The Discrete Choice Experiment Approach to Environmental Contingent Valuation

Richard T. Carson

Department of Economics
University of California, San Diego
and
Centre for the Study of Choice
University of Technology Sydney

Mikołaj Czajkowski

Department of Economic Sciences
University of Warsaw

For Handbook of Choice Modelling

Stephane Hess and Andrew Daly, editors

Draft

September 2012

1. Introduction[1]

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, the travel cost method and other variants of the household production function approach, 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,

[1] This work draws heavily upon our earlier work, and in particular: Carson, Flores and Meade (2001), Carson, et al. (2005), Carson, et al. (2007), Carson, et al. (2011), and Carson (2011).

[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 values are thought to be an important component of many cultural resources (Navrud and Ready, 2002) and a

existence value or stewardship value and often is defined to include bequest value and option/quasi-option value when uncertainty comes into play. In the case of many environmental goods, consumers may hold positive value for them even though they are not directly or indirectly using them. This passive use component is thought to constitute the majority of the total economic value of many 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 data base (EVRI) maintained in conjunction with environmental protection agencies in several other countries including Australia, France, New Zealand, the United Kingdom, and the United States.[4] Discussion of CV is now standard in almost all textbooks on environmental economics at the elementary (e.g., Tietenberg, 2006) and intermediate (e.g., Kolstad, 2011) and graduate level (e.g., Perman, *et al.*, 2003). 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 at the sectorial or regional level (Griffiths, *et al.*, 2012). Sometimes CV studies are done explicitly to look at one or more policy issues but often the results from older studies are 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, 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. This

---

public good like national defense is a classic case of a passive value. Krutilla's key insight was that passive use did not need to leave traces in choices that people made involving marketed goods.

[4] https://www.evri.ca/.

section concludes with a selective overview of the range of applications that are currently being undertaken. 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 received considerable attention 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 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 predictions from neoclassical economic theory 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. Next, a discussion of specific survey design and administration issues related to environmental valuation using CV is provided. For readers not familiar with environmental applications of DCE, a recent empirical example involving species protection in Scotland is provided. 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".

[5] For a number of reasons obtaining reliable estimates of WTA tends to be more difficult than for WTP. Throughout the paper, we will usually refer to obtaining WTP estimates unless there is a clear reason to distinguish conceptually between WTP and WTA.

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 and 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 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 to make 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 being developed there (Clawson and Knetsch, 1966), following an earlier suggestion of Hotelling, showed the two approaches produced similar estimates for an outdoor recreation example.

Over the next years several other economists followed 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

[6] CV also started to spread outside of environmental economics. See for instance the Acton's (1973) pioneering work on valuing health effects and the early work of Throsby (1984) on government support for the arts. There was also unrelated work in transportation that 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 many 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. A major difference with the environmental economics literature can clearly be seen by noting that the environmental economist's focus, at the time, would have been on determining the value of a fishing day.

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 federal participation in project evaluation which specified those methods that were acceptable for use in 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 in the characterization of CV.

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 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. It is the only non-DCE format that currently receives widespread use.[7]

[7] Even this format can be cast in a choice context. It is often modeled using an interval censored data framework with 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.

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. The discrete choice elicitation format for CV got a major boost when Hanemann (1984a; 1984b) put forward a comprehensive statistical framework for the estimation of the standard neoclassical Hicksian welfare measures based on the indirect utility function.[8] Cameron and James (1987) for the probit model and Cameron (1988) for the logit model put forward an alternative framework based on expenditure functions that exploits the fact that cost is randomly assigned.[9] McConnell (1990) showed the Hanemann and Cameron approaches to be duals of the same utility maximization problem when there is no random component but the two approaches can 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 the issues involved in estimating WTP from CV data.

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.[10] In response to a critique of CV put forward at an Exxon sponsored symposia contained in the volume edited by Hausman (1993), the U.S. government convened a panel co-chaired by Nobel Laureates Kenneth Arrow and Robert Solow to perform an assessment of the contingent valuation 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

[8] Hanemann's starting point was the random utility model of McFadden (1974).

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

[10] 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 prevent plan similar to that described in the study and key components of that plan have been deployed several times to prevent a spill.

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). By this point in time though researchers were starting to become aware of the incentive properties of CV survey questions (Hoehn and Randall, 1987; Mitchell and Carson, 1989) and offering multiple levels of a pure public good when only one level of a public good can be provided. This did not, however, initially discourage researchers who starting with Carson (1985) and Hanemann, Kanninen and Loomis (1991) often sought more information from each respondent by asking what is known as a double-bounded binary choice question which asked respondents who said yes initially about their whether WTP was greater than a higher amount. Those who said no initially were asked about their WTP a lower amount. The response to this question, if taken at face value, reduced the length of the interval in which the respondent's WTP lay, and hence, decreased 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 respondents about options involving Kenai King salmon fishing and fit a nested logit model to the data which included a *status quo* no-purchase option.[11]

[11] An earlier effort involving air pollution having respondents completely rank order a set of programs is Rae (1982) but was shown to have substantial econometric problems given the techniques of the time. There have been other periodic efforts to implement conjoint related techniques including ratings of alternatives (e.g., McKenzie, 1993) but these 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.

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 that more complex DCE formats could 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 specifically on the use of DCE for environmental valuation that was included 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 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 and DCE was defined to mean a sequence of multinomial choice questions. In contrast, Hanley, Wright and Adamowicz (1998) use a series of open-ended direct questions which they defined as CV to compare to a set of multinomial choice questions which was labeled as a choice experiment. A sequence of binary choice questions was sometimes called CV and 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 that: (a) DCE with more than two

---

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.

alternatives or multiple choice sets overcame some of the problems critics (e.g., Diamond and Hausman, 1994) asserted existed with CV, (b) that DCE were new which helped foster publication in academic journals and (c) because 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).[12] 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 is more complex than a single binary choice question and could never overcome the problems critics perceived with that format. That is because with enough statistically equivalent samples single binary discrete choice questions could be used to estimate anything that a more complex DCE could be used for 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.[13] One currently sees a variety of different DCE elicitation formats in use 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. Table 1 provides readers, and particularly those outside environmental economics, some idea of the range of goods currently being valued using DCE.

[12] 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 (Carson, 1991), for providing the good(s) including the specific description of the good(s) and their manner of their provision.

[13] Contemporary overviews are provided in 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).

Table 1: List of Representative Current DCE Environmental CV Applications

Good Valued	Location	Citation
Ecotourism by foreign visitors	Nepal	(Baral, Stern and Bhattarai, 2008)
Coastal nature reserves and flood protection	North Norfolk Coast (UK)	(Bateman, et al., 2009)
External benefits of hosting 2012 Olympics	United Kingdom	(Brouwer, et al., 2010)
Protection of birds of prey at managed moorlands	Scottish Uplands (UK)	(Hanley, et al., 2010)
Human and wildlife-related impacts of wind turbines	West Saxony (Germany)	(Drechsler, et al., 2011)
Water quality improvements under EU directive	England and Wales	Metcalfe, *et al.,* 2012)
Food (milk) safety attributes	Seven major Chinese cities	(Ortega, et al., 2012)
Visual aspects of forest management	Portugal	(Madureira, *et al.* (2009)
Benefits of climate change mitigation	Basque Country (Spain)	(Longo, et al., 2012)
Cleaning up irrigation water	Ethiopia	(Weldesilassie, *et al.* (2009)
Context-dependent value of statistical life	Czech Republic & Italy	(Alberini, et al., 2011)
Urban air quality improvement	China	(Wang and Zhang, 2009)

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.[14]

A matching question essentially solves the problem of what quantity makes the consumer indifferent between two situations, typically the *status quo* and an alternative.[15] The quantity that makes the agent indifferent effectively sets utility in the two states of the world equal. 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.

[14] 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.

[15] Correct specification of the *status quo* in both matching and choice formats is important in defining appropriate property rights situations, and by extension, the types of welfare measures obtained.

Obtaining a reliable answer to a matching question has proven difficult for two reasons. First, a direct question (DQ) for WTP (or WTA) tends to yield high non-response rates, most likely because agents in industrialized 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.[16] 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 that was 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 to the payment card (PCARD) as a way to get reasonably close to the desired matching measure. Both the BGAME and the PCARD move away from the DQ toward having the respondent answer one or more simpler choice questions.[17] The DQ, BGAME, and PCARD elicitation formats have all been used to value a single (SINGLE) good as well as 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 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

[16] 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 can effortlessly retrieve those numbers. A Herbert Simon view of the world (Conlisk, 1996) suggests bounded-rationality through satisficing in determining these numbers in particular contexts. This leads to a consideration of processing costs and limitations. Contemporary work in psychology and behavior economics sometimes follows this line of thinking about how agents deploy generally useful heuristics, but tends to put more emphasis on the possibility of cognitive biases.

[17] The direct question, bidding game, and payment card are the three matching formats commonly used in CV environmental applications. Carson and Louviere (2011) note that there are other matching formats, such as the time-tradeoff question used in health economics and the allocation game used in public economics.

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 the response to a SBC question doesn't reveal much preference information.

There are three ways to enhance the amount of preference information.[18] The first has already been noted and that is to 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 experiment 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 second 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*

[18] A key distinction to keep in mind is whether the extra preference information comes from a between respondent design or from a within-respondent design.

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 are also possible. For instance, a complete ranking (RANK) of a set of *j* alternatives can exploded to form a sequence of choice sets (Chapman and Staelin, 1982). Any division of alternatives into more preferred and less preferred subsets can also be used can to determine consumers' preferences by forming choice sets in which one alternative is chosen and other alternatives are not. A increasingly popular elicitation format that falls into this group is where respondents are asked to indicate their best alternative and their worst alternative among a set of $j > 2$ alternatives (Marley, Flynn and Louviere, 2008). This format is known as best-worse choice (BWCHOICE). It can be seen as a less burdensome version of ranking that extracts more information than a SMC.

Matching and DCE questions can sometimes appear in the survey instrument. We refer to them as hybrid methods. Perhaps the most frequent combination is a SBC question followed by a DQ question. This poses some strategic problems because the SBC question tells respondents that the good can be provided for its stated cost, however, the extra information given by the DQ question may be useful in reducing the respondent's WTP interval.

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 the issue was that respondents had problems valuing unfamiliar public goods. The divergence between these estimates is cited as one of the major reasons by some CV critics as to why estimates based on stated preference 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 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 framework for examining the incentive and informational properties of stated preference elicitation formats. The starting point of this framework is the assumption that agents are standard optimizing rational neoclassical agents. Preference elicitation formats are then examined using mechanism design theory.

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 responses to inconsequential questions do not influence 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 theorem of Gibbard (1973) and Satterthwaite (1975) 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 types of respondents to reveal preference information inconsistent with their 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.[19] 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 which 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, the probability the alternative is implemented is weakly monotonically increasing in the percent in favor,

[19] A well-known referendum example that violates this condition are votes on property taxes to support local public schools where the rules are that if the referendum fails to pass that it can only be brought back for another vote at a lower cost. It may be optimal for an agent, who prefers the higher school expenditure level to the *status quo* level of funding and cost, to vote no in the referendum if their ideal funding level lies between the *status quo* and the level of the referendum in order to increase the chance that a subsequent referendum closer to the ideal point will be offered.

can be substituted. This condition includes a 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 rather than a vote by the whole population. Putting these together yields that a SBC in an advisory survey 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 a 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.[20] The two WTP distributions revealed by a DBDC questions should not be perfectly correlated as the standard interval censored estimator assumes.

It is possible to generate more specific prediction about the properties of particular elicitation formats with stronger assumptions about respondent beliefs. For instance, in the DBBC, if the initial amount is perceived as the cost at which the government can supply the good, then one should generally see a downward bias in DBBC WTP estimates relative to the estimate from the first question. The difficulty though with moving into an assumption about beliefs is that different beliefs can yield the same basic result. Continuing with the DBBC elicitation format, another reasonable assumption is that respondents take the stated cost as the point estimate of what they would have to pay but now increase the level uncertainty around this payment obligation. Because one is valuing a public good with a coercive payment mechanism, increasing uncertainty over the payment obligation should translate into an increase in uncertainty over income. This in turn reduces the propensity to favor the program at a stated price relative to the case where there is no change in the level of

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

uncertainty.[21] Interestingly, this effect is indeterminate with quasi-public or private goods because the agent gets to observe the cost before deciding to purchase.

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 provided.[22] 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.[23] 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, 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

[21] In cases like the DBBC where most plausible belief structures lead to the same downward bias and empirical evidence supports this direction of the bias, the researcher may face a classic bias versus variance tradeoff, with the SBC being unbiased but often quite noisy in moderate sized samples and the DBBC being downwardly biased but with a much tighter confidence interval. Alberini, Kanninen and Carson (1997) discuss this situation and various modeling strategies.

[22] There are, of course, other reasons why IIA violations may occur.

[23] 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.

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.

The results for the DBBC give some idea of what can happen when the respondent sees the same good offered at different prices. In sequential DCE, the pattern of price relationships within and across choice sets in conjunction with the pattern of attribute levels seen can give rise to clear incentives for the respondent to not truthfully reveal his preferences. It is not uncommon to see respondents get more price sensitive over time or become more likely to choose the *status quo* level when the price offered does not seem like a good deal relative to alternatives seen earlier (e.g., Day, *et al.*, 2012). Other beliefs about the objectives of the sequential DCE can yield other patterns of response behavior that does not appear to be independent across choice sets. These may sometimes look like various types of learning, various behavioral effects, and changes in estimated scale across choice sets. In general, it is hard to uniquely sort out different reasons for observing some type of lack of independence of responses across choice sets. Because of the large gains in the amount of preference information collected in sequential DCE, what is needed is more work on how to exploit this information in a way that is robust to various types of response effects that might be observed.

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.[24] 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 (Varian, 1992) 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

[24] 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 or with or without being able to buy it on a credit card where the good could be paid for over the course of a year.

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. It envisions the standard goods of neoclassical economic theory of being bundles of attributes. Policy discussions are often interested in how value for the good changes as one or more of the attributes of the good 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.

Let us begin by denoting the item being valued (be it the change in the amount of 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.[25] 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.[26] 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

[25] The income variable that applies here can be defined in different ways. It could be the supernumerary income that is available to the individual after allowing for certain committed expenditures on market or non-market goods. Or the appropriate measure of income may be more closely tied to wealth than current period income. The inherent difficulty is that discretionary permanent income is not observed. Hence, any observable income measure is likely to contain considerable measurement error relative to the ideal measure. With classical measurement error, the estimated coefficient will be attenuated toward zero.

[26] That is, we assume $v(p,q,y)$ is homogeneous of degree zero in p and y, increasing in y, non-increasing in p, and quasi-convex in p.

$$v(p, q^1, y - C) = v(p, q^0, y), \tag{1}$$

and the equivalent measure, *E*, which satisfies

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

Observe that

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

If the change is regarded as an improvement, $C > 0$ and $E > 0$; in this case, *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 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,

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

The WTA function, $\mathrm{WTA}(q^0, q^1, p, y)$, 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. However, the field to a large degree has turned to the use of DCE 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 random utility model (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.[27] 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.[28] It is popular to fit data from the DBBC elicitation formats with models like a bivariate probit where interest often focuses on the correlation component. Empirical evidence 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 an incorrect statistical model (Versonsi, Alberini and Cooper, 2011).

Unobservables could be characteristics of the individual and/or attributes of the items considered for choice and/or they can signify variation in preferences among members of a population or measurement error. Pragmatically, introducing the stochastic component is equivalent of assuming that utility levels are random variables, as it is otherwise impossible to explain why apparently equal individuals (equal in all attributes which can be observed) may choose different options.

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

[27] This is not to say that the economic version of the RUM rules out respondent uncertainty. Indeed in most decision contexts including market purchases, respondents face some uncertainty about product attributes. The economic version of the RUM framework also does not rule out learning and other forms of dynamic behavior, but it does not have agents making optimization errors as is one of the typical explanations for error terms in most economic models. However, from the perspective of the analyst, there may appear to be optimization errors because the agent may choose a decision rule from some available set other than that assumed by the analyst.

[28] 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.

deterministic and the part which is random: $v(p,q,y;\varepsilon)=\vartheta(p,q,y)+\varepsilon$. Empirical implementation of random utility theory requires making an assumption about statistical properties of the random term. It is usually assumed 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 logit model (MNL), also referred to as a conditional logit model (McFadden, 1974), has the convenient closed-form expression of the probability of an individual i choosing alternative j from a set of all available alternatives J:

$$P(j \mid J)=\frac{\exp\left(\vartheta\left(p_{ij},q_{ij},y_{ij}\right)\right)}{\sum_{k=1}^{J}\exp\left(\vartheta\left(p_{ik},q_{ik},y_{ik}\right)\right)}. \tag{5}$$

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. As a result, economic value is clearly anthropocentric; it results from and reflects the preferences of individuals and should not be confused with an idea of fairness. Economists do not judge if consumers' preferences are right or wrong. They want to be able to observe them and express consumers' relative preferences for different goods in monetary units. This is how economic value is defined.

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. The literature has developed an extensive

set of potential motives and potential types of value (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 included 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. A passive use value then is comprised of existence value including its variants such as stewardship value, bequest value, option value and quasi-option value. Each of these can be explained in terms of a specific formulation of the individual's preference structure and a specific conceptualization of what is being valued by them. In addition, we briefly discuss altruism and the concept of 'warm glow' which is sometimes thought to be problematic from a policy perspective.

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. Hedonic pricing using household sales data is an example of a technique that relies on this type of data. Some environmental goods examined by environmental economists are comprised of mostly, if not exclusive, use value. Analysis of DCE in this instance more closely resembles their counterparts in other areas of applied microeconomics.

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 a – say a remote

national park even though they knew they would never visit the park 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." A number of earlier economists had clearly 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.[29] 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 exactly the same specification in utility. Stewardship is one of these and occurs when people believe that the government should be a "good steward" of some environmental 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.[30] 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.

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:

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

[29] The classic case is travel cost analysis of visits to an outdoor recreation site where the agent must pay travel costs to be able to use a site even if it has no entrance fee. Differences in travel cost mimic differences in an entrance fee and allow recovery of the demand curve for the site.

[30] Being a good steward can mean being willing to expend money to keep animals from being killed by human related causes but not natural causes. This conceptually causes no problem from the perspective of economic theory but it should be clear that utility is not defined on the number of animals but rather on the number of animals and the source of the threat to them.

where T[.,.] is a bivariate function, and $\bar{\phi}(\cdot)$ is a sub-function that aggregates the *x*'s. (6) involves a 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 entirely independent of the level of *q*. A consequence of the formulation in (6) is that the ordinary demand functions for the *x*'s are each independent of *q*; they take the form:

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

where the functions $\bar{h}^n(\cdot)$ are the ordinary demand functions associated with the 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 takes the form:

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

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:

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

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:

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

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 $\bar{h}^n(p,q,y),\ n=1,...,N$, will recover only the sub-utility function $\bar{\phi}(x,q)$, but *not* the function T[.,.]. The indirect utility function associated with (9) is:

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

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:

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

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

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

while the non-use component C^{NU} would be defined as the difference:

$$C^{\text{NU}} \equiv C^{\text{TOT}} - C^{\text{NU}}\ . \tag{14}$$

From a CV perspective, the distinction between use and existence 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 any identification would effectively rest on the functional form assumption. As such, *only* information elicited in CV surveys relates to C^{TOT}.

Some early CV researchers approached the problem of trying to separate C^{TOT} elicited through a DQ or other matching type question into components by first asking respondents for their total value and then requesting that they allocate this total among several specific motives.[31] Suppose for the moment that the only motives are use and existence value (other components will be considered below). It is unlikely that total value plus the elicited allocation between the two components could be used to uniquely recover the parameters of the separate components of the utility function (*i.e.*, $\bar{\phi}(x,q)$ and T[.,.]). Ignoring this theoretical difficulty, a number of empirical objections arise to any type of decomposition

[31] For an example see Walsh, Loomis, and Gillman (1984).

approach. From a survey perspective, employing a top-down approach of first asking for a total, without identifying the specific individual items to be included in this total, is an unwise practice. If people think holistically about an item, then on cognitive grounds, it may be difficult for them to produce a meaningful and reliable decomposition of this whole *ex post*. A bottom-up approach avoids some of these issues by introducing respondents to the separate components and asks about each of them, with the adding up occurring subsequently. Mitchell and Carson (1989), Cummings and Harrison (1995), Carson, Flores, and Mitchell (1999) point out that these decompositions are generally not unique since the portion of value allocated to a particular motivation may differ depending on the sequence in which they are elicited and the list of possible motivations is unlikely to be exhaustive from the perspective of some respondents.

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 an existence 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. There are obvious problems to asking the second question in many instances. The second is to collect two sets of information, one being total value from a CV survey and the other being revealed preference data on the demand functions for one or more of the x's. Suppose, for simplicity, there is sufficient revealed preference 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:

$$\begin{cases} x_1 = h^1(p,q,y) = -\left[\partial\bar{\psi}(p,q,y)/\partial p_1\right]/\left[\partial\bar{\psi}(p,q,y)/\partial y\right] \\ \ldots \\ x_N = h^N(p,q,y) = -\left[\partial\bar{\psi}(p,q,y)/\partial p_N\right]/\left[\partial\bar{\psi}(p,q,y)/\partial y\right] \\ T[\bar{\psi}(p,q^1,y-C^{TOT}),q^1] = T[\bar{\psi}(p,q^0,y),q^0] \end{cases}. \tag{15}$$

The advantage of this approach is that the revealed preference data enrich the SP data, the two sets of data are analyzed in a mutually consistent manner, and they permit the

estimation of the separate components of total value; at least in this particular case, where there are only use and existence value components in total value. Pioneering studies following this approach include Cameron (1992) and Adamowicz, Louviere, and Williams (1994).[32]

5.3.3 Bequest value

In addition to existence 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. More generally, Krutilla's bequest value can be seen as concern for a future generation where members of the current generation want 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 to them. q_2 represents 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.

5.3.4 Option value

Weisbrod (1964) put forth a different reason why agents might value a resource but not use it. His formulation relied on uncertainty and what became known as "option value". Under option value, some people who do not currently intend to visit a distant national park may be willing to pay money to protect it from destruction or irreversible damage, because they want to preserve the option of being able to visit it at some point in the future. After the

[32] For a recent review of the issues involved in combining the two sources of data see Azevedo, Herriges, and Kling (2003). A parallel literature exists in transportation and marketing (*e.g.,* Ben-Akiva and Morikawa, 1990). Swait and Louviere (1993) note that one should not expect the underlying variance parameters to be the same in revealed and stated preference data sources and that the possibility of different variances must be statistically allowed for when combining data from different sources (Hensher, Louviere and Swait, 1999).

publication of Weisbrod's paper, an extensive debate took place in the literature as to how to represent Weisbrod's option value in formal utility-theoretic terms.[33] Clearly, Weisbrod's option value involved choice under uncertainty, which is conceptually different than the utility models we have considered so far. The now standard formalization of Weisbrod's option value was provided by Graham (1981). Cameron and Englin (1997) have used Graham's model to look at responses to a CV question on option value. We follow their development here, using the context of protecting a national park from environmental damage as an example.

The core feature of Graham's model is uncertainty about the state of world and we will now adapt our basic utility model to include this.[34] For simplicity, assume that there are two possible future states of the world, one in which the individual will visit the national park and the other where the individual will not visit the national park. In this second state of the world the individual may or may not derive utility from existence of the park. If the first state of the world occurs, then the individual's indirect utility function is $v_a(q,y)$. If the second state of the world occurs, then the individual's utility is given by $v_n(q,y)$. The quality of the park may vary in either state of the world so we let q = 0 denote the case where the park is damaged, and q = 1 is the case whereby as the result of some policy intervention, the park is undamaged. Now let π denote the individual's subjective probability that in the future he will be a user of the park, while 1 - π is the probability of not being a user. Cameron and Englin (1997) allow for the possibility that this probability itself depends on the state of the park, so that $\pi = \pi(q)$; if the park is damaged, the probability of visiting it is lowered, $\pi(0) < \pi(1)$. There may also be uncertainty that the environmental damage will occur so we let P(1) be the individual's subjective probability that the park is undamaged (q = 1), and P(0) = 1-P(1) be the subjective probability that the park is damaged (q = 0). Graham (1981) puts forth a distinction between option price and expected surplus that has been followed by many in the literature.

Expected surplus can be derived as follows. If the individual knew with certainty that she would visit the park then her WTP to protect it from environmental damage is S_a, with the following condition being satisfied:

[33] There is a variant of option value, quasi-option value, where it is possible that new information is gained over time that can be important in an environmental context (Arrow and Fisher, 1974).
[34] For simplicity, we suppress the commodity price vector, p.

$$v_a(1,\ y - S_a) = v_a\left(0, y\right). \tag{16}$$

Conversely, if she knew for sure that she would *not* visit the park, her WTP would be S_n with the following condition being satisfied:

$$v_n(1,\ y - S_n) = v_n\left(0, y\right). \tag{17}$$

These are both *ex post* measures of WTP for q = 1 versus q = 0. *Ex ante*, the expected value of this expression, known as the *expected surplus*, is:

$$E\left(S\right) = \left[\pi\left(1\right)P\left(1\right) + \pi\left(0\right)P\left(0\right)\right] S_a + \left[\left(1 - \pi\left(1\right)\right)P\left(1\right) + \left(1 - \pi\left(0\right)\right)P\left(0\right)\right] S_n. \tag{18}$$

Making use of Hirschleifer's (1965, 1966) extension of the von Neumann-Morgenstern expected utility theorem, the individual's baseline expected utility when no policy intervention is undertaken is given by:

$$E\left(V\right)^* = \pi\left(0\right)\ P\left(0\right)\ v_a\left(0, y\right) + (1 - \pi\left(0\right))\ P\left(0\right)\ v_n\left(0, y\right). \tag{19}$$

Option price (OP) is the sure payment that, regardless of uncertain user status, the individual is willing to pay to ensure that the park does not suffer from environmental damage. *OP* satisfies the following equality:

$$E\left(V\right)^* = \pi\left(0\right)P\left(0\right)v_a\left(0, y - OP\right) + (1 - \pi\left(0\right))P\left(0\right)v_n\left(0, y - OP\right). \tag{20}$$

Empirical measurement of OP requires explicit models for (i) $\pi = \pi(q)$, the subjective conditional probability of participation given the environmental status of the national park, and (ii) P(q), the subjective probability of the status of the park; and (iii) $v_a(q,y)$ and $v_n(q,y)$, the state-dependent utilities defined above.

Cameron and Englin (1997) provide an empirical implementation of this framework to responses from surveys of the general population in four New England states in 1989 which had asked the discrete choice CV question "If acid rain damaged fishing in one fifth of all currently fishable high-altitude lakes in the Northeast, would you be willing to pay $X per year to prevent this?" Cameron and Englin (1997) are careful to point out a key difficulty—because the survey question is not explicit, respondents' answers can be given two possible interpretations. One is that respondents do not harbor any uncertainty. More specifically, if they are currently using the lakes, they will continue to do so in the future, and if they are not current users, then they will not become users in the future. Under this latter interpretation, the response to the survey question is based on the surplus S_a or S_n, that is they answered "yes" if $S_j > X$, $j = a$ or n, and "no" otherwise. The second possibility is that

respondents are unsure of whether they will use the lakes in the future and/or unsure about the future extent of harm to the lakes. In his case, the survey response should correspond to the expression for OP in (20) with the respondent answering "yes" if OP > X, and "no" otherwise.

5.3.5 Other motives

It is sometimes asserted that there are motivations for economic values that are problematic. These are often odd critiques, often by non-economists, where the primary thrust is that economic theory requires people to act as if they have very narrowly defined self-interest (Sagoff, 1988). While this assumption can make analysis simpler in some cases, nothing in neoclassical economic theory requires it and in much of economic analysis, such as the work of Gary Becker, it is not made. One motive which is sometimes invoked were problems may arise is altruism. Here the issue is not that altruism is in any way an illegitimate source of value, but rather that it raises the possibility of double counting. The reason for this is that altruism, as well as its counterpart envy, can involve interdependent utility function. In looking at the possibility of double counting, it becomes clear that the nature of the altruism matters. Paternalistic altruism in the form of my wanting you to have a park available doesn't cause a potential double-counting problem. Non-paternalistic altruism where I care about your utility and want to preserve the park only because I think it increases your utility, can also involve potential double-counting if I don't know that you will also have to pay if the park is preserved. But, this notion of everyone having to pay is typically made explicit in CV surveys. Further, altruism in many environmental valuation contexts is directed toward animals or ecosystems, in which case, the issue of interdependent utility functions does not arise.

Another problematic motivation sometimes raised is "warm glow". This term is often used in a myriad of different ways that amount to a contention that some people get utility by doing something or buying something that makes them feel good (Kahneman and Knetsch, 1992). There is, of course, no problem with this from the perspective of standard economic theory where motives don't matter. Unfortunately, some researchers define the notion of self-interest so narrowly that it rules out much behavior that makes common sense. Many people don't vote for political candidates they despise even if they think the candidate's election would increase their after tax income. Nor do many people buy the

lowest priced product if the company making it engages in practices they find strongly objectionable. Commercial products abound that are marketed to segments with particular tastes. The Toyota Prius is but one well-known example in the environmental marketing arena. There is a form of impure altruism first identified by Olson (1965) and later termed "warm glow" by Becker (1974) that has received extensive attention in the literature through the work of Andreoni (1989) who used the concept to show why there was not complete crowding out of private charities by government provision. Andreoni's work is of clear relevance for CV surveys that use voluntary contributions where people may feel good contributing to a particular group, irrespective of what they get. In the context of government provision with a coercive tax payment mechanism, however, the concept may be of limited use, as it requires people to get utility from paying higher taxes even though they know they will get nothing for it.

6. Theoretical validity of CV results

Almost from its inception, CV has turned up behavior which was 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. The interesting thing to point out here is 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. Thus, there is a clear pattern here. 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 generally appear to produce results that are well-behaved. Problems are

clearly concentrated among studies with poorly defined goods, no clear way of coercing payment, and haphazard survey administration, often involving students. While it was once popular to write a paper using data from one CV survey and from it to make very general statements about the validity of the method, hopefully those days are now long since over (Randall, 1998). No responsible researcher makes claims that a general approach to economic analysis does not work based on an isolated, often poorly executed study, but rather, considers the now considerable weight of the overall body of evidence.

To keep this section focused, we concentrate on the main anomalies that have been put forward in the literature. The first one considered is that different elicitation formats yield different WTP estimates. Next we examine a set of potential anomalies addressed that involve predictions about the relationship between different economic quantities. These include the relationship between WTP and WTA, the magnitude of the income elasticity of WTP estimated in CVM studies, the effect on the sequence in which a good is valued on WTP for it, and an adding-up test proposed Diamond (1996). Then we turn to the issue of sensitivity of CV estimates to the scope of the good being valued which was a major concern of the Arrow, *et al.* (1993) panel. Considerable theoretical clarification and empirical evidence on this issue now exists. Lastly we turn to the issue of how well CV estimates correspond to behavior in actual markets where it is possible to make comparisons.

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.[35] 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 consistently yielding different behaviors as perhaps the major problem with neoclassical theory. The third was that Carson and Groves (2007) showed that if respondents are

[35] 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.

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, Carson and Groves (2007) 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 and, to the extent that one thinks that observed effects are behaviorally motivated, that such behavior is not specific to CV surveys.

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 that there is a considerable difference 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 and the implicit price which makes utility in these two states equal. Most textbook discussions of welfare economics and quantities related to it are based on the first situation and, as such, the intuition of most economists is based on this case. There was always, though, a recognition by some theorists that the second case might behave quite differently. Hicks (1943), in his formulation of current welfare economics, showed that there are four consumer surplus measures. Two of these were routinely ignored because they involved "rationed" goods, a World War II phenomenon of interest to Hicks at the time.[36] However, Cornes and Sandler (1996) were to later show that pure public goods were simply a special case of rationed (i.e., imposed quantity change) goods. And in turn, many of the environmental goods that economists have been most interested in like the level of water quality in a city, the preservation of wilderness areas, or the enactment of climate policies as insurance against global warming, are all pure public goods.

[36] 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.

Much of the inherited wisdom about the price change case comes from Willig's seminal paper (1976) that argued that the difference between WTP and WTA was sufficiently small to be ignorable (for small price changes) where the qualification in parentheses was almost always dropped. Willig, though, was careful not to include quantity changes in his analysis and it took quite some time for the full ramifications of the difference between the price change and imposed quantity change cases to emerge. Starrett (1988), in an advance text of public economics, stops his analysis with the price change case noting that the imposed quantity case was much more difficult. The unraveling of structure of demand for imposed quantity changes that was set in motion by Hanemann's (1991) showing that contrary to the Willig's result, 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, as it meant that one only needed to measure WTP, as this 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. Based on Willig's work, 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 good with the available private goods became small, the difference between WTP and WTA

becomes large. As much of the popular discussion emphasized that perceived substitutability was likely to be very low for some environmental goods, like an endangered species, one would not expect WTP and WTA estimates to always be close together.

A competing explanation was prospect theory (Kahneman and Tversky, 1979) which predicted a divergence between WTP and WTA for behavioral reasons that clashed with neoclassical economic theory. Experiments using actual transactions in a variety of settings soon began to show 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.

Finding that there is a difference between WTP and WTA sets up a potential conflict between proponents of neoclassical and behavioral approaches. Many experimental tests are incapable of distinguishing between neoclassical Hanemann types and behavioral Kahneman and Tversky types. Sugden (1999) puts forward a deep critique of the WTP-WTA divergence arguing that Hanemann's framework is not sufficient to resolve the divergence, but he clearly recognizes that the issues involved are "not specific to CVM surveys". Other researchers took different tracks suggesting that while Hanemann's insights were important, there were other features of WTP versus WTA frames that needed to be taken into account to fully explain the divergence in a neoclassical framework. These include bargaining, strategic behavior, and differential uncertainty (*e.g.,* Carson and Groves, 2007; Kolstad and Guzman, 1999; Kling, List and Zhao, 2010; Zhao and Kling, 2004). The key implications for CV are twofold. First, one should expect to see a difference between WTP and WTA in contrast to claims by CV critics that this is a sign that CV is not valid (Hausman, 1993) and second, that this divergence may 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 to see the income elasticity from a CV study

above one, if it produced valid welfare measures. As before there are several problems with the claim. First, there is no good reason to believe that the environment is a luxury good. The very wealthy have always been able to isolate themselves from adverse environmental impacts and researchers working in developing countries are quick to attest to a myriad of issues that environmental contamination causes the lowest income groups. As such it is not surprising that 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.

It turns out that 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:

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

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, which may well be quite large. Examination of this result shows the income elasticity of demand and the income elasticity of WTP can differ significantly in magnitude and even sign. As a consequence, a good which is a luxury in a demand sense may have a WTP income elasticity which is less than zero, between zero and one, or greater than one. Closer examination suggests that the income elasticity of WTP is likely to be considerably smaller than the income elasticity of demand. To see this, when we let the matrix of cross-price elasticities be the identity matrix, the virtual price income elasticity is now equal to the ordinary income elasticity of demand, scaled by the ratio of income to income *plus* the monetized value of all public goods. Since this ratio is less than one, and likely substantially less than one, the income elasticity of WTP is likely to be 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. Income, in general, is very difficult to measure and subject to many interpretations. Indeed, the relevant income measure that should be used is open to question, ranging from wealth, to some definition of permanent income, to discretionary income after taking prior commitments and expenditures on necessities into account. Agents also have mental accounts (Thaler, 1985) that influence the income available for potential expenditures on environmental goods. From a statistical vantage point, all of these measurement related problems will tend to attenuate 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 in Chicago is given the reversed sequence. They are willing to pay on average about $60 for the Grand Canyon improvement and $70 for the Chicago improvement. Policy makers are, of course, troubled to see such a result, believing that there should be only one “true” monetary value.

By training (Debreu, 1959; Varian, 1992), an economist does not believe that there is 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 much larger than the intuition many would expect and, as a result, critics of CV have claimed that large sequence effects are inconsistent with neoclassical economic theory (Kahneman and Knetsch, 1992). As noted earlier, the question to ask is what does neoclassical economic theory actually predict?

Carson, Flores and Hanemann (1998) examine this question. They show that, if the goods being valued are normal goods and Hicksian substitutes for each other, which is the typical case, then the monetary value of a particular public good should get progressively smaller the later it is valued in a WTP sequence. An implication of this finding is that a package of goods should be valued less than the sum of its independently valued constituents, comporting with the argument first 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 of them.

Carson, Flores and Hanemann (1998) show the opposite effect occurs in a WTA sequence. The agent has fewer substitute goods as the sequence progresses and is compensated for this by more income. As a consequence, the later in a sequence the good is valued, the larger its monetary 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. As it is common to use a WTP measure as a surrogate for the desired WTA measure in natural resource damage assessments, this result suggests a downward bias is induced by this practice, in contrast to the policy case. They also show under very weak assumptions that standard Hicksian welfare measures cannot be context independent. That is order effects should always be observed.

The counter argument of CV critics is that the sequence effects observed in CV studies are too large. They base this on the contention that estimated income effects are small and they believe that goods such as those in the air quality example above are not close substitutes.[37] The standard intuition based on welfare measures for price changes is also faulty here. That is because one should be looking at an inverse demand system with quantity changes rather than at price changes in a demand system.

To see this, 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. Good one is food and good two is 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:

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

[37] See for example Kahneman and Knetsch (1992). McFadden (1994) put forward a similar argument drawing on his empirical analysis of a CV data set from a wilderness area study: “These results indicate that either there are extraordinarily strong diminishing returns to preserving additional wilderness areas, or that there is a context effect that makes responses inconsistent with classical economic preferences.”

The percentage change in demand will be small relative to either good in terms of the percentage change in either its own price or the other good's price. As such, one does not observe large context effects depending on the sequence in which changes are made.

If, however, one restricts quantities in the choice that agents face, then the situation involves partial rationing. 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}), which are the measures of responsiveness, are related inversely, as a system, to the compensated price elasticities (Madden, 1991; Flores, 1995). For our food and clothing example, the virtual price matrix of the substitution terms is:

$$\begin{bmatrix} \sigma_{11}^{v} & \sigma_{12}^{v} \\ \sigma_{21}^{v} & \sigma_{22}^{v} \end{bmatrix} = \begin{bmatrix} \sigma_{11}^{d} & \sigma_{12}^{d} \\ \sigma_{21}^{d} & \sigma_{22}^{d} \end{bmatrix}^{-1} = \begin{bmatrix} -5.60 & -2.55 \\ -7.19 & -9.33 \end{bmatrix}. \tag{23}$$

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 example 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.[38]

So far we have only looked at substitution effects and these clearly are sufficient to drive the sequence effects observed in many CV studies. However, 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. Income measurement issues (Sudman and Bradburn, 1982) may make it difficult to pick up this effect.

CV critics (e.g., Kahneman and Knetsch, 1992) have argued that if sequence effects are large then the value of goods can be manipulated by choosing the sequence order. There

[38] Simple conditional logit models often produce quite large sequence effects. Consider for example, six goods characterized by two attributes x_1 and x_2 which take on binary values (0, 1) and two price levels (p=1, 2), plus the no purchase alternative (0, 0, 0). Let the utility function for a specific alternative be $V = 3*x_1 + 6*x_2 - p$, where the parameters are known. WTP for the least desirable good (1, 0, 2) is over 2000 times higher if it is the first good provided than if it the last of six goods provided.

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 public goods that people want more of. It is not surprising that supporters of a project want it considered first and those with rivals for available funds want them considered after their projects. This is the agenda control problem in politics (Mueller, 1989) and it would be surprising if CV, or for that matter benefit-cost analysis, more generally, was somehow able to avoid it.

Another context-related consistency test, termed an adding-up test, was proposed in the Hausman (1993) volume with the test being 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 involves the implicit assumption that people are indifferent between a program that prevents some number of existing birds from being killed and a hatchery program which produces the same number of new birds (Smith and Osborne, 1996; Kopp and Smith, 1997). This assumption would be rejected by almost any environmentalist. The last problem with the test is that people generally do not conform to the predictions of the adding-up test 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 are designed around violations of the test.

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 (1992), is whether CV estimates are sensitive to the scope of the good being valued.[39] The Arrow, *et al.*

[39] 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.

(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).[40] Nesting occurs when 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, in the sense of using two independent statistically equivalent

[40] 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 neoclassical economic theory, which as noted earlier, predicts that WTP for the same 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 other. Carson and Mitchell (1995) recommend that the term "embedding" not be used because of ambiguity in its meaning and relationship to theoretical predictions.

subsamples each of which values one of the two goods. Or, the test can be internal, in the sense of having the same respondent 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 test 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 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 put forward by the CV critics. They suggest that the critics' claims were in most instances much weaker than they first appeared, and that the surveys on which they were based were 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. Theoretical analysis of the incentive structure for voluntary payments suggests that there is often no reason why responses should be sensitive to scope.

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 seemly perverse result that respondents are willing pay more for less.[41]

Another issue for scope tests is the statistical power to detect plausible differences as sizeable differences are sometimes declared not statistically different. Large sample sizes are particularly important when testing for differences in regions of the preference space where marginal utility may sharply decline (Rollins and Lyke, 1998). It can also be the case that people's preferences are not monotonic. Detailed examination (Heberlein, *et al.*, 2005), for example, found that while people preferred 300 wolves in the wild in Wisconsin, to none they got disutility from increasing the population of wolves to 800. This suggests that researchers need to be careful in trying to determine what people's choices are saying rather than jumping to the conclusion that something is amiss because more is not valued higher than less. A final issue to keep in mind in thinking about external scope tests is that there can be multiple metrics that can characterize the change in the good of interest. A good example of this is the Desvousges, et al. (1993) bird study which has received so much attention. Critics point to the similar WTP for preventing 2,000, 20,000, and 200,000 birds from being killed. However, respondents were also told that the three programs would save "much less than 1%" of the population, "less than 1%" of the population, and "about 2%" of the population. The average WTP responses of Desvousges, *et al.*'s sample resembles that of an ecologist who has positive WTP to prevent birds from being killed by oil but at a sharply

[41] Ironically, this problem is often caused by trying to keep the survey language for the two goods as close as possible, whereas it often takes more effort in terms of survey design to get the perceived probability of provision for the larger good equal to that of the smaller good.

declining rate typically found in most disaster relief plus WTP to prevent a very similar small adverse shock to the population.

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*.[42] 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,[43] though, they are thought to be weaker tests than the external version of a scope test because respondents typically can easily see the difference between goods. However, it should be noted that side-by-side comparisons are a typical characteristic of markets.

6.4 Criterion and convergent validity[44]

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 rather than being 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 underling 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 when theory and the conditions under which it is implemented suggest that there should be consistency and divergence under conditions that predict divergence.

[42] Other variants of internal scope tests are possible where there are covariates in a dataset, such as distance to a recreation site, on which a difference in WTP would be expected if respondents are paying careful attention to the details of the good.

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

[44] 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.

The Carson and Groves (2007) results have 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.[45] 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 fund raising drive would be mounted, which in turn can provide the good 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). Voluntary contribution mechanisms are sometimes used because a coercive payment mechanism does not seem plausible. But 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) and ask respondents about them in surveys.[46] Sometimes the "revealed" behavior is taken from actual markets and sometimes from economic experiments. It has often been argued 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" in the survey 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

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

[46] For examples, see Inforinso (1986), Neil, *et al.,* (1994), and Cummings, Harrison and Rutström (1995). In some economic experiments, agents are allowed to follow "home-grown" preferences/values rather than given induced values (Harrison and List, 2004).

the expected result, though, because the survey, which may still be useful if calibrated correctly, traces out potential demand, not actual demand.[47] 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 household hooked up over a longer time horizon.[48]

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.[49] These surveys can still be quite useful, as responses are often quite informative about the relative attractiveness of different attribute levels since the incentives are to get the firm to supply the preferred configuration of the good at 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 reveled 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.

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

[48] This is consistent with the results of a major World Bank review (Water Demand Research Team, 1993). CV estimates of water system hook-ups in developing countries where CVM estimates initially over-estimated the percent of households subscribing to water services, but they were reasonably accurate and not systematically biased forecasts of behavior over long time horizons.

[49] 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 that the much broader range of attributes in DCE, including price, may be a factor in decreasing marginal WTP estimates. Hensher (2010) discusses a variety of factors that may influence the nature of the divergence.

The third place that researchers have looked at is how well estimates from SP surveys perform relative to actual transactions in experimental economics labs. There are a sizeable number of such tests. 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.[50] While this upward bias is troubling, it is well within the range of uncertainty that characterizes much economic analysis. Further, regulatory cost estimates (Harrington, Morgenstern, Nelson, 2000) are known to have an upward bias. Perhaps more importantly, though, it is not clear what relevance these studies have 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 those used in 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.[51] As such, it is not clear that anything useful from the perspective of a consequential CV survey can be learned from these tests.

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

[50] 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.

[51] Poe and Vossler (2011) provide a useful summary of the growing literature on the role of consequentiality and its relationship to the set of laboratory experiments with purely hypothetical treatments.

correlated (0.78).[52] Carson, *et al.* (1996) do identify a publication bias consistent with divergent beliefs about what sort of results should be expected. Studies are most likely to get published if the CV/RP ratio is close to one or quite large.

The literature has expanded dramatically since then so more recent comparisons have focused on particular classes of goods and examine all of the studies done rather than just those with pairs of CV to RP comparisons. 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. There are also a large number of studies valuing statistical lives. Here Kochi, Hubbell and Kramer (2006) find that CV studies tend to produce smaller estimates than those obtained from hedonic wage studies.

Carson and Groves (2007) identify one survey situation where mechanism design theory suggests that truthful preference revelation is the dominant response in a CV survey. That is, when there is a consequential binary SBC featuring a take-it-or-leave-it offer not tied to any other decision with a coercive payment. A binding referendum with the same auxiliary conditions has the same properties making it natural to try to compare estimates from the two ways of establishing a market where agents can indicate whether they want the program at a given price.[53] Carson, Hanemann and Mitchell (1987) performed the first such test where the ballot proposition involved a water quality bond and the CV survey that mimicked the ballot proposition was administered as part of the Field Institute's California Poll. A close correspondence was found between the percent in favor in the survey and the percent in favor in the actual vote.[54] There have been four subsequent U.S. comparisons of estimates involving CV surveys done close to comparable referendum votes. Champ and Brown (1997) look at a referendum involving road maintenance in Colorado, Vossler, et al.

[52] 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 an average CV/RP ratio of 0.92 with the confidence interval [0.81-1.03]. One might expect the CV estimates to be somewhat higher due to the inclusion of passive use value, although this does not appear to be an important consideration in most cases. Concerns over provision issues tend to drive estimates downward.

[53] Bowen (1943) was the first to propose that surveys could be used to emulate voting in a referendum as a way of recovering the preference information needed to determine the demand for public goods.

[54] 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.

(2003) look at an open-space bond issue in Oregon, Vossler and Kerkvliet (2003) look at a different open space bond issue in Oregon, Johnston (2006) looks at a water supply bond issue in a small town in Rhode Island. All find that the survey results and referenda votes being statistically indistinguishable.[55] 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. Survey design, sampling and administration issues development

CV surveys differ from other surveys on public policy issues in several important ways; they also differ from marketing surveys. To a large degree, this comes from the typical need to present a considerable amount of information about the good the (environmental) policy will deliver and set up a scenario under which it would be paid for and provided. A CV survey can be thought of as having a structured conversation with a respondent during which a large amount of information is conveyed and the respondent becomes engaged in the task of providing preference information related to one or more plans to provide public goods. Surveys on policy issues typically cover many policies without going into much depth and, if they elicit any information about willingness to incur cost for a policy, do so on a cursory manner. The attributes of goods in many marketing surveys are reasonably well-defined in many cases and don't require detailed explanation. Marketing surveys with new products provide a closer comparison. The detailed scenarios presented in CV surveys have some parallels with information acceleration techniques used in some market studies involving new goods.

From a presentation standpoint, CV surveys generally contain an introductory section that identifies the sponsor and helps to put the general topic in the appropriate context, followed by the CV scenario. This scenario includes a description of what the policy is designed to accomplish, how it would be implemented and paid for, and what will happen under the current *status quo* situation if the project were not implemented. The policy is

[55] An issue in all of these studies as well as in Carson, Hanemann, and Mitchell (1987) is how to treat don't know responses. Standard practice in CV studies (Carson, *et al.*, 1998) is to treat all "don't knows" as not in favor and survey firms find that the don't know category tends to break disproportionately toward voting no, which always maintains the *status quo*. Treating "don't know" responses as "not in favor" responses results in the CV estimate being statistically indistinguishable or conservative relative to the referendum. The other major issue in comparing CV estimates to an actual referendum vote is the potential difference between the general voting age population and those that actually vote in a particular election.

described in terms of its attributes relative to the *status quo* option with considerable emphasis on its cost. In a DCE, respondents are then asked one or more choice questions that will provide preference information that can be used to help estimate WTP. The last part of a CV survey includes debriefing questions to help understand how respondents perceived various aspects of the scenario and demographic questions. Mitchell and Carson (1989) provide an overview of many of the issues involved in designing a of CV surveys, while Bateman, *et al.* (2002) provide a useful manual for the practitioner.[56]

The valuation scenario lies at the heart of any CV exercise.[57] The scenario must convey the change in the good to be valued, how that change would come about, how it would be paid for, and the larger context in which the change is being considered. This needs to be done in a way that is both consistent with the underlying scientific/engineering reality and comprehensible to respondents who may have little prior knowledge about the good in question. Respondents need enough information to be able to make an informed decision but should not be overwhelmed by information. They need to feel that the technical details of proposed programs have been worked out by the government[58] but that whether the program will be implemented is not a foregone conclusion, and the public's input via the survey will play an important role in that decision. The survey should make respondents feel comfortable with making either "favoring" or "opposing" a particular program. These are not easy tasks even though it seems easy to some outside observers simply to ask people about their WTP for a good. A CV survey instrument should have "face validity", in the sense of representing a factually accurate, easy to understand description of policy option(s) in a choice environment that facilitates good decision making by respondents.

Survey design problems are made more difficult because few economists are trained in survey research techniques. The description of the good(s) to be valued are often received in

[56] Mitchell (2002) provides a detailed description of the process of designing a CV survey instrument using the Exxon Valdez questionnaire as an example. Whittington (2002) provides a discussion of issues related to the design and administration of a CV survey from a developing country perspective. Louviere, Hensher and Swait (2000) is helpful for those designing DCE involving multiple goods. General references on issues related to survey design and administration are Bradburn, *et al.* (2004), Presser, *et al.* (2004), and Tourangeau, *et al.* (2000).

[57] A substantial amount of development work is usually required to produce a high quality CV survey instrument, particularly for environmental goods where there is little prior experience or where respondent beliefs about the program diverge substantial from those presented in the survey.

[58] Provision of only a brief description of a program runs the risk that respondents will believe they are being asked about whether it is desirable for the government to develop a concrete proposal on the topic. As such, the respondent may not take the stated details of the plan including its cost seriously.

very technical terms that need to be translated into terms that ordinary people can understand and use to make informed choices. What has emerged from decades of doing CV is that much of the value of developing a survey instrument has nothing to do with the actual estimation of the monetized value of policy options. The development of a CV instrument often results in policy makers being forced to determine what the actual policy options are and determining what those technical options would accomplish in terms that the public does cares about.[59] Involvement of policy makers and stakeholders at the design stage of a CV survey can foster interest in the results and make it less likely that economic analysis is an after the fact add-on that may stop a project from going forward but plays little role in helping to craft better policy options.

The designer of a CV survey often faces decisions that can influence WTP estimates. Here we can only give a brief description of some of the major ones that frequently come up. Perhaps the major issue that the designer of a CV survey faces is trying to convince respondents that the government can deliver the good(s) described. When a CV survey is seen as consequential by respondents, WTP for a good should be a function of the likelihood the good is delivered. In general, this biases CV estimates downward to the extent that some respondents believe that the good will be delivered if paid for with some uncertainty. The next issue that always arises is the payment vehicle. Just as goods have to be delivered, they need to be paid for. This requires choosing a specific payment vehicle. Respondents may have preferences over payment vehicles because they perceive them more as influencing the probability of provision or because they are seen as more appropriate or fairer given the characteristics of the good.

Payment for a good can be required in a single lump sum payment, in monthly payments over a year or monthly payments over a longer time horizon. Clearly, the most conservative choice is a single lump sum payment. This has evolved into the generally preferred alternative when the good under consideration would be provided in a one-time action like purchasing land for a wilderness area. Monthly payments are appropriate when

[59] If there is substantial scientific uncertainty or disagreement among stakeholders about policy outcomes, then it is possible to value different outcomes separately in different samples to see how sensitive the results are. Carson, Wilks and Imber (1994) provide an example involving proposed mining in the Kakadu Conservation Zone in Australia. When the policy debate is over what outputs a program should produce, then a DCE where the relevant attributes are explicitly varied may produce useful information for making the decision at hand.

the environmental service like local water quality is paid for through a monthly utility bill and the quality would quickly decline in the absence of the payments.

Provision of environmental services by the government raises a host of other issues. Many people resent taxes believing the government wastes money by diverting funds to activities other than providing the goods they were told would be supplied. This gives rise to various types of protest responses. These are often manifested by respondents indicating they are not in favor of the good at any positive price when, in reality, they would be in favor if they believed the good would be delivered at the stated price. This has often given rise to researchers dropping those indicating some type of protest response. The difficulty though is that it is optimal for someone with positive WTP that is less than their expected cost to register a protest in the hope that this will encourage the government to put the burden on other parties. The conservative decision is to treat the protests as "no" responses.

Plausibility including attribute levels, and particularly cost, is important. For instance, if the stated cost is too low to be plausible, respondents may substitute a higher cost estimate. In the marketing literature this is known as not being able to make a Mercedes into a Yugo. Development work in the form of focus groups and cognitive interviews, as well as debriefing questions on pilot surveys, can help uncover potential problems with respect to respondents' beliefs diverging from those intended by the researcher. It is not possible to overcome these divergent beliefs but it is usually possible to determine the overall direction of these effects on WTP estimates, particularly if choices on design features are made in a consistent direction (e.g., to conservatively influence WTP) as recommended by the Arrow, *et al.* (1993) Panel.[60]

As noted earlier, there has been considerable interest in moving beyond using a single SBC question. Much of this interest is driven by an explicit interest in valuing the attributes of projects (Holmes and Adamowicz, 2003). While a survey focused around a single SBC questions was recommended by Arrow, *et al.* (1993) and is ideal from many perspectives, undertaking a series of studies valuing related goods with a series of large statistical equivalent subsamples using the SBC format, is financially and logistically

[60] The researcher should not try to artificially drive down WTP estimates. Rather when faced with a design decision for which there is no clear choice dictated by the good or decision context, the researcher should consistently pick the option that tends to decrease rather than increase WTP estimates. This should be done as judiciously as possible.

infeasible for a study being done for policy purposes. The challenge in a more complex DCE is to use the attributes of a class of possible environmental policy options in an effective manner that helps to define specific goods that could be provided. There is clearly a lot of development work involved in doing this. As in other CV survey instruments, visual aids can help to convey complex information and help to hold respondent attention.[61] Successful approaches tend to get copied in subsequent studies. There is a history of study instruments that did not really work until some researcher determined a good way to explain or visually depict the policy to be valued or came up with elements of a plan to provide that good that convinced respondents the good could be provided and, if it were provided, that they would have to pay for it. The stock of expertise for particular classes of goods has been building up over time for more complex DCE, but it is still woefully inadequate in many areas. Many features of DCE such as the role of complexity (e.g., DeShazo and Fermo, 2002) cut across application areas. Substantial advances being made on several of these. Experimental designs for DCE and statistical models to fit data being generated from them are among the ones that are always relevant to environmental applications. Readers are referred to other chapters in this Handbook for overviews of the current state-of-the-art.

While marketing and transport surveys often limit their focus to current users of a good, CV surveys typically must face the issue of who are the potential demanders of the good.[62] This happens because improvements like increasing outdoor recreational opportunities may substantially increase the number of new users and because passive use values can exist in the general population. Defining the relevant population where passive use considerations are important can be particularly challenging because they do not need to follow well-defined geographic boundaries. Indeed, while some type of distance decay function is typical (Sutherland and Walsh, 1985), even this raises the question of where distance-wise one should stop interviewing from a cost-efficiency perspective. In some cases, and particularly those involving development versus preservation issues with a very well-known resource, people living far from the site may have higher values for preservation

[61] An issue that seems particularly relevant to environmental applications is that respondents often see causal links between different attributes and how those attributes can influence outcomes (Blamey, *et al.*, 2002). Such relationships need to be uncovered during development work and taken into account in constructing the DCE.

[62] This observation is, of course, not an absolute. Marketing and transportation surveys sometimes try to identify potential consumers of the good, although often stop short of interviewing a random sample of the general public. In the environmental arena, on-site surveys of outdoor recreators are common.

than those living close to the site (Carson, Wilks and Imber, 1994). Often agency guidance plays an important role in determining the population of interest.

The mode of survey administration is important in CV surveys.[63] Where a project has sufficient funding, in-person interviewing is generally desirable. In-person interviewing is ideal when there is a need to present a scenario in a specific order and there are considerable visual aids to help the respondent understand the good. Respondents tend to take in-person interviews more seriously and interviewers keep respondents focused. Further, it is possible to achieve a reasonably high response rate from a random sample of the general public.[64] The drawbacks of in-person surveys revolve around their substantial monetary and time costs, particular if a high response rate is needed. For this reason, mail surveys and, more recently, survey administered via the internet, are frequently used.[65] These survey modes can often provide useful information but issues of sample selection bias and how seriously respondents took the survey can arise.

Sample size is always an important issue in CV studies.[66] It is all too common to see CV studies with inadequate sample sizes to adequately examine the issues that were the stated purpose of the survey. CV surveys using a single SBC question, in particular, tend to require quite large sample sizes because they collect relatively little preference information. The costs of large samples often motivate using a matching format like the payment card or using a DCE comprised of a large number of choice sets to provide a large amount of

[63] Good sampling is important in any survey in order to represent the preferences of the population of interest. General survey sampling references include Cochran (1977), Kish (1965) and Levy and Lemeshow (1999). On the administration of surveys standard references, often with emphasis on particular modes, see Dillman, Smyth and Christian (2008), Fowler and Mangione (1989), Groves, *et al.*, (2004), and Presser, *et al.* (2004). After the survey is compete, various types of non-response issues usually need to be addressed.

[64] This, of course, is a relative statement as response rates from all modes of survey administration have been falling, which is leading to experimentation with offering some type of incentive to survey respondents and to consideration of specially recruited respondent panels.

[65] Telephone surveys have proven difficult to use with CV surveys except in cases where the good is already fairly well-known to respondents. They are sometimes used as the first stage of a hybrid administration where a sample of the population meeting a screening criterion (like engaging in fishing) is identified and then that sample is interviewed using a detailed CV survey delivered via some other mode like mail or the internet.

[66] Other aspects of sampling, such as stratification and clustering, can influence effective sample size. Further, underlying WTP distributions for environmental goods are typically characterized by reasonably large variances and indications of considerable heterogeneity. Both suggest the need for large samples. Sample size requirements are further increased when choice data is collected. Obtaining multiple responses from a single respondent can reduce sample size requirements but such responses will typically be correlated, so much less information is collected than often assumed. Bausell and Li (2002) provide a standard reference on statistical power in an experimental design context.

information on each respondent.[67] The tradeoff here is between the information and incentive issues raised by Carson and Groves (2007; 2011) and the more complex statistical analysis they may entail (e.g., Day, *et al.*, 2011) versus the collection of considerably more preference information per respondent. The need for large samples may also be a factor in the choice of survey administration mode given the large differences in the cost per completed interview. This is especially the case where multiple goods need to be valued. In many instances, it is possible to get a reasonable idea of the required sample size based on past work on related topics. The level of precision required for specific applications can vary considerably and this should be taken into account. Finally, the usual caveats on the role of non-sampling errors need to be taken into consideration in thinking about the entire process of survey design and administration (Groves, *et al.*, 2004).

8. An illustrative environmental DCE example

As some readers of this Handbook will not be familiar with what a DCE valuing an environmental policy looks like, we provide an illustrative example drawn from Hanley, *et al.* (2010) that valued different protection schemes of two species of birds of prey, the Golden Eagle and Hen Harrier. These birds are under increasing pressure from human activities. They breed on heather moorlands in the Scottish uplands in areas that are often managed for commercial shooting of Red Grouse and, as one might expect, are fond of eating grouse (Redpath, *et al.*, 2004). When the Red Grouse populations fall, the heather moorlands become less profitable and are at risk of being converted to forests and other types of landscape. There is ongoing conflict between moorland managers and conservation organizations. This study was done to investigate public preferences for alternative management strategies and to estimate their economic values to Scottish households so they can be used in cost-benefit assessments of potential policy options.

8.1 The choice context

To set up the context for the policy options, the survey needs to convey the location of uplands in Scotland, that these uplands are made up of a number of different habitats, including heather moorland, peat land and rough grassland, and their general uses. This is done using a combination of graphical and textual representations. The focus is on the half

[67] If characterizing heterogeneity in the population is important, then complex DCE will require large samples.

of the heather moorlands that are managed by burning and grazing by sheep, cattle and deer. This creates a unique ecosystem for many species, and, most importantly from an economic perspective, the Red Grouse. Red Grouse shooting contributes £5 million a year to the economy of the remote Scottish uplands and supports about 1,240 jobs. In areas where there are no longer enough grouse to maintain shooting, land owners often switch to forestry and this has resulted in a decline of over 30% in heather moorlands since 1950.

This decline in heather moorland has in turn adversely impacted the populations of two rare birds, Hen Harriers a medium-sized bird of prey of which there are roughly 633 pairs in Scotland, and Golden Eagles, a large bird of prey of which there are 440 pairs in Scotland. Golden Eagles are the top predator and share much of their habitat with Hen Harriers. Both birds are protected but are sometimes killed illegally, particularly in managed grouse moors because they prey upon the Red Grouse. The three birds, the Red Grouse, Hen Harriers, and Golden Eagles are displayed visually to respondents and details of their status and relationship to each other are provided.

8.2 The choice attributes and their levels

The two obvious policy attributes are possible population changes of Hen Harriers and Golden Eagles. The *status quo* alternative needs to be defined and, as is often the case in situations involving endangered species, populations are on a downward trajectory. Consultations with various experts and stakeholders suggested that *status quo* management would likely lead to a decline in populations of Golden Eagles and Hen Harriers on the order of 20% while the new management strategies that had been advanced could maintain populations at current levels or increase them up to 20%. To increase the plausibility that it was possible to change the population of one bird of prey independently of the other, responders were told that policies can be specifically targeted (e.g., moving eggs of particular species). As such, -20%, 0%, and +20% were used as the three possible policy outcomes for each of the two birds of prey. To help facilitate making translation of the policy options into monetary terms, the different policies were assigned a cost in terms of an additional annual tax households would have to pay if a particular alternative was implemented. Initial levels of the cost attribute were determined using focus groups and

updated after conducting a pilot study.[68] While environmental applications of DCE tend to use generic alternatives, people often care not only about the physical outcome (here the population change) but also how it is accomplished (Czajkowski et al., 2009), and that is accomplished here by using labeled alternatives which can be seen as similar to brands in marketing. A key debate among various stakeholders is whether the public has strong preferences over the "how" aspect of policy alternatives.

A number of alternative management options have been proposed to help manage the moorland stakeholder conflict (Redpath, et al., 2010). The three receiving the most attention by stakeholders and scientists at the time were used as alternative-specific labels in this study. They were: (1) stricter law enforcement, (2) establishment of feeding stations for grouse away from the birds of prey, and (3) physically moving eggs or chicks to new sites away from prime grouse moors. Consultation with experts and stakeholders helped to ensure the relevancy of the policy space explored. Focus groups were used to help refine the presentation of attributes, their levels, and the overall policy context to help ensure that respondents understood them in the way intended by the researchers. An example choice card from one of the six choice sets respondents faced is shown in Table 1.

Table 1. An example choice card

CHOICE A	DO NOTHING Maintain current management	LAW Stricter law enforcement	FEED Feeding stations away from grouse	MOVE Move eggs and chicks to new sites
HEN HARRIER	20% population decline	Maintain current population	Maintain current population	Maintain current population
GOLDEN EAGLE	20% population decline	20% population increase	20% population decline	20% population decline
COST	£0	£50	£25	£25
YOUR CHOICE (please tick one only)	☐	☐	☐	☐

[68] The choice experiment utilized a D-efficient design utilizing priors obtained through a pilot study. To account for uncertainty with respect to parameters' priors, they were modeled as random variables following normal distributions. The final design consisted of 8 questionnaire versions, each with 6 choice tasks per respondent.

8.3 Survey administration and structure

The survey was administered in 2009 by mail to a random sample of 1000 Scottish addresses.[69] The introduction explained why the survey was being carried out. It emphasized that the results would be used to inform future government policies regarding the management of the Scottish Uplands and assured confidentiality of respondent answers. The survey then provided respondents an overview of uplands management issues and explained the nature of the human-wildlife conflicts involving the Red Grouse and the two birds of prey included, detailing information about all three bird species. This was followed by a presentation of the new management strategies that could be implemented. Since the amount of information that needed to be conveyed was rather lengthy, the presentation was interrupted at several points with questions to help gauge the level of understanding of the presented issues and to obtain information about respondent attitudes.

The survey then focused on program costs explaining that any of the non-*status quo* management options would require additional funds in the form of higher taxes from Scottish households. Respondents were reminded to carefully consider their budgets and current expenditures in making choices. They were told that expert knowledge on the issues was not needed since the objective was to determine the public's preferences toward the government's policy options. Respondents then completed the six choice tasks and were asked to read back carefully through all of these to make sure they were happy with all of their choices. Finally, respondents were asked a series of socio-economic and behavioral questions including household income and about whether the respondent was a hunter.

8.4 A brief summary of results

Using a mixed logit model with correlated parameters, the study showed that the public was willing to pay between £35-65 for different policy options that would improve Hen Harrier and Golden Eagle populations in Scotland relative to the *status quo*.[70] As expected, respondents are willing to pay considerably more for improving the Golden Eagle population

[69] The sample was stratified by region and rural areas deliberately over-sampled. Two weeks after mailing the questionnaire, a reminder letter was sent to those not responding. Two weeks after this, a second reminder and copy of the questionnaire was sent out to those who still had not responded. The response rate was 22%.

[70] This study dropped three 'protesting' respondents who were identified by meeting three criteria: (1) stating that hen harriers or golden eagles should be protected, (2) choosing the *status quo* option in all six choice tasks and (3) giving as a reason for choosing this option stated that they did not think the plan would work, that the money would not be used appropriately or that other parties should pay for the plan.

than the Hen Harrier population and there is declining WTP for marginal increases of populations relative to the *status quo*. Hunters were willing to pay less than non-hunters and there is not a statistically significant difference between those living in rural versus urban areas. On average, respondents were largely indifferent to *how* conservation objectives are achieved; however, there is considerable heterogeneity in preferences. The results appear to be useful to discussions of policy options taking place among stakeholders. We refer the reader to Hanley, *et al.* (2010) for more detailed discussion of the study and its results.

9. 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 were 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 the theoretical 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 those predictions. There are, though, some strong policy implications. Substituting a WTP measure for a WTA measure can substantially undervalue a good if WTA is the correct property rights perspective and the sequence in which public goods are valued can matter a lot. It is possible to use CV studies to show that 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 and 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 to the public. On the other hand, it 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 was 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 a study 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.

CV can help decision makers to identify environmental policies that are in the public's interest. It is likely to be most useful in two situations. The first is where the benefits of providing a good are large but diffuse and where provision is opposed by a powerful special interest group. An improved level of air quality in a region is a good example. In this case, a countervailing interest group pushing for the good's provision is unlikely to spring up. The second case is where there is a strong lobby in favor of providing the good, with the public as

a whole paying for the project but where their aggregate WTP for it is much smaller than the project's cost. Buying a particular piece of land as a nature preserve that has high value to those adjacent to it might be a reasonable example here. The political process will often supply the good to the detriment of the public's welfare if there is no strong opposition. In both cases, an estimate of the public's WTP for the good can help illuminate the nature of the decision at hand. Both of these uses of CV will engender controversy because knowing the public's preferences can influence the outcome of a public decision in a manner adverse to a strong special interest group.

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.

Alberini, A. and M. Ščasný (2011), 'Context and the VSL: evidence from a stated preference study in Italy and the Czech Republic', *Environmental and Resource Economics*, **49**, 511-538.

Andreoni, J. (1989), 'Giving with impure altruism: applications to charity and Ricardian equivalence', *Journal of Political Economy*, **97**, 1147-1458.

Arrow, K. and A.C. Fisher (1974), 'Environmental preservation, uncertainty, and irreversibility', *Quarterly Journal of Economics*, **88**, 313-319.

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.

Atkinson, G., S. Mourato, S. Szymanski and E. Ozdemiroglu (2008), 'Are we willing to pay enough to back the bid?: valuing the intangible impacts of London's bid to host the 2012 Summer Olympic Games', **Urban Studies**, **45**, 419-444.

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.

Baral, N., M.J. Stern and R. Bhattarai (2008), 'Contingent valuation of ecotourism in Annapurna conservation area, Nepal: Implications for sustainable park finance and local development', *Ecological Economics*, **66**, 218-227.

Bausell, R. Barker and Yu-Fang Li (2002), *Power Analysis for Experimental Research: A Practical Guide for the Biological, Medical and Social Sciences*, New York: Cambridge University Press.

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 MA: 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.

Bateman, I.J., B.H. Day, A.P. Jones and S. Jude (2009), 'Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change', *Journal of Environmental Economics and Management*, **58**, 106-118.

Becker G.M., M.H. DeGroot and J. Marschak (1964), 'Measuring utility by a single-response sequential method', *Behavioral Science*, **9**, 226–232.

Ben-Akiva, M. and T. Morikawa (1990), 'Estimation of mode switching models from revealed preferences and stated intentions', *Transportation Research A*, **24A**, 485-495.

Becker, G. (1974), 'A theory of social interaction', *Journal of Political Economy*, **82**, 1063-1093.

Bennett, J. and R. Blamey (2001), *The Choice Modelling Approach to Environmental Evaluation*, Northampton, MA: Edward Elgar.

Blamey, R.K., J.W. Bennett, J.J. Louviere, M.D. Morrision and J.C. Rolfe (2002), 'Attribute causality in environmental choice modelling', *Environmental and Resource Economics*, **23**, 167-186.

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.

Bradburn, N.M., S. Sudman and B. Wansink (2004), *Asking Questions: The Definitive Guide to Questionnaire Design—For Market Research, Political Polls, and Social and Health Questionnaires*, San Francisco: Jossey-Bass.

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 J. Englin (1997), 'Welfare effects of changes in environmental quality under individual uncertainty about use', *RAND Journal of Economics*, **28**, 45-70.

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, University of California, Berkeley.

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

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 Hist*ory, Northampton, MA: Edward Elgar.

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 in the US, EC, and Developing Countries*, 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', State of Alaska, Anchorage.

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*, 639-46, Pittsburgh: Air & Waste Management Association.

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

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', *Proceedings from W-133 Benefits and Cost Transfer in Natural Resource Planning*, 10th Interim Report.

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.

Clawson, M. and J. Knetsch (1966), 'Economics of outdoor recreation', Baltimore: Johns Hopkins University Press.

Cochran, W.G. (1977), *Sampling Techniques, 3rd ed.*, New York: Wiley.

Conlisk, J. (1996), 'Why Bounded Rationality?', *Journal of Economic Literature*, **34**, 669-700.

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. Graha, (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.

Cummings, R.G., G.W. Harrison and E.E. Rutström (1995), 'Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible?', *American Economic Review*, **85**, 260-266.

Czajkowski, M. and N. Hanley (2009), 'Using labels to investigate scope effects in stated preference methods', *Environmental and Resource Economics*, **44**, 521–535.

Davis, R.K. (1963), 'The value of outdoor recreation: an economic study of the Maine woods', Dissertation, Harvard University.

Day, B., I.J. Bateman, R.T. Carson, D. Dupont, J.J. Louviere, S. Morimoto, R. Scarpa and P. Wang (2012), 'Ordering effects and choice set awareness in repeat-response stated preference studies', *Journal of Environmental Economics and Management*, **63**, 73-91.

Deaton, A.S. (1974), 'The analysis of consumer demand in the united kingdom', *Econometrica*, **42**, 341-367.

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

DeShazo, J.R. and G. Fermo (2002), 'Designing choice sets for stated preference methods: the effects of complexity on choice consistency', *Journal of Environmental Economics and Management*, **43**, 360-385.

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.

Dillman, D.A., J.D. Smyth and L.M. Christian (2008), *Mail and Internet Surveys, 3rd ed*., Hoboken, NJ: Wiley.

Drechsler, M., M. Eichhorn, J. Meyerhoff, J.Monsees and C. Ohl (2011), 'Combining spatial modeling and choice experiments for the optimal spatial allocation of wind turbines', *Energy Policy*, **39**, 3845–3854

Duffield, J.W. and D.A. Patterson (1991), 'Field testing existence values: An instream flow trust fund for Montana Rivers', Paper presented at the American Economic Association Annual Conference.

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

Flores, N.E. (1995), 'The effects of rationing and virtual price elasticities', Discussion Paper 95-20, Department of Economics, University of California, San Diego

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.

Fowler, F.J. and T.W. Mangione (1989), *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*, Beverly Hills, CA: Sage.

Gibbard, A. (1973), 'Manipulation of voting schemes: a general result', *Econometrica*, **41**, 587-601.

Green, J.R. and J.J. Laffont (1978), 'A sampling approach to the free rider problem', in Agnar Sandmo (ed.), *Essays in Public Economics*, Lexington, MA: Lexington Books.

Graham, D.A. (1981), 'Cost benefit analysis under uncertainty', American Economic Review, **71**, 715-725.

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.

Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer and R. Tourangeau (2004), *Survey Methodology*, New York: Wiley.

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. (1994), 'Valuing the environment through contingent valuation', *Journal of Economic Perspectives*, **8** (4), 19-43.

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., M Czajkowski, R. Hanley-Nickolls and S. Redpath (2010), 'Economic values of species management options in human–wildlife conflicts: Hen Harriers in Scotland', *Ecological Economics*, 70, 107-113.

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.

Harrington, W., R.D. Morgenstern and P. Nelson (2000), 'On the accuracy of regulatory cost estimates', Journal of Policy Analysis and Management, **19**, 297-322.

Harrison, G.W and J.A. List (2004), 'Field experiments', *Journal of Economic Literature*, **42**, 1009-1055.

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

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.

Hirschleifer, J. (1965), 'Investment decision under uncertainty: choice theoretic approaches', *Quarterly Journal of Economics*, **79**, 509-536.

Hirschleifer, J. (1966), 'Investment decision under uncertainty: applications of the state-preference approach', *Quarterly Journal of Economics*, **80**, 252-277.

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.

Holmes, T.P. and W.L. Adamowicz (2003), 'Attribute Base Method', in P.A. Champ, K.J. Boyle, and T.C. Brown (eds.), A Primer on Nonmarket Valuation, Norwell, MA: Kluwer.

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.

Infosino, W.J. (1986), 'Forecasting new product sales for likelihood of purchase ratings', *Marketing Science*, **5**, 372-390.

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.

Kish, L.F. (1965), *Survey Sampling*, New York: Wiley.

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.

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. (2011), *Environmental Economics, 2nd ed.*, New York: Oxford University Press.

Kolstad, C.D. and R.M. Guzman (1999), 'Information and the divergence between willingness to accept and willingness to pay', *Journal of Environmental Economics and Management*, **38**, 66-80.

Kopp, R.J. and V.K. Smith (1997), 'Constructing measures of economic value', in R.J. Kopp, W.W. Pommerehne and N. Schwarz (eds.), *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, Boston: Kluwer.

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.

Levy, Paul and Stanly Lemeshow, (1999), *Sampling of Populations: Methods and Applications*, New York: Wiley.

Longo, A., D. Hoyos and A. Markandya (2012), "Willingness to pay for ancillary benefits of climate change mitigation', *Environmental and Resource Economics*, **51**, 119-140.

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.

Madureira, L., L.C. Nunes, J.G. Borges, and A.O. Falcao (2009), 'Assessing forest management strategies using a contingent valuation approach and advanced visualisation techniques: A Portuguese case study', *Journal of Forest Economics*, **17**, 399-414.

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 and 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. (2002), 'On designing constructed markets in valuation surveys', *Environmental and Resource Economics*, **22**, 297-321.

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.

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

Mur 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, MA: Edward Elgar.

Neill, H.R., R.G. Cummings, P.T. Ganderton, G.W. Harrison and T. McGuckin (1994), 'Hypothetical Surveys and Real Economic Commitments,' *Land Economics*, **70**, 145-154.

Olson, M. (1965), *The Logic of Collective Action: Public Goods and the Theory of Groups*, Cambridge, MA: Harvard University Press.

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.

Perman, R., Y. Ma, J. McGilvray and M. Common (2003), Natural Resource and Environmental Economics, 3rd ed., Boston, MA: Addison-Wesley.

Poe, G. and C.A. Vossler (2011), 'Consequentiality and Contingent Valuation: An Emerging Paradigm', in J. Bennett (ed.), *International Handbook on Non-Market Valuation*, Cheltenham, UK: Edward Elgar.

Presser, S., J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin and E. Singer (2004), *Methods for Testing and Evaluating Survey Questionnaires*, New York: Wiley.

Rae, D.A., (1982), *Benefits of Visual Air Quality in Cincinnati*, Report to the Electric Power Research Institute, Boston: Charles River Associates.

Randall, A. (1998), 'Beyond the crucial experiment: mapping the performance characteristics of contingent valuation', *Resource and Energy Economics*, **20**, 197-206.

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', Paper presented at American Economic Association Meeting, Washington, DC.

Redpath, S., B. Arroyo, F. Leckie, P. Bacon, N. Bayfield, R. Gutierrez and S. Thirgood (2004), 'Using decision modelling to resolve human–wildlife conflicts: a case study with raptors and grouse', *Conservation Biology*, **18**, 350–359.

Redpath, S.M., A. Amar, A. Smith, D.B.A. Thompson and S.J. Thirgood (2010), 'People and nature in conflict: can we reconcile hen harrier conservation and game management?', In D.B.A. Thompson (ed.), *Species Management: Challenges and Solutions for the 21st Century*, London: UK Stationery Office.

Rollins, K. and A. Lyke (1998), 'The case for diminishing marginal existence values', *Journal of Environmental Economics and Management*, **36**, 324-344.

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.

Sagoff, M. (1988), *The Economy of the Earth*, New York: Cambridge Economic Press.

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.

Satterthwaite, M. (1975), 'Strategy-proofness and arrow conditions: existence and correspondence theorems for voting procedures and welfare functions', *Journal of Economic Theory*, 10, 187-217.

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, VK. 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.

Starrett, D.A. (1988), *Foundations of Public Economics*, New York: Cambridge University Press.

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

Sugden, R. (1999), 'Alternatives to the neo-classical theory of choice', in I.J. Bateman and K.G. Willis, (eds.), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the USA, EC, and Developing Countries*, New York: Oxford University Press.

Sutherland, R.J. and R.G. Walsh (1985), 'Effect of distance on the preservation value of water quality', *Land Economics*, **61**, 281-291.

Swait, J., and J.J. Louviere (1993), 'The role of the scale parameter in the estimation and use of multinomial logit models', *Journal of Marketing Research*, **30**, 305-314.

Thaler, R.H. (1985), 'Mental accounts and consumer choice', *Marketing Science*, **4**, 199-214.

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.

Tietenberg, T. (2006), *Environmental and Natural Resource Economics, 5th ed.*, Boston: Addison-Welsley.

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

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 (NTIS Document No. PB94-154101), Washington DC: U.S. Environmental Protection Agency.

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.

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

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, Department of Agriculture and Natural Resource Economics, Colorado State University.

Wang, Y. and Y.S. Zhang (2009), 'Air quality assessment by contingent valuation in Ji'nan, China', *Journal of Environmental Management*, **90**, 1022-1029.

Water Demand Research Team, World Bank (1993), 'The demand for water in rural areas: determinants and policy implications', *World Bank Research Observer*, 8, 47-70.

Weldesilassie, A.B., O. Frör, E. Boelee and S. Dabbert, 2009), 'The economic value of improved wastewater irrigation: a contingent valuation study in Addis Ababa', Ethiopia', *Journal of Agricultural and Resource Economics*, **34**, 428–449

Weisbrod, B.A., (1964), 'Collective consumption services of individual consumption goods', *Quarterly Journal of Economics*, **78**, 471-477.

Willig, R. (1976), 'Consumer's surplus without apology', *American Economic Review*, 66, 589-597.

Whittington, D. (2002), 'Improving the performance of contingent valuation studies in developing countries', *Environmental and Resource Economics*, **22**, 323-367.

Zhao, J. and C.L. Kling (2004), 'Willingness to pay, compensating variation, and the cost of commitment,' *Economic Inquiry*, **42**, 503-517.