to be cognizant of when analyzing choice data. The assumption that respondents truthfully answered all questions and treated choice sets independently should never have been a plausible assumption.

At first, more complex DCE seemed different in substantive ways from other variants of CV because of their focus on attributes and multiple choice sets. These differences were sometimes seen as overcoming problems critics had attacked CV for, but this was not to be the case. Any problem seen in a CV study using the SBC elicitation format generally has a counterpart when looking at more complex DCE. Protest responses and nay/yea-saying with an SBC look like *status quo* effects in a DCE with three or more alternatives.

More complex DCE, if anything, require even more skill on the part of researchers applying them. Conveying more levels of attributes in a way that is understandable to respondents is a difficult task, as is making the bundling of them together in a wide array of configurations plausible. Asking multiple choice sets also takes up scarce interview time in a survey relative to asking a single SBC. Experimental design issues loom considerably larger as the design used can influence what parameters are statistically identified and the efficiency of the choice sets used in collecting relevant information on preferences. Still, the gains from high quality implementation of more complex DCE can be very large. It allows exploration of a much larger set of policy options and can help identify the marginal value of changing the attributes of particular plans. This is close to the holy grail of environmental economics of being able to set marginal benefits equal to marginal costs.

References

- Acton, J.P. (1973), *Evaluating Public Progress to Save Lives: The Case of Heart Attacks*, Rand Research Report R-73-02, Santa Monica CA: RAND Corporation.
- Adamowicz, W.L., J.J. Louviere and M. Williams (1994), 'Combining revealed and stated preference methods for valuing environmental amenities', *Journal of Environmental Economics and Management*, **26**, 271-292.
- Adamowicz, W., P. Boxall, M. Williams and J.J. Louviere (1998) 'Stated preference approaches for measuring passive use values: choice experiments and contingent valuation', *American Journal of Agricultural Economics*, **80**, 64-75.
- Alberini, A., B.J. Kanninen and R.T. Carson (1997), 'Modeling response incentives in dichotomous choice contingent valuation data', *Land Economics*, **73**, 309-324.
- Andreoni, J. (1989), 'Giving with impure altruism: applications to charity and Ricardian equivalence', *Journal of Political Economy*, **97**, 1147-1458.
- Arrow, K., R. Solow, R., P.R. Portney, E.E. Learner, R. Radner and H. Schuman (1993), 'Report of the NOAA panel on contingent valuation', *Federal Register*, **58**, 4601-4614.
- Azevedo, C.D., J.A. Herriges and C.L. Kling (2003), 'Combining revealed and stated preferences: consistency tests and their interpretations', *American Journal of Agricultural Economics*, **85**, 525-537.
- Bateman, I.J., R.T. Carson, B. Day, W.M. Hanemann, N.D. Hanley, T.Hett, M.W. Jones-Lee, G. Loomes, S. Mourato, E. Özdemiroglu, D.W. Pearce, R. Sugden and J. Swanson (2002), *Economic Valuation with Stated Preference Techniques*, Northampton: Edward Elgar.
- Bateman, I.J., M. Cole, P. Cooper, S. Georgiou, D. Hadley and G. Poe (2004), 'On visible choice sets and scope sensitivity', *Journal of Environmental Economics and Management*, **47**, 71-93.
- Bateman, I.J., A. Munro, B. Rhodes, C. Starmer and R. Sugden (1997), 'Does part-whole bias exist?: an experimental investigation', *Economic Journal*, **107**, 322-332.
- Becker G.M., M.H. DeGroot and J. Marschak (1964), 'Measuring utility by a single-response sequential method', *Behavioral Science*, **9**, 226-232.
- Bennett, J. and R. Blamey (2001), *The Choice Modelling Approach to Environmental Evaluation*, Northampton, MA: Edward Elgar.
- Bishop, R.C. and T.A. Heberlein (1979), 'Measuring values of extra market goods', *American Journal of Agricultural Economics*, **61**, 926-930.
- Bowen, H.R. (1943), 'The interpretation of voting in the allocation of economic resources', *Quarterly Journal of Economics*, **58**, 27-48.
- Brown, G.M. and J. Hammack (1972), 'A preliminary investigation of the economics of migratory waterfowl', in J.V. Krutilla, (ed.), *Natural Environments: Studies in Theoretical and Applied Analysis*, Baltimore MD: Johns Hopkins University Press.
- Brownstone, D., D.S. Bunch and K. Train (2000), 'Joint mixed logit models of stated and revealed references for alternative-fuel vehicles', *Transportation Research B*, **34**, 315-338.
- Brownstone, D. and K. Small (2005), 'Valuing time and reliability: assessing the evidence from road pricing demonstrations', *Transportation Research A*, **39**, 279-293.
- Cameron, T.A. (1988), 'A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression', *Journal of Environmental Economics and Management*, **15**, 355-379.
- Cameron, T.A. (1992), 'Combining contingent valuation and travel cost data for the valuation of nonmarket goods', *Land Economics*, **68**, 302-317.

- Cameron, T.A., and D.D. Huppert (1989), 'OLS versus ML estimation of non-market resource values with payment card interval data', *Journal of Environmental Economics and Management*, **17**, 230-246.
- Cameron, T.A. and M.D. James (1987), 'Efficient estimation methods for 'closed-ended' contingent valuation surveys', *Review of Economics and Statistics*, **69**, 269-276.
- Carson, R.T. (1985), *Three Essays on Contingent Valuation*, Ph.D. Dissertation, UC Berkeley.
- Carson, R.T. (1997), 'Contingent valuation and tests of insensitivity to scope', in R.J. Kopp, W. Pommerehne and N. Schwartz, (eds.), *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, Boston: Kluwer.
- Carson, R.T. (2011), *Contingent Valuation: A Comprehensive Bibliography and History*, Northampton: Edward Elgar.
- Carson, R.T. (2012), 'Contingent Valuation: A Practical Alternative when Prices Aren't Available', *Journal of Economic Perspectives*, **26**(4), 27-42.
- Carson, R.T., N.E. Flores, and W.M. Hanemann (1998), 'Sequencing and valuing public goods', *Journal of Environmental Economics and Management*, **36**, 314-323.
- Carson, R.T., 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*, Oxford: Oxford University Press.
- 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 N.F. Meade (2001), 'Contingent valuation: controversies and evidence', *Environmental and Resource Economics*, **19**, 173-210.
- Carson, R.T. and T. Groves (2007), 'Incentive and informational properties of preference questions', *Environmental and Resource Economics*, **37**, 181-210.
- Carson, R.T. and T. Groves (2011), 'Incentive and Information Properties of Preference Questions: Commentary and Extensions', in J. Bennett (ed.), *International Handbook of Non-Market Environmental Valuation*, Northampton, MA: Edward Elgar.
- Carson, R.T., W.M. Hanemann (2005), 'Contingent valuation', in K. G. Mäler and J. R. Vincent, (eds.), *Handbook of Environmental Economics*, Elsevier: Amsterdam.
- Carson, R.T., W.M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud and V.K. Smith (1994), *Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight*, Report to the National Oceanic and Atmospheric Administration.
- Carson, Richard 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**, 151-163.
- Carson, R.T., W.M. Hanemann, R.J. Kopp, R.C. Mitchell, S. Presser and P.A. Ruud (2003), 'Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill', *Environmental and Resource Economics*, **25**, 257-286.
- Carson, R.T., W.M. Hanemann and R.C. Mitchell (1987), 'The use of simulated political markets to value public goods', Discussion Paper 87-7, Department of Economics, University of California, San Diego.

- Carson, R.T., W.M. Hanemann and D. Steinberg (1990), "A discrete choice contingent valuation estimate of the value of Kenai king salmon," *Journal of Behavioral Economics*, **19**, 53-68.
- Carson, R.T. and J. Louviere (2011), 'A common nomenclature for stated preference elicitation approaches', *Environmental and Resource Economics*, **49** (4), 539-559.
- 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., W.M. Hanemann, R.J. Kopp, R.C. Mitchell, S. Presser and P.A. Ruud (1992), 'A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill', Anchorage: State of Alaska.
- 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., R.C. Mitchell and P.A. Ruud (1990), 'Valuing air quality improvements: simulating a hedonic equation in the context of a contingent valuation scenario', in C. V. Mathai, (ed.), *Visibility and Fine Particles*, 639-946, Pittsburgh: Air & Waste Management Association.
- Champ, P.A., K.J. Boyle and T.C. Brown (eds.) (2003), *A Primer on Non-Market Valuation*, Boston: Kluwer Academic Press.
- Champ, P.A. and T.C. Brown (1997), 'A comparison of contingent and actual voting behavior', *W-133 Proceedings: Benefits and Cost Transfer in Natural Resource Planning*.
- 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.
- Chapman, R.G. and R. Staelin (1982), 'Exploiting rank ordered choice set data within the stochastic utility model', *Journal of Marketing Research*, **19**, 281-299.
- Chilton, S.M., and W.G. 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.
- Ciriacy-Wantrup, S.V. (1947), 'Capital returns from soil-conservation practices', *Journal of Farm Economics*, **29**, 1181-1196.
- Cornes, R. and T. Sandler (1996), *The Theory of Externalities, Public Goods and Club Goods*, New York: Cambridge University Press.
- Corso, P.S., J.K. Hammitt and J.D. Graham, (2001), 'Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation', *Journal of Risk and Uncertainty*, **23**, 165-184.
- Cummings, R.G., D.S. Brookshire and W.D. Schulze (1986), 'Valuing environmental goods: an assessment of the contingent valuation method', Totowa NJ: Rowman and Allanheld.
- Cummings, R.G., and G.W. Harrison (1995), 'The measurement and decomposition of nonuse values: a critical review', *Environmental and Resource Economics*, **5**, 225-247.
- Davis, R.K. (1963), 'The value of outdoor recreation: an economic study of the Maine woods', Dissertation, Harvard University.
- Deaton, A.S. (1974), 'The analysis of consumer demand in the united kingdom', *Econometrica*, **42**, 341-367.

- Desvousges, W.H., K.J. Boyle, R.W. Dunford, S.P. Hudson, F.R. Johnson and K.N. Wilson (1993), '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.
- 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 (1994), 'Contingent valuation: is some number better than no number?', *Journal of Economic Perspectives*, **8** (4), 45-64.
- Duffield, J.W. and D.A. Patterson (1991), 'Field testing existence values: An instream flow trust fund for Montana Rivers', Paper presented at American Economic Association Annual Conference.
- Farquharson, R. (1969), *Theory of Voting*, New Haven: Yale University Press.
- 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.
- Green, J.R. and J.J. Laffont (1978), 'A sampling approach to the free rider problem', in Agnar Sandmo (ed.), *Essays in Public Economics*, Lexington, MA: Lexington Books.
- Griffiths, C.W., H. Klemick, M. Massey, C. Moore, S. Newbold, D. Simpson, P. Walsh and W. Wheeler (2012), 'U.S. Environmental Protection Agency's valuation of surface water quality improvements', *Review of Environmental Economics and Policy*, **6**, 130-146.
- Haab, T.C., and K.E. McConnell (2002), *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*, Northampton: Edward Elgar.
- Hammack, J. and G.M. Brown, Jr. (1974), 'Waterfowl and wetlands: toward bioeconomic analysis', Baltimore MD: Johns Hopkins University Press.
- Hanemann, W.M. (1984a), 'Discrete/continuous models of consumer demand', *Econometrica*, **52**, 541-561.
- Hanemann, W.M. (1984b), 'Welfare evaluations in contingent valuation experiments with discrete responses', *American Journal of Agricultural Economics*, **66**, 332-341.
- 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. and B.J. Kanninen (1999), 'The statistical analysis of discrete-response cvm data,' in I.J. Bateman and K.G. Willis, (eds.), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries*, 302-441, Oxford UK: Oxford University Press.
- Hanemann, W.M., J.B. Loomis and B.J. Kanninen (1991), 'Statistical efficiency of double-bounded dichotomous choice contingent valuation,' *American Journal of Agricultural Economics*, **73**, 1255-1263.
- Hanley, N.D., R.E. Wright and W.L. Adamowicz (1998), 'Using choice experiments to value the environment - design issues, current experience and future prospects', *Environmental and Resource Economics*, **11**, 413-428.
- Hanley, N.D., S. Mourato and R.E. Wright (2001), 'Choice modelling approaches: a superior alternative for environmental valuation?', *Journal of Economic Surveys*, **15**, 435-462.
- Hausman, J. (ed.), (1993), *Contingent Valuation: A Critical Assessment*, Amsterdam: Elsevier.
- Heberlein, T.A., M.A. Wilson, R.C. Bishop and N.C. Schaeffer (2005), 'Rethinking the scope test as a criterion for validity in contingent valuation', *Journal of Environmental Economics and Management*, **50**, 1-22.

- Hensher, D.A. (2010), 'Hypothetical bias, choice experiments and willingness to pay', *Transportation Research, Part B*, **44**, 735-744.
- Hensher, D.A., J.J. Louviere and J. Swait (1999), 'Combining sources of preference data', *Journal of Econometrics*, **89**, 197-221.
- Hensher, D.A., J.M. Rose, and W.H. Greene (2005), *Applied Choice Analysis: A Primer*, New York: Cambridge University Press.
- Hensher, D.A., and P. Truong (1985), 'Valuation of travel time savings: a direct experimental approach', *Journal of Transport Economics and Policy*, **19**, 237-261.
- 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.K., and K.E. McConnell, 2002, "A review of wta/wtp studies", *Journal of Environmental Economics and Management*, **44**, 426-447.
- Hoyos, D. (2010), 'The state of the art of environmental valuation with discrete choice experiments,' *Ecological Economics*, **69**, 1595-1603.
- Johnston, R.J. (2006), "Is hypothetical bias universal?: validating contingent valuation responses using a binding public referendum", *Journal of Environmental Economics and Management*, **52**, 469-481.
- 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 A. Tversky, 1979, 'Prospect theory: an analysis of decision under risk', *Econometrica*, **47**, 263-291.
- Kanninen, B. (eds.), (2007), *Valuing environmental amenities using stated choice studies: a common sense approach to theory and practice*, Dordrecht: Springer.
- Kling, C.L, J.A. List and Jinhua Zhao (2010), 'A dynamic explanation of the willingness to pay and willingness to accept disparity', National Bureau of Economic Research, Working Paper No. 16483.
- Kling, C. L., D.J. Phaneuf and J. Zhao (2012), 'From Exxon to BP: has some number become better than no number?,' *Journal of Economic Perspectives* **26**(4):3-26.
- 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**, 227-238.
- Knetsch, J.L. and R.K. Davis (1966), 'Comparisons of methods for resource evaluation', in A.V. Kneese and S.C. Smith, (eds.), *Water Research*, Baltimore MD: Johns Hopkins University Press.
- 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.
- Kochi, I. B. Hubbell and R. Kramer (2006), 'An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis', *Environmental and Resource Economics*, **34**, 385-406.
- 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.
- Krutilla, J.V. (1967), 'Conservation reconsidered', *American Economic Review*, **57**, 777-786.

- Lancaster, K. (1966), 'A new approach to consumer theory,' *Journal of Political Economy*, **84**, 132-157.
- Louviere, J.J., (1974), "Predicting the evaluation of real stimulus objects from an abstract evaluation of their attributes: the case of trout streams', *Journal of Applied Psychology*, **59**, 572-577.
- Louviere, J.J., T. Flynn and R.T. Carson (2010), 'Discrete choice experiments are not conjoint analysis', *Journal of Choice Modelling*, **3** (3), 57-72.
- Louviere, J.J. and D.A. Hensher (2001), 'Combining sources of preference data', in: D.A. Hensher (ed.), *Travel Behaviour Research: The Leading Edge*, Oxford: Pergamon.
- Louviere, J.J., David A. Hensher and Joffre D. Swait (2000), *Stated Choice Methods: Analysis and Application*, New York: Cambridge University Press.
- Mackenzie, J. (1993), 'A comparison of contingent preference models', *American Journal of Agricultural Economics*, **75**, 593-603.
- Madden, P. (1991), 'A generalization of Hicksian substitutes and complements with application to demand rationing', *Econometrica*, **59**, 1497-1508.
- Marley, A.A.J., T.N. Flynn and J.J. Louviere (2008), 'Probabilistic models of set-dependent and attribute-level best-worst choice', *Journal of Mathematical Psychology*, **52**, 281-296.
- Manski, C. (1977), 'The Structure of Random Utility Models', *Theory & Decision*, **8**, 229-254.
- McConnell, K.E. (1990), 'Models for referendum data: the structure of discrete choice models for contingent valuation', *Journal of Environmental Economics and Management*, **18**, 19-34.
- McFadden, D.L. (1974), 'Conditional logit analysis of qualitative choice behavior', in P. Zarembka (ed.), *Frontiers in Econometrics*, New York: Academic Press.
- McFadden, D.L., (1994), 'Contingent Valuation and Social Choice', *American Journal of Agricultural Economics*, **76**, 689-708.
- Metcalfe, P.J. W. Baker, K. Andrews, G. Atkinson, I.J. Bateman, S. Butler, R.T. Carson, J. East, Y. Gueron, R. Sheldon and K. Train (2012), 'An assessment of the nonmarket benefits of the water framework directive for households in England and Wales', *Water Resources Research*, **48**.
- Mitchell, R.C. and R.T. Carson (1981), *An Experiment in Determining Willingness to Pay for National Water Quality Improvements*, Report to the U.S. Environmental Protection Agency, Washington DC: Resources for the Future.
- Mitchell, R.C. and R.T. Carson (1989), *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Baltimore: Resources for the Future.
- Murphy, J.J., G. Allen, T.S. Stevens and D. Weatherhead (2005), 'A meta-analysis of hypothetical bias in stated preference valuation', *Resource and Environmental Economics*, **30**, 313-325.
- Navrud, S. and R.C. Ready (eds). (2002), *Valuing Cultural Heritage: Applying Environmental Valuation Techniques to Historic Buildings, Monuments and Artifacts*, Northampton: Edward Elgar.
- Neill, H.R., R. Cummings, P. Ganderton, G.W. Harrison and T. McGuckin (1994), 'Hypothetical Surveys and Real Economic Commitments,' *Land Economics*, **70**, 145-154.
- Ortega, D.L., H. H. Wang, N.J. Olynk, L. Wu and J. Bai (2012), 'Chinese consumers' demand for food safety attributes: a push for government and industry regulations', *American Journal of Agricultural Economics*, **94**, 489-495.
- Rae, D.A., (1982), *Benefits of Visual Air Quality in Cincinnati*, Report to the Electric Power Research Institute, Boston: Charles River Associates.

- Randall, A., B.C. Ives and C. Eastman, 1974, "Bidding games for the valuation of aesthetic environmental improvements", *Journal of Environmental Economics and Management*, **1**, 132-149.
- Randall, A., J.P. Hoehn and G.S. Tolley (1981), 'The structure of contingent markets: some experimental results', Presented at American Economic Association Meeting.
- Rosenberger, R.S, and J.B. Loomis (2000), 'Using meta-analysis for benefit transfer: in-sample convergent validity tests of an outdoor recreation database', *Water Resources Research*, **36**, 1097-1107.
- Samuelson, P.A. (1954) 'The pure theory of public expenditures', *Review of Economics and Statistics*, **36**, 387-389.
- Samples, K.C., J.A. Dixon and M.M. Gower (1986), 'Information disclosure and endangered species valuation', *Land Economics*, **62**, 306-312.
- Scarpa, R., T.J. Gilbride, D. Campbell and D.A. Hensher (2009), "Modelling attribute non-attendance in choice experiments for rural landscape valuation," *European Review of Agricultural Economics*, **36**, 151-174
- Seip, K. and J. Strand (1992, "Willingness to pay for environmental goods in Norway: a cvm study with real payment', *Environmental and Resource Economics*, **2**, 91-106.
- Smith, V.K. and W.H. Desvousges (1986), *Measuring Water Quality Benefits*, Boston: Kluwer.
- 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.
- Thayer, M.A. (1981), 'Contingent valuation techniques for assessing environmental impacts: Further Evidence', *Journal of Environmental Economics and Management*, **8**, 27-44.
- Throsby, C. D., 1984, 'The measurement of willingness to pay for mixed goods', *Oxford Bulletin of Economics and Statistics*, **46**, 279-89.
- Thurstone, L. (1927), 'A law of comparative judgment', *Psychological Review*, **34**, 273-86.
- Train, K. and M. Weeks (2005), 'Discrete choice models in preference space and willingness-to-pay space', in R. Scarpa and A. Alberini (eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, Dordrecht: Springer
- Tversky, A., P. Slovic and D. Kahneman (1990), 'The causes of preference reversals," *American Economic Review*, **80**, 204-217.
- U.S. Environmental Protection Agency (1994), *President Clinton's Clean Water Initiative: Analysis of Benefits and Costs*, EPA 800-R-94-002, Washington: U.S. EPA.
- U.S. Water Resources Council (1979), 'Procedures for Evaluation of National Economic Development (NED) Benefits and Costs in Water Resources Planning (Level C), Final Rule', *Federal Register*, **44** (242), 72892-72976.
- Veronesi, M., A. Alberini and J.C. Cooper (2011), 'Implications of Bid Design and Willingness-To-Pay Distribution for Starting Point Bias in Double-Bounded Dichotomous Choice Contingent Valuation Surveys', *Environmental and Resource*, **49**, 199-215.
- Vossler, C.A., and J. Kerkvliet (2003), 'A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum', *Journal of Environmental Economics and Management*, **45**, 631-649.
- Vossler, C.A., J. Kerkvliet, S. Polasky and O. Gainutdinova (2003), 'Externally validating contingent valuation: an open-space survey and referendum in Corvallis, Oregon', *Journal of Economic Behavior and Organization*, **51**, 261-277.
- Walsh, R.G., J.B. Loomis and R.A. Gillman (1984), 'Valuing option, existence and bequest demands for wilderness', *Land Economics*, **60**, 14-29.

- Walsh, R.G., L.D. Sanders and J.B. Loomis (1985), *Wild and Scenic River Economics: Recreation Use and Preservation Values*, Report to the American Wilderness Alliance.
- World Bank, Water Demand Research Team (1993), 'The demand for water in rural areas: determinants and policy implications', *World Bank Research Observer*, **8**, 47-70.
- Willig, R. (1976), 'Consumer's surplus without apology', *American Economic Review*, **66**, 589-597.
- Zhao, J. and C. Kling (2009), Welfare measures when agents can learn', *Economic Journal*, **119**, 1560-1585.