

A Common Nomenclature for Stated Preference Elicitation Approaches

Richard T. Carson · Jordan J. Louviere

Accepted: 23 December 2010 / Published online: 12 January 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract It is often difficult to determine what actually was done in work involving data collected with stated preference surveys because the terms used to describe various procedures have ambiguous and sometimes conflicting meanings. Further, terms used to describe data collection procedures often are confounded with terms used to describe statistical techniques. We call for the use of a common nomenclature to describe what was done in a data collection effort for stated preference studies, and take a first step at setting out such nomenclature. We only seek to improve clarity in the communication of research results and take no position here on appropriateness of particular procedures.

Keywords Contingent valuation · Discrete choice modeling · Survey questions

JEL Classification Q51 · D6 · H4 · I18 · M31

1 Introduction

One often reads that a contingent valuation survey was undertaken or choice experiment performed, yet this provides little information about what was actually done. Several papers compare results from contingent valuation studies to choice experiments, yet this also conveys little information about what was actually compared. This state of affairs is widespread, including works by the present authors. Perhaps even worse is the fact that researchers like us sometimes use the same method but refer to it with different names. As there is now a

R. T. Carson (✉)
Department of Economics, University of California, San Diego, CA, USA
e-mail: rcarson@ucsd.edu

R. T. Carson · J. J. Louviere
Centre for the Study of Choice, University of Technology, Sydney, NSW, Australia

large and rapidly growing literature¹ on stated preference methods of various types, we see a compelling need for common nomenclature for methods associated with stated preference work. Thus, the purpose of this paper is to provide a relatively comprehensive nomenclature that can be used with stated preference work which should lead to less confusion and more understanding amongst researchers.

Where possible we use existing terminology if it is sufficiently well-defined, or we make small modifications to ensure clarity. In other cases we suggest that the field should abandon a few oft-used terms like “conjoint” due to a multiplicity of possible and conflicting meanings and/or the terms that, in and of themselves, convey no real meaning. Throughout the paper we denote our preferred nomenclature in italics.

Why does one need a common nomenclature for stated preference work? One key reason is that different stated preference methods have long been known to produce different results (e.g., [Desvousges et al. 1987](#); [Carson et al. 2001](#)); and differences are predicted by neoclassical economic theory ([Carson and Groves 2007](#)), as well as by common psychological alternatives to that theory (e.g., [Tversky et al. 1990](#); [Bateman et al. 1995](#)). Further, the method by which one elicits information related to preferences in any given study matters because they differ in the amounts and nature of information about consumer preferences obtained, the assumptions required to calculate standard measures of economic welfare, and the statistical methods that are appropriate for analysis of the elicited preference data.

A further complication is that there is often confusion between the particular preference *elicitation method* used and the associated (implied) statistical model (e.g., probit, mixed logit, Turnbull) used to fit the information related to preferences elicited. It often is the case (and desirable) that a data collection effort be designed with a particular statistical estimation method in mind to be used to recover information about preferences. However, the data collection effort and the approach used to elicit preferences need to be able to stand on their own with the recognition that different assumptions about the underlying *data generation process* may imply different statistical estimation methods.

2 Stated Preference, Contingent Valuation, Choice Experiments and Conjoint

The literature contains four fairly generic oft-used terms: stated preference, contingent valuation, choice experiment and conjoint.² All are somewhat problematic. We propose retaining the first two, making a small modification to the third (adding the prefix “discrete”) and abandoning the fourth altogether.

Stated preference (SP) is a long-standing, widely used term that encompasses many different methods that can be used in several different contexts. The “stated” preference part has nice historical ties to doing surveys in-person and (later) via the telephone, but typically serves as a foil to “*revealed*” preference (RP) data from observed behavior. Both of the terms “SP” and “RP” are somewhat deceptive insofar as one typically does not obtain data on actual preferences per se (either stated or observed), but instead on data that provides information from which one may infer preferences.³ Further, the statistical methods

¹ [Carson \(2010\)](#) delineates over 7,500 contingent valuation papers and studies from over 130 countries.

² See [Mitchell and Carson \(1989\)](#); [Smith \(2004\)](#) and [Carson \(2010\)](#).

³ There is a long standing debate on whether the more appropriate starting primitive is preference or choice for deriving theoretical results. We follow the standard formulation (e.g., [Varian 1992](#)) of using preference without taking a stance on which is more appropriate, in large part because of the potential for confusion given the heavy use of the term “choice” in specific elicitation formats.

used to extract information about preferences from SP and RP data are often identical. Thus, a simple and straightforward definition is that a stated preference survey is a survey that asks agents questions that embody information about preferences.⁴ It should be clear that this definition does not necessarily have direct ties to preferences defined in an economic sense; consequently, a question that asks agents whether they prefer to have their grandmother or grandfather appear in their dreams is a perfectly good SP question. Information about preferences can be elicited in many forms in SP surveys, but links between elicited information about preferences and preference parameters of an indirect utility function are at best ill-defined without more specificity.

We suggest that the term “stated preference question/survey” be retained, in part because it is a reasonable generic term to describe one type of information that can be collected in a survey, and, in part because of its contrast with RP data which is often also collected in surveys by asking questions about past or current behavior. Note, however, that despite widespread usage, the term “RP” also conveys nothing specific about the data collected, the nature of the good, or how the data can be used to obtain information about preferences. Thus, the terms ‘SP’ and ‘RP’ should be used to signal the general nature of the data collection effort but not as an indicator that a particular elicitation technique or economic approach has been used.

The original meaning of *contingent valuation* (CV) was using a survey approach to place an economic value on a public good (Mitchell and Carson 1989).⁵ Contingent valuation conveys three main elements: (1) information related to preferences is obtained using an SP survey, (2) the study’s purpose is placing an economic value on one or more goods, and (3) the good(s) being valued are public ones (pure public or quasi-public).⁶ The term has a long and rich history and has permeated introductory level textbooks.

The problem with the term “contingent valuation” is that more recently it is often used to refer to one specific elicitation technique, namely a stand-alone single binary choice task (e.g., Hanley et al. 1998; Bennett and Blamey 2001). Indeed, it often is contrasted with choice experiments that involve something more than a single *choice set* and allow (in contrast to CV) one to estimate the *marginal value* of changing *attributes* as well as the *total value* of a good.⁷ This distinction is odd because most early CV studies (e.g., Randall et al. 1974) asked respondents about their valuation of multiple goods and they valued changes in attributes, which was not surprising in light of long-standing economic interest in valuing marginal changes. Indeed, one of the most heavily used CV applications for policy purposes (US Environmental Protection Agency 1994) has been the Carson and Mitchell (1993) water quality study which provided a functional relationship between maximum *willingness-to-pay* (WTP) and changes in water quality based on a series of answers to payment card questions

⁴ One sometimes sees the term stated preference survey used in the literature to refer to a survey that poses choice question(s) to respondents that involve making what are considered to be economic tradeoffs. This would appear though to be confusing a very general term with one of its specific manifestations.

⁵ Mitchell and Carson (1989) discussed a range of elicitation methods and did not see the term “contingent valuation” as specifying a specific elicitation technique.

⁶ 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) described to including the manner of provision.

⁷ Attributes are characteristics or features of goods that can be seen as a bundle of attributes. Sometimes interest is focused on the implications of changing the *level* of a particular attribute and sometimes on implications of supplying a particular configuration of a good.

that varied the level of water quality.⁸ Since the contentious debate over contingent valuation of the 1990's (Carson et al. 2001) there has been sustained interest in obtaining WTP estimates for marginal changes.

In the 1998 *American Journal of Agricultural Economics* article "Stating Preference Approaches for Measuring Passive Use Values: Contingent Valuation and Choice Experiments", which played a large role in propagating the CV-choice experiment distinction, Adamowicz, Boxall, Louviere and Williams carefully noted in the introduction that:

The most common elicitation approach is the two-alternative (referendum) contingent valuation method (CVM) (see Mitchell and Carson). 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.

Unfortunately, in recent years many authors, including us, have sometimes taken the term "CV" to mean a standalone binary choice question; while the term "choice experiment" often is used as if it refers to a totally different approach.

Perhaps as importantly, the term choice experiment often is used to refer to both a single multinomial choice question and a sequence of binary choice questions, which obscures the original definition by Adamowicz et al. (1998) that a choice experiment is a sequence of multinomial choice questions. We propose that the term "contingent valuation" be applied according to its original usage, which is independent of any particular elicitation method; while modifying the term "choice experiment" to specifically refer to a class of elicitation methods.

We propose modification of the term choice experiment to "*discrete choice experiment*" (or, "DCE").⁹ There are two reasons for this proposed modification; (1) the term "choice experiment" is commonly used in some sciences (e.g., biology and physics),¹⁰ and the use of this term has little or nothing to do with what economists or psychologists think of as a choice experiment. Adding another adjective should help eliminate this confusion; (2) the particular adjective "discrete" emphasizes that one is dealing with a particular class of elicitation methods.

There are two essential elements to a DCE: (1) a respondent is asked to make a discrete choice between two or more alternatives in a *choice set*, and (2) the alternatives presented for choice are constructed by means of an *experimental design* that varies one or more attri-

⁸ Carson et al. (1990) use a series of payment card questions to estimate a non-parametric response surface between WTP and two attributes related to air quality, visibility and morbidity related health effects, demonstrating that an experimental design on attributes can be used in conjunction with an elicitation approach intended to elicit a continuous response. More recently, Ready et al. (2006) provide a matching approach with respect to attribute levels. They allow respondents to continuously adjust an attribute level given a fixed cost structure and show how marginal WTP for a change in the attribute level at the respondent's ideal point (conditional on cost) can be obtained.

⁹ Another term in common use for what we are terming a DCE is "stated choice". The title of the well known book "Stated Choice Methods" by Louviere et al. (2000) can be seen as a merging of an elicitation approach "stated choice/DCE" with a suite of modeling techniques often referred to as "choice modeling" that are generally as applicable to SP data as RP data. The term "experiment" is an important adjective to include in the name of this elicitation format since it emphasizes that the researcher exogenously controls and varies one or more elements of the choice situation that agents face. As an elicitation format, a DCE can be implemented in either an SP or RP context.

¹⁰ See for instance Forsgren (1992) for a biology example of how the term choice experiment is used in the context of selecting sexual mates under different conditions, and Hellmuth et al. (1987) for a physics example of how the term choice experiment is used in the context of quantum mechanics.

butes *within-* and/or *between-*respondents to be able to estimate economic quantities tied to preference parameters.¹¹ A *single binary choice* question (SBC) satisfies both these elements. One might be tempted to define the difference between an SBC and what is often called a choice experiment as the capability of valuing changes in more than one attribute; however, as noted earlier, one does not need to ask people to choose between alternatives to value attributes. Eliciting WTP for different programs that differ on attributes conceptually also works. Nor is it the case that one must give people more than one choice set; for example, a single multinomial choice question long has been the workhorse for obtaining choices in RP data. Moreover, SBC formats also can be used to estimate the value of changes in attributes other than *cost*, if one asks statistically equivalent subsamples of SBC questions that differ on the attributes. The simplest variant of this is the well-known *external scope test* (Carson and Mitchell 1995).¹²

We propose the following definition. A *discrete choice experiment* (DCE) is a general preference elicitation approach that asks agents to make choice(s) between two or more discrete alternatives where at least one attribute of the alternative is systematically varied across respondents in such a way that information related to preference parameters of an indirect utility function can be inferred. This definition encompasses many elicitation techniques including both SBC and a sequence of multinomial choice questions. In a later section we offer additional nomenclature to help distinguish between the members of this class.

The term “conjoint” seems to have multiple and conflicting meanings. Its origin was in mathematical psychology (e.g., Krantz and Tversky 1971), where “conjoint measurement” referred to a formal mathematical theory of ordered relations on sets of attribute levels that allowed one to identify and test various algebraic functional forms of decision rules. The term “conjoint analysis” was introduced to the marketing literature by Green and Rao (1971), who used it to refer to the specific method of eliciting preferences derived from conjoint measurement.¹³ Since that time many different variants of so-called “conjoint analysis” methods have been proposed and applied. These various methods have in common goods with multiple “conjoined” attributes and a way to determine the relative value (on some scale) of a change in a good’s attribute; and typically, some form of experimental design was used to construct the goods studied.

Much of the conjoint literature has little in common with economic literature on consumer theory, and many approaches used clearly are ad hoc and inconsistent with standard utility theory. Indeed, there are so many different flavors of “conjoint” in marketing that the term is largely meaningless in referencing an elicitation method or even a class of elicitation

¹¹ The prominent role of the experimental design in being able to infer economic quantities of interest lies behind the inclusion of “experiment” in the acronym DCE.

¹² *Internal scope tests* are based on comparing estimates for goods varying in scope, based on information about preferences supplied by the same respondents.

¹³ There are two major difficulties with conjoint measurement theory. The first is that while its axioms have some relationship to utility theory, this relationship is very restrictive. As a consequence, conjoint measurement theory has largely been abandoned in favor of more general standard neoclassical utility theory (Varian 1992) and its variants, such as prospect theory (Kahneman and Tversky 1979) that allow particular types of deviations. Second, conjoint measurement theory had no error theory associated with it, statistical or otherwise; hence, discovering that preference data do not correspond to simple conjoint models was a contributing factor to the development of statistical error theories in psychology (e.g., Anderson 1970), although those error theories are not used in current conjoint analysis. Instead non-DCE conjoint applications tend to use ad hoc error assumptions. Lastly, we note that conjoint measurement theory borrowed the notion of a full factorial design from statistics as a way of ensuring that all possibly consistent ranking patterns could be classified. Commercial conjoint applications in marketing quickly abandoned the full factorial approach as being too constraining and moved in the direction of estimating simple linear additive models.

methods (Louviere et al. 2010). Economists naturally would think of such a relationship as a hedonic (Rosen 1974) or Lancaster (1971) characteristics model rather than “conjoint analysis”. Most of what economists now call conjoint analysis is what Louviere (1988) grudgingly called “choice-based conjoint analysis” as a way to gain greater acceptance in marketing for the approach. From an economic perspective, ‘choice based conjoint analysis’ was simply an application of a random utility model (McFadden 1974) to suitable discrete choice data from a DCE. Because a DCE elicitation format more accurately reflected the nature of the choices that people faced in markets, it soon became a commonly applied variant of the more traditional ratings or ranking-based conjoint analysis in marketing.¹⁴

Much of the attraction of doing “conjoint” work in environmental or health economics involves doing something that is “not CV”; yet, as earlier noted, this clearly is not the case. Researchers may well want to do something different than observing a single binary choice question, but it is better to state exactly what one is doing rather than to say that one ‘did’ a conjoint study. The dominant “flavor” of conjoint in economics is a DCE with tight links to economic theory; hence, we think it’s best to use the term “DCE” if that is what one is doing. So, we recommend that all use of the term “conjoint” be abandoned for virtually all purposes.¹⁵

2.1 Stated Preferences and Contingent Valuation: A Broader Context

It may be useful to put SP and CV in a broader context of questions asked of respondents in a survey. Arbitrarily, one can divide survey questions into two groups: SP questions and other question types. We do not define ‘other question types’ because they are not of primary interest in this paper. It suffices to say that ‘other question types’ range from standard demographic questions like age to questions that ask people how happy they are on a 10-category rating scale. Neither question type embodies any direct information about preferences.¹⁶ One way to distinguish SP questions is the area in which they are used. For example, marketing and transport are areas where SP questions are widely used. Marketing tends to use SP questions to help price existing private goods and to determine what new private goods consumers are interested in. Transport tends to use SP questions to help provide insights into consumer travel decisions in situations where a mix of private and public alternatives exist.

Contingent valuation can be thought of as short-hand for using SP questions to address standard applied microeconomic welfare issues characterizing public goods (pure and quasi-public) in arenas like culture, environment and health. Continuing across this spectrum, the political realm has an array of questions like “do you prefer candidate X or Y for President”. We put in a catch-all “other category” and also note that lines between different fields may be fuzzy for particular applications (Fig. 1).

¹⁴ Marketing commercial conjoint analysis evolved (Wittink and Cattin 1989) away from respondents ranking a complete set of alternatives (which is amenable to analysis within a DCE paradigm) toward having respondents assign ratings to alternatives. Many conjoint rating applications only involve pairs of alternatives.

¹⁵ To be clear, it is the term conjoint that we urge no longer be used due to its hopelessly ambiguous meaning and lack of a clear tie in many current applications to its original roots. It is beyond the scope of this paper to argue that any particular elicitation technique should or should not be used. We discuss, below, specific techniques other than DCE that are also often lumped in with conjoint such as rating and ranking.

¹⁶ Note, however, that demographic variables like age may be a useful predictor of agents having different preferences, and that it may be possible to make inferences about economic preferences by regressing happiness scores on variables like income.

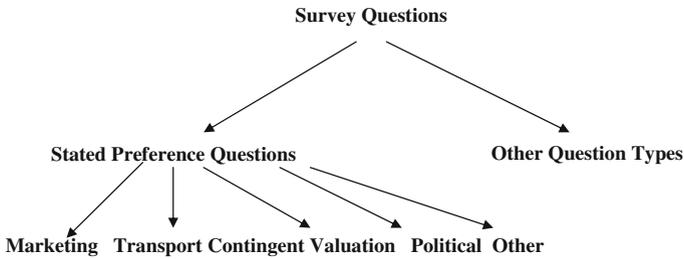


Fig. 1 Typology of survey question types

3 A Simple Typology

Much confusion in the literature arises from using general terms like “stated preference” and “contingent valuation” to refer to specific preference elicitation methods. Elicitation techniques can be categorized by a simple 2×2 table shown in Table 1. The rows of the table represent the nature of the information collected; i.e., one uses a “*matching*” response or a “*discrete choice*” response.¹⁷ We refer to these as matching methods and discrete choice experiments. Columns are represented by the number of SP questions asked, either a *single* stand-alone question or a *sequence* of related questions. While this typology is not perfectly clean,¹⁸ it corresponds well to common practice. We consider the matching/choice distinction first, as this covers most of the basic question types. In most cases, sequential versions are straightforward generalizations of stand-alone counterparts, but there are some important new question formats that emerge in the DCE family when one allows for the possibility of a sequence of questions.

3.1 Matching Methods Versus Discrete Choice Experiments

Two major classes of elicitation methods are heavily used in contingent valuation studies, and both see substantial use in other areas. The first are *matching methods* (MM), where respondents effectively are asked to provide a number (or numbers) that will make them indifferent in some sense. The second are DCEs that effectively ask respondents to pick their most preferred alternative from a set of options.

Table 1 General Classification of Elicitation Techniques

Response	Question number	
	Single stand-alone question	Sequence of questions
Matching		
Discrete choice		

¹⁷ The matching versus choice distinction has long been made by psychologists and they often put it forth as a major underlying source of preference reversals (Tversky et al. 1990). Other terms for matching which are sometimes used are continuous response, equivalence, and pricing.

¹⁸ For instance, a multinomial choice question can always be recast under the standard (but sometimes dubious) independence of irrelevant alternatives assumption as a sequence of binary choices of the chosen alternative versus each of the other alternatives (Chapman and Staelin 1982). For this reason, the typology put forth should not be seen as having direct implications for the properties of elicitation formats along the lines explored by Carson and Groves (2007).

3.2 Matching Questions

The canonical form of the matching approach is the *direct question* (DQ) that asks respondents to state their WTP to obtain a particular public good that they do not currently possess.¹⁹ The matching quantity elicited is the Hicksian compensating surplus measure that makes a respondent indifferent between obtaining the good and giving up the money. One can, of course, change the quantity being matched to a Hicksian *minimum willingness to accept* (WTA) compensation measure by changing the property right. Further, one need not match on money.²⁰ Other metrics like time or a readily tradable commodity also can be used as matching questions to obtain useful information for economic valuation purposes.²¹ In the case of uncertainty, a matching variable also can be the probability of provision of some option relative to a certain alternative.²² A commonly used variant in the health literature is the *standard gamble* (Pliskin et al. 1980; Torrance 1986; Gafni 1994).

The difficulty with direct questions is that few people think about their WTP for a good, whether a marketed or non-marketed one, which often leads to high non-response rates (Mitchell and Carson 1989).²³ Part of this problem may stem from the fact that in most industrialized countries, consumers are used to facing posted prices. The earliest approach (Davis 1963) for getting a matching estimate using this logic was the iterative *bidding game* (BGAME) in which respondents were asked about a tradeoff involving an initial amount and then asked a series of questions with higher or lower amounts depending upon the initial response until an initial yes (no) turned into a no (yes). In principle, without psychological or neoclassical strategic/information-motivated anchoring effects, the bidding game approach can get arbitrarily close to the desired direct question matching response.

A commonly used matching approach is the *payment card* (PCARD) originally proposed by Mitchell and Carson (1981).²⁴ The PCARD approach gives respondents a card with a rectangular display of monetary amounts and asks them for the matching response in the form of an amount on the card or any amount in between. It can be seen as an attempt to overcome the response problem of the direct question and the anchoring effects of the bidding

¹⁹ Another common name for this elicitation format is the open-ended question. We rejected this name because of its frequently used, more general definition in the survey research literature. In this sense, the direct (matching) question can be seen as a special case of an open-ended question where a very specific response is requested rather than letting the respondent express their information about preferences in any form.

²⁰ For instance, Samples and Hollyer (1989) use volunteer time to help preserve endangered species in Hawaii, Shyamsundar and Kramer (1996) use bushels of rice for a park project in rural Madagascar as the matching variables and Echessah et al. (1997) look at willingness to contribute labor in addition to money to control tsetse flies in Kenya. In the health policy literature, a commonly used format described below, asks for a response in terms of life-years. Valuation in non-monetary terms is, of course, also possible with a DCE.

²¹ In some instances, the matching question may be an integer/count, as in the case of a *contingent behavior* question (Englin and Cameron 2002) that asks for how many recreation trips are likely to be taken under a particular set of conditions.

²² In a more general sense, the introduction of uncertainty can change the nature of the welfare measure obtained (Graham 1981). A DQ variant that exploits uncertainty in a different way is the *BDM procedure* (Becker et al. 1964) whereby the agent gives a matching response but whose obligation is related to the outcome of a random device. Carson and Groves (2007) argue that inability to credibly keep the agency from exploiting the continuous information provided in a survey context may result in the BDM procedure having different incentive properties than it does in an experimental economics context.

²³ Moreover, economic theory only suggests that consumers can make decisions when faced with a choice in a specific context.

²⁴ Hammack and Brown (1974) earlier asked a multiple choice question that asked respondents to pick from monetary ranges.

game.²⁵ While it is often assumed that data from the PCARD variant of a matching method is continuous in nature, the response might be better thought of as falling within fairly narrow intervals defined by the payment card (Cameron and Huppert 1991). Clearly, as these intervals become coarser, this approach starts to move in the direction of eliciting a discrete response.

The *time-tradeoff* (TTRADEOFF) approach (Torrance et al. 1972; Bleichrodt and Johannesson 1997) asks respondents for a matching response of how many years of life in one health state (typically better) would be equal to a fixed number of years of life in another (typically worse) state. The technique often uses a visual representation, typically a line with numbers of years, and asks respondents to mark/name a point on the line where they would be indifferent between that many years in the alternative health state relative to the specified number of years in the reference health state. Variants of this approach are similar to either the bidding game (if a starting point is suggested) or the payment card approach (if a specific number is requested) but with time substituted for money as the matching variable. Utility measures derived from such questions are not standard Hicksian money metrics, but instead are expressed in units of the time variable. The *standard gamble* along with the TTRADEOFF approach are common ways to obtain information used to derive *quality adjusted life years* (QALYS).

The last category of matching methods is the *allocation game* (AGAME) (Strauss and Hughes 1976; Mitchell and Carson 1989). Allocation games ask respondents to allocate a fixed quantity between different categories, such as allocating a given monetary amount between different budget categories, or allocating a given amount of time to various activities. In many instances, the AGAME elicitation procedure can be seen as utility maximization subject to a constraint, but translation of allocation game responses to standard economic welfare measures can be difficult.²⁶

3.3 Sequential Versions of Matching Methods

There are single question and sequential question variants of matching methods. Single question versions are denoted by a prefix letter “S” and sequential versions by a suffix “Seq”. Table 2 shows available matching elicitation methods. Sequential versions of stand-alone single question matching methods are straightforward; they ask subsequent versions of an initial question, typically with a variant of the good initially asked. These variants may be ad hoc, randomly assigned, assigned using some type of statistical design or follow a pattern obvious to a respondent. An example of such an ‘obvious pattern’ is when subsequent goods are large (small) or more (less) comprehensive than the initial good. If the scope of the good increases then one has a *bottom-up sequence*. If the scope of the good decreases then one has a *top-down sequence*. Any move from a single question to a series of questions raises issues as to whether respondents treat their responses to different questions as *independent*, and whether they “*learn*” something. The learning here can either be about their preferences or about something that helps them to develop a strategic response as they go through the sequence of questions.

²⁵ Payment cards can be unanchored (just a rectangular array of numbers) or anchored where expenditures on other goods (public or private) are provided for some of the numbers on the card. In addition to the role of anchors, another early concern with the payment card was that the particular amounts used on the card might influence the nature of the matching response (e.g., Rowe et al. 1996). Carson and Groves (2007) have recently examined some of the properties of the payment card.

²⁶ Some of these problems may be overcome by embedding an allocation game in a larger discrete choice experiment (Crouch et al. 2007).

Table 2 Matching methods

Elicitation technique	Single Stand-alone question (acronym)	Sequence of questions (acronym)
Direct question	Single direct question (SDQ)	Direct questions-sequence (DQ-Seq)
Bidding game	Single bidding game (SBGAME)	Bidding games-sequence (BG-Seq)
Payment card	Single payment card (SPCARD)	Payment cards-sequence (PCARD-Seq)
Time tradeoff	Single time tradeoff (STTRADEOFF)	Time tradeoffs-sequence (TTTRADEOFF-Seq)
Allocation game	Single allocation game (SAGAME)	Allocation games-sequence (AGAME-Seq)

3.4 Discrete Choice Experiments

Two basic types of DCE questions can be asked as single stand-alone questions: a *single binary choice* (SBC) question and a *single multinomial choice* (SMC) question. In the case of the multinomial choice question, the number of *alternatives* is $k \geq 3$. We use the prefix letter “S” for ‘single’ for both elicitation methods to help avoid confusion with commonly used sequential variants, which use the suffix “Seq”. In the case of both SBC and SMC elicitation methods, the *scenario* containing the choice alternatives can be laid out in *narrative* form describing the *attributes* of the choice, or in a form that explicitly compares attributes *side-by-side*.²⁷

The distinction between an SBC and SMC is sometimes important from an incentive perspective, particularly in the case of a (pure) public good where only one level of the good can be supplied (Carson and Groves 2007). In other ways the two question formats are quite similar; for example, in the most basic context all that can be observed is the fraction of a sample that favors each alternative. To obtain welfare measurements, an attribute (typically cost) must be varied in some way. The simplest model consists of a set of alternatives, each described by a bundle of not directly comparable attributes and cost. Cost is systematically varied across respondents in such a way that changes in the probability of choosing an alternative as its cost changes can be estimated.²⁸ Of course, one also can make attributes other than cost comparable across the alternative and systematically vary those. Thus, with a sample of sufficient size, both SBC and SMC formats can obtain the full set of information about preferences recoverable from any DCE.

The next DCE to be considered involves a SMC in which respondents are asked not just for their most preferred alternative but also a *complete ranking exercise* (RANKING) over all alternatives presented in the SMC. This elicitation technique has a very long history and is sometimes not thought of as a DCE. However, it is straightforward to show that in a complete ranking exercise a respondent provides information about all possible pair-wise comparisons in the SMC. Under the *independence of irrelevant alternatives* (IIA) assumption, the pair-wise comparisons contain all relevant information. Under more general assumptions, the ranking exercise gives the implied choice for each pair of alternatives, but appropriate estimation requires conditioning on the other choices in the SMC. There are two major issues with RANKING: (1) respondent task difficulty grows as k increases; and if the IIA property

²⁷ Attributes are sometimes referred to as “factors” in the statistical literature on experimental design, as features in the marketing literature, and more generally, as characteristics (Louviere et al. 2000).

²⁸ As an example, consider a choice set comprised of national parks, each of which is unique with respect to its own attributes but has a different cost of visiting from the respondent’s perspective.

is violated, the effective number of parameters that must be estimated can become large because one needs to condition on other alternatives (McFadden et al. 1977); and (2) the implied SBC can have difference variances.²⁹

A less burdensome version of ranking extracts more information than the SMC (but less than RANKING) by asking for both the most and the least preferred out of k alternatives. This elicitation technique is known as the *best-worst choice* (BWCHOICE) question (Marley and Louviere 2005). If there are three alternatives, this format gives a complete ordering of the alternatives and use of BWCHOICE dramatically increases the number of implied binary comparisons (which convey the information content of the response) when $k \geq 3$. Besides increasing statistical efficiency, a rationale for this elicitation technique is that the easiest choices for respondents are likely to be their most and least favorites, particularly as k increases. Further, indicating the least favorite choice may involve less strategic behavior and reveal more statistical information than indicating another choice like a respondent's second favorite.³⁰

A final type of DCE is to ask respondents to divide alternatives into two (or more groups) based on some criteria related to preferences. A popular name (Holbrook et al. 1982) for this approach is "*pick any*" (PICKANY) because respondents are asked to pick any alternatives in the SMC that are acceptable (unacceptable). The presumption is that an alternative in the acceptable set is preferred to alternatives not in this set. Once again, this method can be seen to generate a set of implied binary comparisons. There are a number of variations on this approach that effectively divide alternatives into a larger number of subsets. For example, respondents could be asked to first choose their most preferred alternative, and then to choose any other alternatives that also are 'acceptable'.³¹

3.5 Sequential Versions of DCEs

As in the case of matching methods, from a mechanical perspective, sequential versions of the SBC and SMC are straightforward to develop as one moves from a *single choice set* to *multiple choice sets*. Also, as with matching methods, the key question in moving to a sequence of questions is whether respondents take the questions as independent and/or whether some type of learning occurs as they go through the set of questions. Table 3 displays the five basic types of discrete choice experiments. We use the oft-used term *sequence of binary choice questions* (BC-Seq) instead of the commonly used term 'sequence of paired comparisons' because the latter term often is seen in a more general context (particularly in psychology) to indicate choosing between two objects/stimuli in a way that does not necessarily convey information about preferences, such as choosing which of two images is brighter. Sequential versions of other DCE are straightforward and beyond their being named in Table 3 are not discussed separately.

²⁹ Arguments that respondents evaluate ranking tasks differently are examined by Bateman et al. (2007).

³⁰ There are a number of other variants of the best-worst approach (Marley and Louviere 2005; Flynn et al. 2007; Marley et al. 2008) whereby instead of indicating the best and worst choices, one indicates the best and worst attribute levels in an alternative. That is, the choice set consists of all the attribute levels in a profile and respondents choose the best and worst levels, such as the most and least attractive levels.

³¹ There are also variants that combine RP and SP information. For instance, under the assumption that an alternative chosen by a respondent in a market context such as visiting a particular recreational fishing site is their most preferred alternative (which should be the case if no new information is gained about the site they are visiting), survey respondents can be asked about what other fishing sites from an SMC list they might have visited had the site (at which they are being interviewed) was closed.

Table 3 Discrete choice experiments

Elicitation technique	Single Stand-alone quest (acronym)	Sequence of questions (acronym)
Binary choice question	Single binary choice (SBC)	Binary choice-sequence (BC-Seq)
Multinomial choice question	Single multinomial choice (SMC)	Multinomial choice-sequence (MC-Seq)
Complete ranking exercise	Complete ranking (RANKING)	Complete ranking-sequence (Ranking-Seq)
Best-worst choice question	Best-worst choice (BWCHOICE)	Best-worst choice-sequence (BWCHOICE-Seq)
Alternative subsetting	Alternative subsetting (PICKANY)	Alternative subsetting-sequence (PICKANY-Seq)

One distinction that often matters in sequential DCE's is whether one of the alternatives is the current status quo. Always offering the *status quo alternative* as an option facilitates comparisons between alternatives and facilitates welfare measurement because the status quo alternative is always a *feasible choice* from the respondent's perspective.³² Little of relevance can be learned by a choice set that asks a respondent whether they prefer a Rolls Royce to a Ferrari if the respondent can afford neither. Correct specification of status quo alternatives is critical to defining appropriate *property rights* situations in DCEs, and by extension, the types of welfare measures obtained.³³

A special case of the BC-Seq in Table 3 deserves particular mention due to its extensive use and associated literature, namely the *double-bounded binary choice* question (DBBC) (Carson 1985; Hanemann et al. 1991). This format starts with an initial binary choice question, which is followed by a second binary choice question; the difference in the choices offered in the first and second question is based on responses to the first question. The difference usually involves a change in cost (increased for an initial yes response and decreased for an initial no response) that allows one to *bound* the region where WTP lies to a narrower *interval*.³⁴ The DBBC can be seen to be an example of an *adaptive* DCE experiment whereby the choice set offered on the second condition is *conditioned* on the response to the first question and an experimental design determined in advance. When the two questions are answered independently, substantial gains are obtained from a statistical perspective and analysis is straightforward (Hanemann et al. 1991). Unfortunately, this does not appear to be the typical situation (Alberini et al. 1997), making analysis more difficult although gains in statistical efficiency may be possible. Several reasons for lack of independence between the two questions have been posited (e.g., Carson and Groves 2007).

³² For a pure public good, the appropriate reference level is typically the status quo. Rose et al. (2008) consider experimental designs for DCE that explicitly include a reference alternative where that reference alternative can differ for each respondent.

³³ The most commonly used welfare measures in DCE's are the *Hicksian* ones. Defining the status quo as the current situation yields the *Hicksian compensating surplus* for an imposed quantity change and a *compensating variation* for a price change. *Equivalence surplus* and *equivalence variation* measures are applicable if the status quo property right refers to a different level of consumption than the one currently experienced.

³⁴ It is possible to ask subsequent questions to try to further narrow the interval. Bateman et al. (1995) use a *triple-bounded binary choice* approach. The distinction between this approach and the BGAME elicitation technique become blurred as step size decreases and the number of questions asked increases. Difficulties with the independence of response assumption typically made for the DBBC elicitation method, particularly on the down side, have led to proposals for a *one-and-a-half bound* approach whereby only respondents who say yes initially asked about their willingness to pay a higher amount (Cooper et al. 2002; Bateman et al. 2009).

3.6 Hybrid Methods

One can include matching and DCE questions in the same survey instrument, which we will refer to as *hybrid methods*. Indeed, one sometimes sees a SBC question followed by a DQ question (Whittington et al. 1992). The latter approach has some strategic problems because the SBC question tells respondents that the good can be provided for its stated cost but the extra information given by the DQ question may be useful in reducing the respondent's WTP interval. Many other variants are possible, such as asking an initial matching question to establish a constraint followed by a DCE that offers choices defined by that constraint. Of course, it may be that WTP is ill-defined without a concrete option.

3.7 Ratings Measures

Some elicitation methods cannot give direct estimates consistent with standard (ordinal) utility theory, of which the most obvious are *ratings measures*. Ratings measures often are used by marketers as if they are cardinal measures; and some researchers have tried to extract additional information (beyond the ordinal ranking to which they often reduce) based on strong assumptions about how people use ratings scales. Ratings-type measures are used in many other contexts, such as visual analogue scales sometimes used to measure QUALYs. A different rating variant is a *certainty scale* sometimes used to modify responses from SBC questions statistically. In all these cases, the basic problem is that ratings scales require one to assume that the resulting data are *interval* in nature, not merely *ordinal*. If this assumption is violated, it is unclear how to interpret the numbers provided by rating scales; for example, on a 10 point scale, Liam's response of 3 may mean the same as Isabella's response of 8; the difference is simply how Liam and Isabella treat the scale and/or interpret the scale numbers or verbal labels.

4 Further Terms Useful for Describing What was Done

One can quickly describe and explain what was valued and how it was valued by judiciously including a bit of core information and some consistent terminology that fall into four main groups: (1) characteristics of the good, (2) the experimental design, (3) the survey, and (4) the statistical analysis.

4.1 The Good

Starting with the good, it is important to note the type of good being valued. Essentially, there are three types of goods. (Pure) *public goods* have the standard economic definition of non-excludability and provision of a single level. *Quasi-public goods* are publically provided; and because people can be excluded (e.g., entrance fees or location), different people can experience different levels. *Private goods* are privately provided and by standard definition have the excludability property. While this typology is standard in the literature, dividing lines are sometimes not as sharp. Government regulation can influence what private goods are available and their characteristics and private goods can incorporate attributes that have clear implications for public goods. Contingent valuation is sometimes used in both cases. Also as Carson and Groves (2007) note, these different types of goods can have quite different incentive properties even if one uses the same elicitation technique and it is common

(but incorrect) to make inference about how well contingent valuation works for public goods by reference to its performance with private goods.

Goods need to be paid for, and there are two basic types of *payment mechanisms*, namely *coercive* and *voluntary*. Under a coercive payment mechanism an agent must pay for the good before receiving it, while under a voluntary payment mechanism an agent can potentially enjoy the good without having to pay for it. There are two general types of coercive payment mechanisms: *general taxes* and *user fees*. General taxes are taxes like income taxes not directly tied to the use of the good. User charges are related to the cost of using the good. The price of a good is the simplest example but other examples include commodity taxes, increments to utility bills, gasoline taxes, and entrance fees. The cost of a good can be paid in four ways temporally: (1) a *one-time lump sum* payment; (2) a *continuing payment* over some time interval (e.g., *monthly* or *annual*), possibly indefinite; (3) *payment at time of use*; and (4) whether a commitment is made *ex ante* to pay for the good if it is provided. The latter payment method often applies to a pure public good because the response to the survey can influence the likelihood that the government supplies it, or the respondent gets to decide *ex post* after a good is made available whether they will pay for it and obtain it (common with private goods and quasi-public goods).

4.2 The Experimental Design

Much of the experimental design “language” comes from the statistics literature where properties of experiments for both continuous and discrete response variables have long been and continue to be studied. With specific reference to DCEs involving SP data, the standard reference is Louviere et al. (2000). However, it should be noted that this is an area of active and evolving research (e.g., Ferrini and Scarpa 2007; Street and Burgess 2007) with many unresolved issues and problems, particularly from an applied research viewpoint (Carson et al. 2010). For DCEs, the experimental design is one of two critical elements of the entire study process (the other is correct identification and presentation of attributes and levels). All too often published papers pay little to no attention to details of the experimental design that underlies the data analyzed. Without full disclosure of the details of an experimental design, it is impossible to understand what was done, much less determine whether the analysis results make logical and empirical sense. We now discuss key aspects of experimental design and associated terminology.

4.3 Factorial Designs

In the case of one attribute being varied, such as cost in a SBC task, we refer to this as a “*one factor experiment*”. The values of cost or another attribute being varied are termed “*levels*”. Levels may be associated with inherently *continuous* or *qualitative* attributes, but, regardless of whether attributes are inherently continuous or qualitative, the levels associated with each attribute actually varied in a study are necessarily discrete. Generally speaking, attribute levels represent a sample of the range of variation possible in the attribute. Levels of continuous attributes often are assigned to be evenly spaced between two extreme levels, but more complex designs sometimes are used that rely on *prior* information about preference parameters or latent WTP distributions. Levels of qualitative attributes, including ordinal or mixed ordinal/qualitative attributes are assigned to represent relevant discrete realizations of each attribute.

If two or more attributes are simultaneously varied in such a way that all possible combinations of the attribute levels are presented, we call this a “*factorial design*”. One rarely

sees complete factorial designs in applied research, but this largely is due to past practice and does not mean that factorials cannot be used. Indeed, in many cases one should be able to observe an entire factorial by dividing it into mutually exclusive and exhaustive “versions” and randomly assigning people to one version. An experimental design that involves less than all possible combinations of attribute levels is generally known as a “fractional factorial”. Fractional factorials are discussed in detail in many statistics texts (e.g., [Street and Burgess 2007](#)), so we simply note that the vast majority of fractions used in applications are called “orthogonal main effects plans”, or “OMEs”. These designs are suitable if it is reasonable to assume that the indirect utility function associated with a particular choice process is strictly additive, an assumption that may be frequently incorrect. Thus, it is fair to say that OMEs are used far more often than they should, and their use is especially insidious because they signal that non-additive indirect utility functions are rare, perhaps even non-existent. With OMEs, one cannot reliably estimate any specification other than strictly additive.

Other fractional factorial designs are available that are characterized by the number and types of *interactions* one can estimate. For example, one can construct designs in which the interactions of all attributes with one specific attribute can be estimated; one can construct designs in which selected interactions among attributes can be estimated and one can construct designs in which all possible 2-way *interactions* can be estimated, with many other possibilities. Fractional designs can be “asymmetric” or “symmetric”. Symmetric designs exhibit complete balance in all levels within and between attributes; that is, each level of each attribute occurs equally often and co-occurs with the levels of other attributes equally often. Asymmetric designs may not exhibit within attribute level balance and often co-occurrence of attribute levels is not equal, which typically leads to estimates that differ in precision. Balanced, symmetric designs are “orthogonal”, meaning that each pair of attribute columns is statistically independent because their levels co-occur equally often. Many asymmetric designs also are orthogonal. Designs differ in relationships between omitted attribute effects and observed attribute effects. Such relationships are easy to identify and calculate in symmetric designs; in asymmetric designs omitted effects can be complex linear combinations of omitted effects, making it difficult to calculate or understand.

A key distinction in the experimental design literature involves the nature of the *efficiency criterion* used, where greater efficiency in this context implies reducing the length of the confidence interval(s) for one or more estimated parameters for a given sample size.³⁵ The most commonly used is *d-optimality* (maximizing the determinant of the inverse of the Fisher information matrix) which is a good overall measure if multiple parameters are of interest. A strong case can be made, however, for the use of an alternative measure of efficiency if the final objective is to estimate summary statistics like WTP for one or more specific alternatives. Another key distinction involves whether researchers are willing to assume *non-zero priors* on the parameters of the estimated model as experimental design for non-linear models almost always involves unknown parameters. Often such priors are reasonable given theoretical predictions and knowledge from prior work. With informative priors, substantial improvements in efficiency (i.e., tighter confidence intervals for fixed sample sizes or smaller sample sizes required for specific confidence intervals) are possible. Recent advances in the literature have begun to allow for differences in *heterogeneity* in preference parameters or random components. [Rose and Bliemer \(2010\)](#) summarize the history of the development of experimental design for DCE’s and current issues.

³⁵ Efficiency is sometimes cast equivalently as a more efficient design being capable of achieving a given confidence interval(s) with a smaller sample size.

4.4 Types of DCEs

Generally, a DCE is “*labeled*” or “*alternative-specific*”, or “*generic*”. A *generic* DCE is one in which the options being chosen have no particular meaning other than to identify a bundle of attribute levels; so, generic DCEs typically ask respondents to choose between options 1 to J, or options A to J, where J is the maximum number or letter used to identify alternatives in a choice set. In contrast, a *labeled* DCE always has one or more named options that mean something specific to respondents. For example, in marketing applications, the options often are brand names, whereas in environmental applications the options may be names of specific recreational lakes. The primary distinction in the two types of DCEs lies in the form of the indirect utility function; *generic* DCEs lead to a common functional form for all choice options, whereas *labeled* DCEs allow for different indirect utility estimates for each choice alternative.

4.5 Auxiliary Questions

In addition to describing the good itself, survey instruments typically include several auxiliary questions. The earliest and most common type of auxiliary question is one designed to predict differences in responses to preference elicitation question(s) about the good to help gauge the construct validity of the estimates (Mitchell and Carson 1989). These questions typically can be categorized as (1) *demographic* questions (e.g., age, education, income), (2) *knowledge* about or *use* of a good questions, and (3) *attitudinal* questions related to the good, competing goods, or their method of provision, including general questions about the role of government.

Another type of auxiliary survey question is one that is used to examine sensitivity to or to modify the estimate derived from one or more SP questions in some way. These are sometimes called “debriefing” questions because they follow the main valuation question(s),³⁶ and there tend to be three basic types of these questions: *motivation assessment*, *scenario assessment* and *certainty assessment*. An early and still common example of a motivation assessment question is one designed to identify *protest zeros*; i.e., respondents who might have been willing to pay if some aspect of how the good is provided/paid for differed. For example, asking a respondent “why” they were willing or not willing to pay for a good falls into this class.

Scenario assessment questions focus on specific aspects of a scenario including the characteristics of the good(s) and payment mechanisms. Common examples include asking respondents if they believed that the good would actually be provided and the number of years that they would be required to pay (e.g., Carson et al. 2004).

Moving away from the actual good and its payment and provision mechanism, one also can ask respondents certainty assessment question(s) about their response(s).³⁷ For example,

³⁶ Such questions are distinct from auxiliary language used before the valuation question(s) designed to remind respondents about some aspect of the scenario to which they may have paid insufficient attention. The earliest and most common variant due to recommendation by Arrow et al. (1993) reminds respondents of their *budget constraint* in emphasizing the payment obligation. A more recent variant is “cheap talk” (Cummings and Taylor 1999) that in its softer forms closely resemble a traditional budget constraint reminder and in its more controversial harder forms can alter the perceived nature of the good offered (or the focal choice in multinomial choice questions) by conveying that other individuals regret their original response and do not subsequently pay for the good.

³⁷ Another way to allow for uncertainty is to explicitly provide a “don’t know” response category that is explicitly modeled as falling between “yes” and “no” responses (Wang 1997) instead of taking a (very) conservative view that treats any “don’t know” response as “nos” (Carson et al. 1998).

one can try to elicit a formal *probabilistic* response (e.g., Li and Mattsson 1995) or an *ordinal* scale variant (Champ et al. 1997). An alternative to asking auxiliary questions is to directly incorporate such questions into the elicitation format. Ready et al. (1995) extended a binary choice question into an ordered *polychotomous choice* question with six categories (definitely yes, probably yes, maybe yes, maybe no, probably no, definitely no). Welsh and Poe (1998) modified the payment card elicitation format by asking respondents to classify amounts on the card into a set of five categories (definitely yes, probably yes, unsure, probably no, definitely no); they used these data to obtain WTP estimates based on responses to this *multiple-bounded polychotomous choice* format (Alberini et al. 2003). Treatment of uncertainty is currently an active research area. Uncertainty also characterizes most marketed goods, and it is clear that context matters. For example, in some situations, agents must make decisions at that point in time and thereby *ex ante* decide how to resolve any uncertainty; whereas in other instances agents have an option to wait.

4.6 Survey Administration and Sampling

Survey *instruments* must be administered in some way. Basic administration modes are *in-person surveys*, *telephone surveys*, *mail surveys* and *internet surveys*. Mixed mode surveys also are possible, such as a first-stage telephone survey followed by a second-stage mail survey. A standard in-person survey consists of an *interviewer* verbally asking respondents questions, although in some instances, particularly with complex DCE's, the interview may be computer-assisted. However, it is also possible that interviewers hand out surveys to be *self-administered*, common in on-site recreation surveys or shopping mall intercepts. Surveys completed in this way are not called in-person surveys. Mail and internet surveys are self-administered surveys. In-person and telephone interviews can be conducted by *professional interviewers* or other personnel, often *student interviewers*.

Surveys have *samples* obtained in different ways. *Random samples* can be drawn from a *sampling frame* intended to be representative of a particular *population* or there can be *convenience samples*. It is difficult to extrapolate from a convenience sample to a *population of interest*. Even with random sampling, the sampling frame used may deviate from the population of interest in substantive ways. For instance, *non-response* to the survey by some chosen to be in the sample may influence the ability to extrapolate to the population of interest. The following issues are common to all surveys to varying degrees, and there is a large literature on how to deal with them. Sampling may deviate from *simple random sampling* by *stratification* on observable characteristics (typically geographically) to reduce the risk of drawing a very unrepresentative sample and/or by clustering which increases the number of interviews obtained in a given location that reduces costs but increases sampling variability. *Quotas* on numbers of respondents with particular characteristics often are imposed on samples to help ensure that a sample is representative of the population of interest. After a sample is obtained, *weighting* can be used to correct for non-response under the assumption that it is random, conditional on covariates used to construct the weights.³⁸ More complex corrections for *sample selection bias* may be appropriate if that assumption does not hold and non-response is substantial. Calculation of response rates became increasingly complex over time as new modes of survey administration were introduced, the ability to reach households changed, and

³⁸ The role of weighting has grown in importance over time as response rates from all modes of survey administration have fallen considerably over the last couple of decades. In particular it is worth noting that the 70% response rate recommended by the NOAA Blue Ribbon Panel (Arrow et al. 1993) seems no longer achievable by researchers doing high quality policy research with general population samples in industrialized countries.

the nature of non-response became more complex. Some guidance on calculating response rates is provided by the American Association of Public Opinion Research (2009).

5 Some Concluding Thoughts

We ended our effort to develop a common nomenclature for stated preference approaches before reaching statistical techniques, yet it is clear that statistical estimation plays an important role in valuation estimates. Indeed, we would argue that the number of statistical techniques used to analyze stated preference data with the aim of estimating economic value is sufficiently large that a separate effort to impose some discipline on their naming conventions is needed.³⁹

Finally, we note that while there may be a “best” mode of survey administration, elicitation method or statistical estimator for a particular application, there is unlikely to be a “best” approach for all applications. That is, it may well be that different methods can yield useful information for making decisions about particular issues. What matters is to communicate what was done in a study such that it is broadly understandable.⁴⁰ The field of stated preference research, and specifically contingent valuation, has long been sufficiently mature that some standardization of terms used to describe techniques used should be beneficial to such communications. We hope this paper contributes to that endeavor.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Adamowicz V, Boxall P, Louviere JJ, Williams M (1998) Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *Am J Agric Econ* 80(1):64–75
- Alberini A, Boyle K, Welsh M (2003) Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. *J Environ Econ Manag* 45(1):40–62
- Alberini A, Kanninen B, Carson RT (1997) Modeling response incentive effects in dichotomous choice contingent valuation data. *Land Econ* 73(3):309–324
- American Association for Public Opinion Research (2009) Standard definitions: final dispositions of case codes and outcome rates for surveys, 6th edn. American Association for Public Opinion Research, Deerfield, IL
- Anderson NI (1970) Functional measurements and psychophysical judgment. *Psychol Rev* 73(3):153–170
- Arrow K, Solow R, Portney PR, Leamer EE, Radner R, Schuman H (1993) Report of the NOAA panel on contingent valuation. *Fed Regist* 58:4601–4614
- Bateman IJ, Day BH, Dupont D, Georgiou S (2009) Procedural invariance testing of the one-and-one-half-bound dichotomous choice elicitation method. *Rev Econ Stat* 91(4):807–820
- Bateman IJ, Day BH, Loomes G, Sugden R (2007) Can ranking techniques elicit robust values? *J Risk Uncertain* 34(1):49–66
- Bateman IJ, Langford IH, Turner RK, Willis KG, Garrod GD (1995) Elicitation and truncation effects in contingent valuation studies. *Ecol Econ* 12(2):161–191

³⁹ We are somewhat less sanguine about the success of any such effort because much confusion is associated with biometricians and econometricians giving different names to the same statistical procedures and occasionally the same name to different procedures. Further, use of these estimation techniques is not limited to, and in many instances, not even concentrated on SP data, even within economic valuation.

⁴⁰ Making available a detailed (in print or on-line) appendix containing the survey instrument, as well as information on the experimental design used (if any) and the sampling scheme and its execution, is another obvious step we support.

- Becker GM, DeGroot MH, Marschak J (1964) Measuring utility by a single-response sequential method. *Behav Sci* 9(3):226–232
- Bennett J, Blamey R (eds) (2001) The choice modelling approach to environmental valuation. Edward Elgar, Northampton
- Bleichrodt H, Johannesson M (1997) Standard gamble, time trade-off, and rating scale: experimental results on the ranking properties of QALYs. *J Health Econ* 16(2):155–175
- Cameron TA, Huppert DD (1991) Referendum contingent valuation estimates: sensitivity to assignment of offered values. *J Am Stat Assoc* 86(416):910–918
- Carson RT (1985) Three essays on contingent valuation. Ph.D. dissertation. University of California Berkeley
- Carson RT (1991) Constructed market. In: Braden J, Kolstad C (eds) *Measuring the demand for environmental commodities*. NorthHolland, Amsterdam
- Carson RT (2010) *Contingent valuation: a comprehensive bibliography and history*. Edward Elgar, Northampton
- Carson RT, Conaway MB, Hanemann WM, Krosnick JA, Mitchell RC, Presser S (2004) *Valuing oil spill prevention: a case study of California's central coast*. Kluwer Academic Press, Boston
- Carson RT, Flores NE, Meade NF (2001) Contingent valuation: controversies and evidence. *Environ Resour Econ* 19(2):173–210
- Carson RT, Groves T (2007) Incentive and informational properties of preference questions. *Environ Resour Econ* 37(1):181–210
- Carson RT, Hanemann WM, Kopp RJ, Krosnick JA, Mitchell RC, Stanley Presser S, Ruud PA, Smith VK (1998) Referendum design and contingent valuation: the NOAA Panel's no-vote recommendation. *Rev Econ Stat* 80(2):335–338
- Carson RT, Louviere JJ, Phileas D (2010) Design of discrete choice experiments: a discussion of issues that matter in future applied research. *J Choice Model*
- Carson RT, Mitchell RC (1993) The value of clean water: the public's willingness to pay for boatable, fishable, and swimmable quality water. *Water Resour Res* 29(7):2445–2454
- Carson RT, Mitchell RC (1995) Sequencing and nesting in contingent valuation surveys. *J Environ Econ Manag* 28(2):155–1731
- Carson RT, Mitchell RC, Ruud P (1990) Valuing air quality improvements: simulating a hedonic pricing equation in the context of a contingent valuation scenario. In: Mathai CV (ed) *Visibility and fine particulates*. Air and Waste Management Association, Pittsburgh
- Champ RA, Bishop RC, Brown TC, Daniel W, McCollum DW (1997) Using donation mechanisms to value nonuse benefits from public goods. *J Environ Econ Manag* 33(2):151–162
- Chapman RG, Staelin R (1982) Exploiting rank ordered choice set data within the stochastic utility model. *J Market Res* 19(3):288–301
- Cooper JC, Hanemann WM, Signorello G (2002) One-and-one-half bound dichotomous choice contingent valuation. *Rev Econ Stat* 84(4):742–750
- Crouch GI, Oppewal H, Huybers T, Dolnicar S, Louviere JJ, Devinney T (2007) Discretionary expenditures and tourism consumption: insights from a choice experiment. *J Travel Res* 45(3):247–258
- Cummings R, Taylor L (1999) Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *Am Econ Rev* 89(3):649–665
- Davis RK (1963) The value of outdoor recreation: an economic study of the maine woods. Ph.D. Dissertation, Harvard University
- Desvougues WH, Smith VK, Fisher A (1987) Option price estimates for water quality improvements: a contingent valuation study for the Monongahela River. *J Environ Econ Manag* 14(3):248–267
- Echessah PN, Swallow BM, Kamara DW, Curry JJ (1997) Willingness to contribute labor and money to tsetse control: application of contingent valuation in Busia District, Kenya. *World Dev* 25(2):239–252
- Englin J, Cameron TA (2002) Augmenting travel cost models with contingent behavior data. *Environ Resour Econ* 7(2):133–147
- Ferrini S, Scarpa R (2007) Designs with a priori information for nonmarket valuation with choice experiments: a monte carlo study. *J Environ Econ Manag* 53(3):342–363
- Flynn T, Louviere JJ, Peters TJ, Coates J (2007) Best-worse scaling: what it can do for health care research and how to do it? *J Health Econ* 26(1):171–189
- Forsgren E (1992) Predation risks affects mate choice in Gobiid fish. *Am Nat* 140(6):1041–1049
- Gafni A (1994) The standard gamble method: what is being measured and how it is interpreted. *Health Serv Res* 29(2):207–224
- Green P, Rao V (1971) Conjoint measurement for quantifying judgmental data. *J Mark Res* 8(3):355–363
- Graham DA (1981) Cost-benefit analysis under uncertainty. *Am Econ Rev* 71(4):715–725
- Hammack J, Brown GM (1974) *Waterfowl and wetlands: toward a bioeconomic analysis*. Resources for the Future, Washington, DC

- Hanley N, MacMillian D, Wright RE, Bullock C, Simpson I, Parsisson D, Crabtree B (1998) Contingent valuation versus choice experiments: estimating the benefits of environmentally sensitive areas in Scotland. *J Agric Econ* 49(1):1–15
- Hanemann WM, Loomis J, Kanninen B (1991) Statistical efficiency of double bounded dichotomous choice contingent valuation. *Am J Agric Econ* 73(4):1255–1263
- Hellmuth T, Walther H, Zajonc A, Schleich W (1987) Delayed choice experiments in quantum interference. *Phys Rev A* 35(6):2532–2541
- Holbrook MB, Moore WL, Winer RS (1982) Constructing joint spaces from pick-any data: a new tool for consumer analysis. *J consumer Res* 9(1):99–105
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47(2):263–292
- Krantz DH, Tversky A (1971) Conjoint measurement analysis of composition rules in psychology. *Psychol Rev* 78(2):151–169
- Lancaster K (1971) Consumer demand a new approach. Columbia University Press, New York
- Li CZ, Mattsson L (1995) Discrete choice under preference uncertainty: an improved structural model for contingent valuation. *J Environ Econ Manag* 28(2):256–269
- Louviere JJ (1988) Analyzing decision making: metric conjoint analysis. Sage, Newbury Park, CA
- Louviere JJ, Hensher DA, Swait J (2000) Stated choice methods and applications. Cambridge University Press, New York
- Louviere JJ, Flynn T, Carson RT (2010) Discrete choice experiments are not conjoint analysis. *J Choice Model* 3(3):57–72
- Marley AAJ, Flynn T, Louviere JJ (2008) Probabilistic models of set-dependent and attribute-level best-worst choice. *J Math Psychol* 52(5):281–296
- Marley AAJ, Louviere JJ (2005) Some probabilistic models of best, worst, and best-worst choices. *J Math Psychol* 49(6):464–480
- McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) *Frontiers in Econometrics*. Academic Press, New York
- McFadden D, Train KE, Tye WB (1977) An application of diagnostic tests for the independence of irrelevant alternatives property of the multinomial logit model. *Transp Res Rec* 637:39–45
- Mitchell RC, Carson RT (1981) An experiment in determining willingness to pay for national water quality improvement. Report to the US Environmental Protection Agency. Resources for the Future, Washington, DC
- Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Resources for the Future, Washington, DC
- Pliskin JS, Shepard DS, Weinstein MC (1980) Utility functions for life years and health status. *Oper Res* 28(1):206–224
- Randall A, Ives B, Eastman C (1974) Bidding games for the valuation of aesthetic environmental improvements. *J Environ Econ Manag* 1(2):132–149
- Ready R, Fisher A, Guignet D, Stedman R, Wang J (2006) A pilot test of a new stated preference valuation method: Continuous attribute-based stated choice. *Ecol Econ* 59(3):247–255
- Ready RC, Whitehead JC, Blomquist GC (1995) Contingent valuation when respondents are ambivalent. *J Environ Econ Manag* 29(2):181–196
- Rose JM, Bliemer MCJ (2010) Stated choice experimental design theory: the who, the what and the why. Paper presented at the Sawtooth conference on the use of conjoint analysis in health care, Newport Beach, CA
- Rose JM, Bliemer MCJ, Hensher DA, Collins AT (2008) Designing efficient stated choice experiments in the presence of reference alternatives. *Transp Res B Methodol* 42(4):395–406
- Rosen S (1974) Hedonic prices and implicit prices: product differentiation. *J Political Econ* 82(1):34–55
- Rowe RD, Schulze WD, Breffle WS (1996) A test for payment card biases. *J Environ Econ Manag* 31(2):178–185
- Samples KC, Hollyer JR (1989) Volunteer time as a compensation vehicle in contingent valuation studies of endangered species. In: W-133 benefits and costs in natural resources planning, Western Regional Research Publication Interim Report
- Shyamsundar P, Kramer RA (1996) Tropical forest protection: an empirical analysis of the costs borne by local people. *J Environ Econ Manag* 31(2):129–144
- Smith VK (2004) Fifty Years of contingent valuation. *International yearbook of environmental and resource economics*. Edward Elgar, Cheltenham
- Strauss RP, Hughes GD (1976) A new approach to the demand for public goods. *J Public Econ* 6(3):191–204
- Street D, Burgess L (2007) The construction of optimal stated choice experiments: theory and methods. Wiley, New York

- Torrance GW (1986) Measurement of health state utilities for economic appraisal: a review. *J Health Econ* 5(4):1–30
- Torrance GW, Thomas WH, Sackett DL (1972) A utility maximization model for evaluation of health care programs. *Health Serv Res* 7(2):118–133. Contingent weighting in judgment and choice
- Tversky A, Slovic P, Kahneman D (1990) The causes of preference reversals. *Am Econ Rev* 80(1):204–217
- US 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). Office of Water, United States Environmental Protection Agency, Washington DC
- Varian H (1992) *Microeconomic analysis*. 3rd edn. Norton, New York
- Wang H (1997) Treatment of “Don't Know” responses in contingent valuation surveys: a random valuation model. *J Environ Econ Manag* 32(2):219–232
- Welsh M, Poe G (1998) Elicitation effects in contingent valuation: comparison to multiple bounded discrete choice approach. *J Environ Econ* 36(2):170–185
- Whittington D, Smith VK, Okorafor A, Okore A, Liu JL, McPail A (1992) Giving respondents time to think in contingent valuation studies: a developing country application. *J Environ Econ Manag* 22(3):205–222
- Wittink DR, Cattin P (1989) Commercial use of conjoint analysis: an update. *J Mark Res* 53(3):91–96