

## 6. Stated preference methods and their applicability to environmental use and non-use valuations

**Daniel McFadden<sup>1</sup>**

---

[. . .]it appears that the CVM is likely to work best for those kinds of problems where we need it least; that is, where respondents' experience with changes in the level of the environmental good have left a record of trade-offs, substitutions, and so forth, which can be the basis of econometric estimates of value. But for the problems for which we need something like the CVM most, that is, where individuals have little or no experience with different levels of the environmental good, CVM appears to be least reliable. (Freeman, 1986, p. 160)

### INTRODUCTION

*Stated preference* (SP) methods collect data on consumer tastes by direct elicitation, in contrast to *revealed preference* (RP) methods that infer tastes from observed market demand behavior. Leading SP methods are *choice-based conjoint* (CBC) experiments and surveys, widely used in market research to forecast demand for new or modified products, and *contingent valuation method* (CVM) elicitations, employed by environmental economists to estimate use, non-use, or total values of non-marketed natural resources. CBC and CVM are defined and illustrated in the second section of this chapter. The main subject of this book is CVM, and since the critique of CVM in Hausman (1993), the progress, or lack of progress, in refining this method to the point where it can produce reliable, reproducible, and plausible valuations. This chapter is different, concentrating instead on SP studies of demand for ordinary consumer goods and services where actual market experience provides a proving ground for accuracy of SP methods, and drawing lessons from these market applications for use and non-use valuation of environmental goods.

<sup>1</sup> E. Morris Cox Professor of Economics, University of California, Berkeley.

There are several reasons experience with SP methods in market research matters for CVM. First, one can ask whether the users of CVM could improve their valuations by adopting more of the CBC technology. Second, proponents of CVM for environmental valuations have defended the method by claiming on one hand that CVM is sufficiently close to CBC applied to ordinary market goods so that the demonstrated successes of the latter are support for CVM, and on the other hand that hypersensitivity to context and behavioral inconsistencies found in CVM responses are also seen in CBC studies of ordinary market goods. There is some truth to both premises – CBC studies of demand for ordinary market goods often do exhibit context and behavioral effects, and despite these problems have been relatively successful in demand forecasting for ordinary market goods. However, a closer examination of SP methods for market goods finds a sharp reliability gradient. Forecasts that are comparable in accuracy to RP forecasts can be obtained from well-designed SP studies for familiar, relatively simple goods that are similar to market goods purchased by consumers, particularly when calibration to market benchmarks can be used to correct experimental distortions. However, studies of unfamiliar, complex goods give erratic, unreliable forecasts. For relatively simple environmental goods such as hunting licenses and beach access that are similar to regular market goods, it seems possible to obtain reliable, reproducible use values from well-designed SP studies. However, valuations of relatively complex and unfamiliar environmental goods, particularly for non-use values that have no real market equivalents, are at the bad end of the reliability gradient, and neither CVM nor more robust CBC methods seem capable of producing consistent results. A deeper understanding of the relationship between consumer well-being and stated choices or votes, and major innovations in SP methodology, are needed before SP methods can hope to reliably value complex goods that do not have close market analogues.

Looking back over the 30 years since the Freeman quote that starts this chapter, it is disappointing that CVM has not evolved to overcome its performance issues, so that the concerns of that time are still on the table.<sup>2</sup> Perhaps this is because the task that CVM takes on to elicit consistent non-use values from consumers is in truth impossible to complete – consumers may simply not be up to the job of forming consistent preferences for unfamiliar environmental goods in an experimental setting. Some

<sup>2</sup> The outstanding issues addressed in this book were already concerns expressed by Cummings et al. (1986, pp. 106–9), Freeman (1986), Bishop and Heberlein (1986), Smith et al. (1986), a panel composed of Kenneth Arrow, Danny Kahneman, Sherwin Rosen, and Vernon Smith (in Cummings et al., 1986, pp. 181–236), and Diamond and Hausman (1993).

of the lack of CVM innovation may also be due to over-optimistic assessments by environmental economists of the potential of CVM, followed by a “circling-of-the-wagons” defense against legitimate as well as off-the-mark criticisms.<sup>3</sup> A history of skepticism about stated preference methods within the economics community, described next, may have contributed to the defensiveness of CVM proponents, and to their reluctance to incorporate developments and insights from cognitive psychology, behavioral economics, survey research, and market research that might improve the reliability of at least some environmental valuation tasks.

## HISTORY OF SP METHODS

Stated preference methods date back to the 1930s, when the iconic psychologist Leon Thurstone (1931) made a presentation to the second meeting of the Econometric Society proposing direct elicitation of indifference curves:

Perhaps the simplest experimental method that comes to mind is to ask a subject to fill in the blank space [to achieve indifference] in a series of choices of the following type: “eight hats and eight pairs of shoes” versus “six hats and \_\_\_ pairs of shoes”. . . One of the combinations such as eight hats and eight pairs of shoes is chosen as a standard and each of the other combinations is compared directly with it.

Thurstone introduced psychophysical axioms for preferences that led, via Fechner’s law,<sup>4</sup> to indifference curves that could be interpreted as coming from a log-linear utility function. He collected experimental data on hats vs shoes, hats vs overcoats, and shoes vs overcoats, fit the parameters of the log-linear utility function to data from each comparison, treating responses as bounds on the underlying indifference curves, and used these estimates to test the consistency of his fits across the three comparisons.

At the time of Thurstone’s presentation, empirical demand analysis was in its early days. Frisch (1926) and Schultz (1925, 1928) had published pioneering studies of market demand for a single commodity (sugar), but there were no empirical studies of demand for more than one product. Least-squares estimation was new to economics, and required tedious hand

<sup>3</sup> See, for example, Randall (1986), Mitchell and Carson (1989, pp. 295–7), Carson et al. (2001), and Carson (2012).

<sup>4</sup> Fechner’s law (propounded in *Elemente der Psychophysik*, 1860), also called Weber’s law, states that subjective sensation, quantified in terms of just noticeable differences, is proportional to the logarithm of stimulus intensity. The law was advanced for perceptions of weights and auditory and visual stimuli, and has been verified for a variety of human perceptions, with some exceptions at extremes.

calculation. Consolidation of the theory of demand for multiple commodities was still in the future; for example, Hicks (1939) and Samuelson (1947). Given this setting, Thurstone's approach was path-breaking. Nevertheless, his estimates were rudimentary, and he failed to connect his fitted indifference curves to market demand forecasts and changes in well-being. In retrospect, these flaws were correctable: denote by  $H$ ,  $S$ ,  $C$ , respectively, the numbers of hats, pairs of shoes, and coats consumed, let  $M$  denote the money remaining for all other goods and services after paying for the haberdashery, and let  $Y$  denote total income. If Thurstone had asked subjects for the amounts of  $M$  that made comparison bundles  $(H, S, C, M)$  indifferent to a standard bundle  $(H_0, S_0, C_0, M_0)$ , he could have estimated the parameters of the log-linear utility function  $u = \log M + \theta_H \log H + \theta_S \log S + \theta_C \log C$  by a least squares regression of  $\log(M/M_0)$  on  $\log(H_0/H)$ ,  $\log(S_0/S)$ , and  $\log(C_0/C)$ , and from this forecast the demand for hats at price  $p_H$  and income  $Y$  using the formula  $H = \frac{\theta_H}{1 + \theta_H + \theta_S + \theta_C} \cdot \frac{Y}{p_H}$  derived by utility maximization subject to the budget constraint, with similar formulas for the other goods. He could have plugged these demand functions into the utility function to obtain the log-linear indirect utility function, and from this determined the *net* reduction in income *after* a change in the price of hats from  $p_H'$  to  $p_H''$  that leaves the consumer indifferent to the change,  $HCV = Y \left[ 1 - (p_H''/p_H')^{\theta_H/(1+\theta_H+\theta_S+\theta_C)} \right]$  the *Hicksian compensating variation*. Technical questions could have been raised about the applicability of Fechner's law and the restrictiveness and realism of the log-linear utility that it implies, lack of accounting for heterogeneity in tastes across consumers, and lack of explicit treatment of consumer response errors. Decades later, these issues did arise when the Stone-Geary generalization of this demand system was applied to revealed preference (RP) data.

According to Moscati (2007), Harold Hotelling and Ragnar Frisch panned Thurstone's presentation from the floor. They objected that Thurstone's indifference curves as constructed were insufficient to forecast market demand response to price changes, failing to recognize that extending Thurstone's elicitation to include residual expenditure would have solved the problem. They also pointed out that the knife-edge of indifference that Thurstone tried to elicit is not well determined in comparisons of bundles of discrete commodities. Beyond these objections, Frisch and Hotelling were generally skeptical that experimental, non-market data could be used to predict market behavior. The orthodoxy of that era, formed partly as a reaction to the casual introspections of Bentham and the utilitarians, was that empirical economics should rely solely on revealed market data. Wallis and Friedman (1942) summarized this attitude in an attack that forcefully dismissed Thurstone's method or any other attempt to use experimental data for market demand analysis, pointing out

difficulties in designing experiments that mimic the environment of real market choices: “[Thurstone’s] fundamental shortcomings probably cannot be overcome in any experiment involving economic stimuli and human beings.”

Following the Thurstone presentation, there was no mention of his method in the demand analysis literature until MacCrimmon and Toda (1969). Looking back, this seems narrow-minded, but there was some reason for it. The language of economic analysis, then and now, is prediction of market demand, and assessment of market failures in terms of dollars of equivalent lost income deduced from demand as consumer surplus. Any measurement method that uses experimental data on preferences has to produce convincing results in this language by showing that stated preferences collected outside the market have the same predictive power for market behavior as implied preferences reconstructed from market data. With the advent of behavioral economics, we have learned that people are often not relentless utility maximizers, either in markets or in experiments, undermining the tight links neoclassical consumer theory postulates between consumer utility and demand behavior.<sup>5</sup> This has led to calls for less focus on market demand behavior, and assessment of consumer welfare in terms other than dollars of lost income; see Kahneman et al. (1999) and Kahneman and Krueger (2013). This approach may eventually gain acceptance, but at present market prediction and valuation remain the yardsticks against which any method for eliciting consumer preferences and inferring consumer welfare has to be judged.

The first sustained use of SP methods came out of the theory of conjoint measurement derived from psychophysical axioms by Luce and Tukey (1964) and Luce and Suppes (1965). This method was adapted and named “conjoint analysis” by market researchers like Green and Rao (1971), Johnson (1974, 1999), Shocker and Srinivasan (1974), Green and Srinivasan (1978), Green et al. (1981), Louviere (1988), and Srinivasan (1988) and applied to the study of consumer preferences among familiar market products (e.g., carbonated beverages, automobiles); see Louviere et al. (2000), Rossi et al. (2005), and Ben-Akiva et al. (2016). A central feature of conjoint analysis is use of experimental designs that allow at least a limited mapping of the preferences of each subject, and multiple measurements that allow estimates of preferences to be tested for consistency.

Early conjoint analysis experiments described hypothetical products in terms of levels of attributes in various dimensions, and asked subjects to rank attributes in importance, and rate attribute levels. Market

<sup>5</sup> See, for example, Fischhoff and Manski (1999), Kahneman and Tversky (2000), Camerer et al. (2004), and McFadden (2014a).

*Table 1 A typical CBC menu: paper towels*

Attribute	Alternative 1	Alternative 2	Alternative 3
Price/roll	\$2.29	\$1.54	\$1.25
Sheets/roll	110	58	117
Absorptive capacity (oz. of water per sheet)	2X	3X	1X
Strength when wet (compared to standard towels)	2.5X	1X	1.5X
Brand	Bounty	Viva	Brawny
Check here _____ if you would not pick any of these. Otherwise, check your choice on the right	_____	_____	_____

researchers used these measurements to classify and segment buyers, and target advertising, but they proved to be unreliable tools for predicting market demand. However, Louviere and Woodworth (1983) and Hensher and Louviere (1983) introduced CBC elicitations that directly mimicked market choice tasks, offering respondents repeated menus of products with various attribute levels and prices, and asking them for the choice they would make if the menu offering was fulfilled. McFadden et al. (1986) and McFadden (1986) showed how these elicitations could be analyzed using the tools of discrete choice analysis and the theory of random utility maximization. Subjects would be presented with a series of menus of products. Each product offered in each menu would be described in terms of price and levels of attributes, and perhaps be offered as a sample to handle or taste. Subjects would be asked to choose their most preferred product in each menu. For example, as illustrated in Table 1, subjects might be offered menus of paper towels, with each product described in terms of price, number of towel sheets per roll, a measure of the absorption capacity, a measure of strength when wet, and brand name. Choice data from these menus, within and across subjects, could then be handled in the same way as real market choice data to estimate money-metric indirect utility functions and use them to calculate Hicksian compensating variations for changes in product availability, attributes, and prices. Choice-based conjoint surveys analyzed using discrete choice methods have become widely used and accepted in market research to predict the demand for consumer products, with a sufficient track record so that it is possible to identify many of the necessary conditions for successful prediction; see Green et al. (2001), Cameron et al. (2013), and McFadden (2014b).

Environmental economists developed independently an SP technique

**BOX 1 A TYPICAL CVM REFERENDUM ELICITATION<sup>a</sup>**

There is a population of several million seabirds living off the Pacific coast, from San Diego to Seattle. The birds spend most of their time many miles away from shore and few people see them. It is estimated that small oil spills kill more than 50,000 seabirds per year, far from shore. Scientists have discussed methods to prevent seabird deaths from oil, but the solutions are expensive and extra funds will be required to implement them. It is usually not possible to identify the tankers that cause small spills and to force the companies to pay. Until this situation changes, public money would have to be spent each year to save the birds. We are interested in the value your household would place on saving about 50,000 seabirds each year from the effects of offshore oil spills.

If you could be sure that 50,000 seabirds would be saved each year, would you agree to pay \$5 in extra federal or state taxes per year to support an operation to save the seabirds? (The operation will stop when ways are found to prevent oil spills, or to identify the tankers that cause them and make their owners pay for the operation.)

Yes \_\_\_\_\_ No \_\_\_\_\_

Note: a. The specified payment in each elicitation was a randomly assigned value in \$5, \$25, \$60, \$150, \$400.

termed the *contingent valuation method* (CVM), and applied it to valuing natural resources. For a complete definition of CVM see Randall et al. (1974) or Cummings et al. (1986). In a typical example, taken from Green et al. (1998), CVM asks each respondent one question, in what is termed “referendum format,” as illustrated in Box 1.

CVM questions are typically embedded in a survey that instructs the respondent on the nature of the good, payment arrangements and the circumstances under which the hypothetical offer might be fulfilled, and also collects data on the respondent’s background. This method traces its beginnings to a proposal by Ciriacy-Wantrup (1947) and an article and PhD thesis by Robert Davis (1963a, 1963b) on the use-value of Maine woods. The development of CVM in essentially its current form was due to Randall et al. (1974). Its first published applications for valuation of environmental public goods seem to have been Hammack and Brown (1974), Brookshire et al. (1976, 1980), and Bishop and Heberlein (1979). While CVM has been widely used in environmental economics and beyond, its methodological development has occurred almost entirely within a tight circle of environmental economists who emphasize the unique features of environmental applications and have been selective in incorporating findings from research in marketing, cognitive psychology, and behavioral economics; see Carson et al. (2001).

CVM can be viewed as a truncated form of CBC analysis with three important differences. First, CVM most commonly elicits a single or a small number of stated votes on hypothetical referendums, and consequently does not have the experimental design features of CBC that allow extensive tests for the structure and consistency of stated preferences. Second, CVM as it has developed has been utilized primarily for valuation of environmental public goods, and as a consequence often does not have predictive accuracy in markets as a direct yardstick for reliability. Third, the environmental goods in a CVM experiment are usually complex and unfamiliar, and are described in words, numbers, and/or pictures that are very difficult to present in a way that is complete and balanced, and at the same time sufficiently succinct and graphic to keep the subjects' attention.

Other elicitation methods for stated preferences, termed "vignette analysis" and measurement of "subjective well-being," have become popular among some applied economists and political scientists; see Rossi (1979), King et al. (2004), Caro et al. (2012), Kahneman and Krueger (2013). Vignette analysis uses detailed story descriptions of alternatives, often visual, and may improve consumer information and understanding. Vignette presentations of alternatives can be used within conjoint analysis experiments, and may improve subject attention and understanding of alternatives.

Subjective well-being methods elicit overall self-assessments of welfare, often on Likert or rating scales similar to those used in the early days of conjoint analysis; see Kahneman and Krueger (2013). The conditions under which these methods are reliable enough for policy conclusions are still largely undetermined. In the instances where vignette and subjective well-being methods have been tested, they have proven to be strongly influenced by context and anchoring effects that would tend to reduce forecast accuracy in market demand forecasting applications; see Deaton (2012).

To this day, SP methods, and particularly CVM, remain controversial in the economics community, and SP results are often dismissed, frequently with cause but sometimes without. In the remainder of this chapter, I discuss the track record of CBC for forecasting demand for ordinary market products, and experimental design features and circumstances that seem to be required for reliable demand forecasts. I conclude by drawing lessons from this record for applications of CBC and CVM methods for use and non-use valuations of environmental public goods.

## CHOICE-BASED CONJOINT STUDY DESIGN

A CBC analysis offers subjects a series of menus of alternative products with profiles giving levels of their attributes, including price, and asks them



to identify which product they most prefer in each menu. The menus of products and their descriptions are designed to realistically mimic a market experience, where a consumer is presented with various competing alternatives and chooses one of the options. By changing the attribute levels available for the included products and presenting each consumer with several menus, the researcher obtains information on the relative importance that the consumer places on each of the attributes. The classic CBC setup in marketing might be a laboratory experiment where subjects are asked to sample actual products with the different profiles, and then asked for their choices from different menus. For example, subjects might be given tastes of cola drinks from menus, with various degrees of sweetness, carbonation, flavor, and price for the different products, and asked to pick one from each menu. However, CBC can also be used for familiar products whose features are simply described in words and pictures, with subjects asked to choose from a menu of products based on these descriptions. The paper towel example in Table 1 mimics the menus a consumer sees when going to Amazon.com to look for these products. A major application of CBC in market research has been to experiment on automobile brand and model choice. These studies describe alternatives in terms of price and attributes such as horsepower, fuel consumption, number of seats, and cargo space, and in some cases give subjects experience in a driving simulator and the opportunity to consult reviews and man-on-the-street opinions. These studies can determine with considerable predictive accuracy the distributions of preference weights that consumers give to various vehicle features, and the automobiles they will buy; see Urban et al. (1990, 1997); Brownstone and Train (1999), Brownstone et al. (2000), and Train and Winston (2007).

The levels of attributes of the products offered on different menus can be set by experimental design so that it is possible to separate statistically the weights that consumers give to the different attributes. In the early days, menu designs were often of a “complete profile” form that mimicked classical experimental design and allowed simple computation of “part-worths” from rating responses, but currently the emphasis is simply on ensuring that menus are realistic and incorporate sufficient independent variation in the attributes so that the impact of each attribute on choice can be isolated statistically.

Conjoint analysis methods can be expected to work relatively well for preferences among consumer market goods when the task is choice among a small number of realistic, relatively familiar, and fully described alternatives, with clear and well-understood incentives for truthful response. The idea behind incentives is that when subjects have a realistic chance of really getting what they say they prefer, and they understand this, they have a

positive disincentive to misrepresent their preferences and risk getting an inferior outcome. Studies of conjoint methods show that they are in general less reliable and less directly useful for predicting behavior when the task is to rate products on some scale, or to adjust some attribute (e.g., price) to make alternatives indifferent; these seem to induce cognitive “task-solving” responses different from the task of maximizing preferences; see Wright and Kriewall (1980), Chapman and Staelin (1982), and Elrod et al. (1992). Asking follow-up questions within a single menu also seems to induce a different mind-set than simple choice. For example, a study might follow up a stated choice with a question about the second-best choice among the remaining alternatives, a question as to whether the consumer would stay with first stated choice if the price of one of the alternatives were reduced, or questions about the perceived attributes of various alternatives. Empirical experience is that such follow-up questions elicit responses that are not always consistent with the initial stated choices, even though they do not differ much in their framing from market experiences; see Beggs et al. (1981). The explanation may be that the initial menu “anchors” perceptions and shadows subsequent responses, or that follow-up questions induce a “bargaining” mind-set that invites strategic responses; see Hanemann et al. (1991) and Green et al. (1998). Responses to follow-up questions on preferences can also be colored by self-justification.

Conjoint methods can be expected to be less reliable when the products are unfamiliar or incompletely described, or involve public good aspects that induce respondents to incorporate social welfare judgments; for example, when preferences for automobile models are stated in an elicitation that emphasizes the energy footprint of the models, and environmental consequences. Valuation of non-use aspects of natural resources are particularly challenging for conjoint methods because these applications seek to measure preferences that are outside normal market experiences of consumers.

There are six important issues that need to be considered when designing a CBC study. These also apply to CVM considered as a variety of CBC analysis, although the focus on market forecasting reliability applies only to CVM valuation of lost use, since non-use valuations are neither constrained nor disciplined by market benchmarks:

- Familiarity is important. If subjects are experienced with the products or services, and the attributes that are being assessed, then they seem to make more consistent and predictive choices. If possible, subjects should have the opportunity to test for themselves their subjective sensations from different attribute levels. For example, in a study of consumer choice among streaming music services with

various attributes, it should improve prediction to give subjects a hands-on experience with different features. These opportunities to acquire and validate information and impressions of products should resemble their opportunities to investigate and experience these features in a real market. This might be done with mock-up working models of the products, or with computer simulation of their operation. However, there is a trade-off: attempting to train consumers, and providing mock-ups, can inadvertently create anchoring effects. Consumers who are unfamiliar with a product may take the wording in the training exercises about the attributes, and the characteristics of the mock-ups, as clues to what they should feel about each attribute. Even the mention of an attribute can give it more prominence in a subject's mind than it would have otherwise. The researcher needs to seriously weigh the often-conflicting goals of making the subject knowledgeable about the products and avoiding influencing their relative values of attributes.

- The researcher needs to decide whether to offer an “outside alternative” in the choice sets, and, if so, how to characterize it to the subjects. The inclusion of a realistic “no purchase” option allows estimation of market shares and price elasticities, while experiments without this option can only be used to estimate demand conditioned on a purchase. If the outside option is included, it is important that the meaning of the option be clearly delineated to subjects. For example, in a car choice exercise, does “no purchase” mean that the subject would use a vehicle that the household currently owns and reconsider options next year, or what? In a CVM referendum response, will there be opportunities later to support interventions the subject considers more appropriately scaled or more cost-effective? A danger is that the “no purchase option” can be interpreted differently by different subjects, and can easily become a way for subjects to avoid the effort of resolving difficult trade-offs. Whether and how to include an outside option is an important experimental design decision. If it is not included, it will be necessary to use external market share data to constrain or calibrate the choice model fitted to the CBC data so that it can make complete market demand predictions.
- If possible, the conjoint study should be “incentive compatible”; that is, subjects should have a positive incentive to be truthful in their responses. For example, suppose subjects are promised a Visa cash card, and then offered menus and asked to state whether or not they would purchase a product with a profile of attributes and price, with the instruction that at the end of the experiment, their choice from

one of their menus will be delivered, and the price of the product in that menu deducted from their cash card balance. If they never choose a product, then they get the full Visa balance. If subjects learn, perhaps with training or experience, that it is in their interest to say they would choose a product if and only if its value to them is higher than its price, then they have a positive incentive to be truthful, and the experiment is said to be “incentive aligned” or “incentive compatible.”

- In many situations it will not be practical to provide an incentive-compatible format while maintaining the objectives of the analysis. The researcher might want to consider combinations of attributes that are not currently available, such as testing consumers’ reactions to new features during the design phase of a manufacturers’ product development. For existing but expensive products, a lottery that offers a chance of receiving a chosen alternative may be incentive compatible in principle, but the probabilities required to make it practical may be so low that subjects do not take the offer seriously. For example, suppose a CBC experiment on preferences for automobiles asks a subject to choose between a car with a selling price of \$40,000 and \$40,000 in cash, and told that she has a 1 in 10,000 lottery chance of receiving her choice. If she declines the car when her true value  $\$V > \$40,000$ , then her *expected* loss is  $\$V/10,000 - \$4$ , a small number. This incentive is still enough in principle to induce a rational consumer to state truthfully whether she prefers the car. However, misperceptions of low-probability events, mistrust of lotteries, and attitudes toward risk may in practice lead the consumer to ignore this incentive or view it as insufficient to overcome other motivations for misleading statements.
- The researcher needs to decide how “far down” to explore stated preference orderings. Subjects’ first choice (i.e., most preferred option) is most natural to consumers, since it mimics their regular purchasing task. Second choice, third choice, and so on, can be colored by framing dynamics and may be less reliable for predicting market behavior; see Beggs et al. (1981), McFadden (1981), Louviere’s (1988) “best-worst” choice setup, Green et al. (1998), Hurd and McFadden (1998), and List and Gallet (2001).
- Where possible, CBC results should be tested against and calibrated to consumer behavior in real markets. In some cases, CBC menus will coincide with product offerings in existing markets. In this case, it is useful to compare models estimated from the CBC study and the market data to assess whether people are weighing attributes similarly. Improved forecasts may be obtained by imposing real market

constraints such as product shares on the estimation of choice models from CBC data, by calibrating CBC model parameters to satisfy market constraints, or by combining CBC and market choice data and estimating a combined model with scaling and shift parameters for CBC data as needed; see Hensher et al. (1999).

- CBC studies should when possible embed tests for response distortions that are commonly observed in cognitive experiments, such as anchoring to cues in the elicitation format, reference point or status quo bias, extension neglect, hypersensitivity to context, and shadowing from earlier questions and elicitations. While some of these cognitive effects also appear to influence market choices, many are specific to the CBC experience and have the potential to reduce forecasting accuracy. Ideally, a well-designed CBC study will not show much sensitivity of its bottom-line willingness-to-pay (WTP) values to these sources of possible response distortion.

The following subsections expand on some of these conditions and other important requirements for reliable demand prediction using CBC data.

### **Sampling and Recruitment**

Target populations may differ depending on the objectives of the study – for example, current users, current and potential users, and the general population. An important consideration is whether the target population is individuals, families considered as unitary decision-makers, or family or social group with related but not identical preferences, and in the latter cases how to identify a knowledgeable spokesperson for the group. It is important that the sampling frame draw randomly from the target population, without excessive weighting to correct for stratification and non-response, or use of convenience samples that can contain unobserved sampling biases. However, not all members of the target population may have the background needed to make informed product choices. Then it may be more informative to study the preferences of experienced users, and separately study the differences in users and non-users. An example might be study of consumer demand for relatively esoteric technical attributes of products, say the levels of encryption built into telecommunications devices, where only technically savvy device users will appreciate the meaning of different encryption levels. In this case, a good study design may be to conduct an intensive conjoint analysis on technically knowledgeable users, and separately survey the target population to estimate the extent and depth of technical knowledge, and the impact of technical information on the purchase propensities of general users and non-users.

Relatively universal Internet access has led to less expensive and more effective surveying via the Internet than by telephone, mail, or personal interview. However, it is risky to use Internet convenience samples recruited from volunteers, as even with weighting to make them representative in terms of demographics, they can behave quite differently than a target population of possible product buyers. Better practice is to use a reliable method such as random sampling of addresses, then recruit subjects for the Internet panel from the sampled addresses. It is important to compensate subjects for participation at sufficient levels to minimize selection due to attrition; see McFadden (2012). Experience with “professional” subjects who are paid to participate in Internet panels is positive: subjects who view responding as a continuing “job” with rewards for effort are more attentive and consistent in their responses.

### **Experimental Design**

The design of a conjoint experiment establishes the number of menus offered to each subject, the number of products on each menu, the number of attributes and attribute levels introduced for each product, and the design of the profiles of the products placed on each menu. Some other aspects of a conjoint study, the setup and introduction to the experiment given to each subject, subject training, and incentives, might be considered components of the design, but will here be treated separately. There are four distinct considerations that enter conjoint experimental designs.

The first consideration is that for good statistical identification of the valuations of separate attributes, the design needs to allow considerable linearly independent variation in the levels of different attributes, and a considerable span of attribute levels. The classical statistical literature on experimental design focused on analysis of variance and emphasized orthogonality properties that permitted simple computation of effects, and treatments that provided minimum variance estimates. Designs that reduce some measure of the sampling variance under specified model parameters (such as the determinant of the covariance matrix for “D-efficiency”) have been implemented in market research by Kuhfield et al. (1994), Bliemer and Rose (2009), Rose and Bliemer (2009), and others. It is important that conjoint studies be designed to yield good statistical estimates, but there is relatively little to be gained from adherence to designs with classical completeness and orthogonality properties. First, with contemporary computers, the computational simplifications from orthogonal designs are usually unimportant. Second, for the non-linear models used with CBC, orthogonality of attributes does not in general minimize sampling variance. Unlike classical analysis of variance problems, it is not usually

possible in non-linear choice models to specify efficient designs in advance of knowing the parameters that are the target of the analysis.

The second consideration is the formatting, clarity, and prominence of attributes and prices of products presented in CBC studies. These presentations are critical aspects of real market environments, and are correspondingly important in realistic hypothetical markets. Advertising and point-of-sale product presentations in real markets often feature “hooks” that attract the consumers’ attention and make products appealing, and understate or shroud attributes that may discourage buyers. Thus, prices may not be prominently displayed, or may be presented in a format that shrouds the final cost; for example, promotions of “sales” or “percent-off” discounts without stating prices, statements of prices without add-ons such as sales taxes and baggage fees, and subscriptions at initial “teaser” rates. Products like mobile phones, automobiles, and hospital treatments are often sold with total cost obscured or shrouded, often through ambiguous contract terms, through a two-part tariff that combines an access price and a usage fee, or through framing (e.g., “pennies a day”). A CBC study that is reliable for forecasting evidently needs to mimic the market in its presentation of product costs, incorporating the same attention-getting, persuasion, ambiguities and shrouding that consumers see in the real market.

The third consideration is that relatively mechanistic statistical approaches to setting attribute levels may lead to profiles that are unrealistic, or are dominated by the profiles of other products on a menu. Considerable care is needed to balance statistical objectives with realism of the experiment; see Huber and Zwerina (1996). Menus and their framing that are unlike familiar market settings invite cognitive responses that differ from those that appear to determine preferences and drive choices in market settings. There is a tendency for subjects to approach surveys as if they were school exams – they cast about for “correct” answers by making inferences on what the experimenter is looking for. While some may use their responses to air opinions, most give honest answers, but not necessarily to the question posed by the experimenter. They may “solve” problems other than recovering and stating their true preferences, indicating instead the alternative that seems the most familiar, the most feasible, the least cost, the best bargain, or the most socially responsible; see Schkade and Payne (1993).

The fourth consideration is that prominence and ease of comparison are known to be factors that influence the attention subjects give to different aspects of decision problems; for example, there is a claim that subjects in their stated choices systematically place more weight on price relative to other product attributes than they do in real markets, perhaps because this

dimension is clearly visible and comparisons are easy in a conjoint analysis menu, whereas prices in real markets often come with qualifications and may not be displayed side by side. Widespread folklore in marketing is that subjects have trouble processing more than six attributes and more than four or five products, and begin to exhibit fatigue when making choices from more than 20 menus; see Johnson and Orme (1996). Beyond these limits, they appear to use *filtering* heuristics, taking “short cuts” by eliminating consideration of some products and attributes using simple heuristics, and considering trade-offs only on the remainder. Often conjoint analysts will address this behavior by limiting the dimensionality of the attribute profile, explicitly or implicitly asking subjects to assume that in all other dimensions, the products are comparable to brands currently in the market. This leaves subjects free to make possibly heterogeneous and unrealistic assumptions about these omitted attributes, or requires them to digest and remember lengthy specifications for omitted attributes and their assumed levels. These design restrictions may make responses more consistent and easy to analyze, but they may not improve prediction. Filtering heuristics also seem to be used in real markets with many complex products, such as the market for houses. If the primary focus of the conjoint study is prediction, then the best design may be to make the experiment as realistic as possible, with approximately the same numbers and complexity of products as in a real market with similar products, and possibly sequential search, so that consumers face similar cognitive challenges and respond similarly even if decision-making is less single-minded than neoclassical preference maximization. However, if the primary focus is measurement of consumer welfare, there are deeper problems in linking well-being to demand behavior influenced by filtering. While it may be possible to design simple choice experiments that eliminate filtering and give internally consistent statements of consumer welfare, there is currently no good theoretical or empirical framework for using filtering-influenced consumer choice in either real or hypothetical markets to calculate neoclassical economic measures of well-being.

### **Subject Training**

Extensive experiments from cognitive psychology show that context, framing, and subject preparation can have large, even outside, effects on subject response. It is particularly important that subjects have familiarity with the products and features they are being asked to evaluate that is comparable to their real market experiences, as attention, context, and framing effects are particularly strong when subjects are asked to respond in unusual or unfamiliar circumstances. Familiarity may be automatic if



the target population is experienced users of a particular line of products. For inexperienced users, tutorials on the products and hands-on experience can reduce careless or distorted responses, but may also influence stated preferences in ways that reduce forecasting accuracy.

It is useful to recognize that training of subjects can occur at several levels, and that training can manipulate as well as educate, leading to unreliable demand predictions. First, subjects have to get used to answering questions that may be difficult or intrusive, and learn that it is easier or more rewarding to be truthful than to prevaricate. Some of this is mechanical: practice with using a computer for an Internet-based conjoint survey, and moving through screens, buttons, and branches in a survey instrument. Second, subjects need to be educated as to what the task of stating preferences is. Subjects can be taught in “Decision-Making 101” how to optimize outcomes with assigned preferences, and how to avoid mistakes such as confusing the intrinsic desirability of a product with process issues such as availability or dominance by alternatives. Such training can be highly manipulative, leading to behavior that is very different from and not predictive for real market choices. But real markets are also manipulative, providing the “street” version of “Decision-Making 101” that teaches by experience the consequences of poor choices. The goal of a conjoint study designed for prediction should be to anticipate and mimic the training that real markets provide. Third, the study designer needs to determine what information will be conveyed to the subject, in what format, and assess what information the subject retains and understands. Typically a conjoint survey begins by describing the types of products the subject will be asked to evaluate, their major attributes, and the structure of the elicitation, asking for most preferred alternatives from a series of menus. Details may be given on the nature and relevance of particular attributes. Instructions may be given on the time the subject has to respond, and what rules they should follow in answering. For example, the survey may either encourage or discourage the subject from consulting with other family members, finding and operating past products in the same line, or consulting outside sources of information such as Internet searches. Finally, subjects need to be instructed on the incentives they face, and the consequences of their stated choices. At various stages in subject training, they may be monitored or tested to determine if they have acquired information and understand it. For example, a protocol in market research called “information acceleration” gives subjects the opportunity to interactively access product descriptions, consumer reviews, and media reports, and through click-stream recording and inquiries during the choice process collects data on time spent viewing information sources and its impact on interim propensities. This protocol seems to improve subject

attention and understanding of product features, and also identify the sources and content of information that has high impact on stated choices; see Urban et al. (1990, 1997).

In summary, while training may educate subjects so they are familiar with the products being compared, it is difficult to design training that is neutral and non-manipulative. Real markets are in fact often manipulative, via advertising and peer advice, and one goal for CBC is to achieve accurate prediction by mimicking the advertising and other elements of persuasion the consumer will encounter in the real market. One caution is that particularly in cases where preferences are not well formed in advance, subjects will be particularly vulnerable to manipulation, and training that embodies manipulation that is not realistic risks inducing stated responses that are not predictive for real market behavior.

Training and context seem to have particularly strong effects on subjects asked to value complex and unfamiliar environmental goods, particularly non-use valuations. The suggested benchmark for market goods, that training and information presentations in hypothetical elicitation be designed to mimic these processes in real markets, is obviously not available for non-use valuations. A fundamental question in this case is whose preferences are being solicited, the perhaps poorly formed and unformed preferences of untrained consumers in the general population, or the presumably informed, but possibly not representative, educated preferences of trained jurors. In the first case, CVM mimics direct democracy by referendum, as practiced in Switzerland or California, and in the second case, elicitation from trained respondents mimic the choices of presumably informed grand jurors, judges, or legislators. In practice, direct referendums and decisions by untrained juries are seldom cited as models for thoughtful, consistent public decision-making, although our judicial system mandates their use to avoid systematic biases that can enter the decisions of professional experts. If one is in a well-functioning, effective, and informed representative democracy, admittedly a big if, then one would expect that professional legislators aided by their experts would understand the issues and the preferences of their constituents, and would provide the most reliable mechanism for expressing values for environmental goods. Then, well-trained respondents in carefully designed SP experiments might be envisioned as providing the “expert valuations” that legislators need as input to rational public policy decisions on environmental issues.

### **Incentive Alignment**

Economic theorists have developed mechanisms for incentive-compatible elicitation of preferences for both private and public goods. The simplest

offer the subject a positive probability that every stated choice will be *pivotal* and result in a real transaction. If subjects understand the offer, the probabilities are sufficiently high so that the subject does not dismiss them as negligible, and subjects view the transactions as being paid for from their own budgets rather than in terms of “house money” that they feel is not really theirs, then it is a dominant strategy for the subject to honestly state whether or not a product with a given profile of attributes is worth more to them than its price; this is shown by Randall et al. (1974) and Green et al. (1998) for a leading variant of this mechanism due to Becker et al. (1964). For more general settings, including menus with multiple alternatives, McFadden (2012) shows that the Groves-Clarke mechanism (Groves and Loeb, 1975) is incentive-compatible when consumers empaneled in an informed jury embrace the incentives. However, the experimental evidence is that people have difficulty understanding, accepting, and acting on the incentives in these mechanisms. Thus, it can be quite difficult in practice to ensure incentive alignment in a CBC, or to determine in the absence of strong incentives whether subjects are responding truthfully. Fortunately, there is also considerable evidence that while it is important to get subjects to pay attention and answer carefully, they are mostly honest in their responses irrespective of the incentives offered or how well they are understood; see Bohm (1972), Bohm et al. (1997), Camerer and Hogarth (1999), Yadav (2007), and Dong et al. (2010). This provides some encouragement for applications where it is impractical to provide effective incentives. However, the argument for a simple link between stated choices or referendum votes and consumer welfare is particularly weak in the absence of incentives to be truthful.

Incentive compatibility has been a particular issue for CVM. While it is not difficult in principle to design incentive-compatible elicitation mechanisms, the biggest problem is to get subjects to understand and accept the offered incentives. This is particularly problematic for elicitations of values for large-scale environmental goods, such as protection of endangered species, where subjects are likely to be justifiably skeptical that the environmental change as stated would actually be delivered at the stated cost, or that their response has a non-negligible chance of being pivotal.

### **Reconciliation and Validation**

An advantage of CBC experimental designs is that through the presentation of a slate of menus, there is an opportunity to test the consistency of individual stated choices with neoclassical preference theory, to confront respondents and ask them to explain and reconcile stated choices, and to incorporate menus that allow direct cross-validation between stated and revealed market choices. For example, menus can be offered that allow for

the possibility of testing whether stated choices are consistent with the axioms of revealed preference, and specifically whether they violate the transitivity property of preferences. Even under the more relaxed standard that consumers have stochastic preferences with new preference draws for each choice, their responses can be tested for the regularity property that adding alternatives cannot increase the probability that a previous alternative is chosen. If menus contain current market alternatives, and past purchase behavior of the subjects is known, then one can test whether revealed and stated preferences for the same alternatives are consistent in their weighting of attributes. For example, Morikawa et al. (2002) find that there are systematic differences in weights given to attributes between stated and revealed choices, and that predictions from stated choices can be sharpened by calibrating them to revealed preferences; see also Ben-Akiva and Morikawa (1990), Hensher and Bradley (1993), and Brownstone et al. (2000). This step of testing and validating CBC is important particularly in studies where verisimilitude of the conjoint menus and congruity of the cognitive tools respondents use in experimental and real situations are in question, for example when the products being studied are complex and unfamiliar, such as choices of college, house to purchase, cancer treatment to pursue, or remedies for environmental damages. A large literature compares and tests stated preference elicitation methods, and is relevant to questions of CBC reliability: see Huber (1975, 1987), Rao (1977, 2014), Carmone et al. (1978), Hauser and Koppelman (1979), Hauser and Urban (1979), Jain et al. (1979), Acito and Jain (1980), Neslin (1981), Segal (1982), Akaah and Korgaonkar (1983), Bateson et al. (1987), Train et al. (1987), Louviere (1988), Reibstein et al. (1988), Srinivasan (1988), Huber et al. (1993), McFadden (1994), Huber and Zwerina (1996), Orme (1999), Huber and Train (2001), Hauser and Rao (2002), Raghavarao et al. (2010), and Miller et al. (2011).

In marketing applications, it is possible to validate CBC forecasts against actual market performance of new or modified products, judged by market shares in the population or in population segments. I have found only selective surveys of the performance of forecasts from CBC studies. Natter and Feurstein (2002) compare consumer-level CBC forecasts with scanner data on actual purchases, and conclude that accounting for individual heterogeneity leads to market-level forecasts no better than aggregate models. However, they do not use a statistical method that accounts for the unreliable estimation of individual preference weights. Moore (2004) compares CBC with other elicitation and forecasting methods, and concludes that CBC data analyzed using the methods described in this chapter outperformed the alternatives. Wittick and Bergestuen (2001) cite studies in which CBC forecasts of market shares of data terminals,

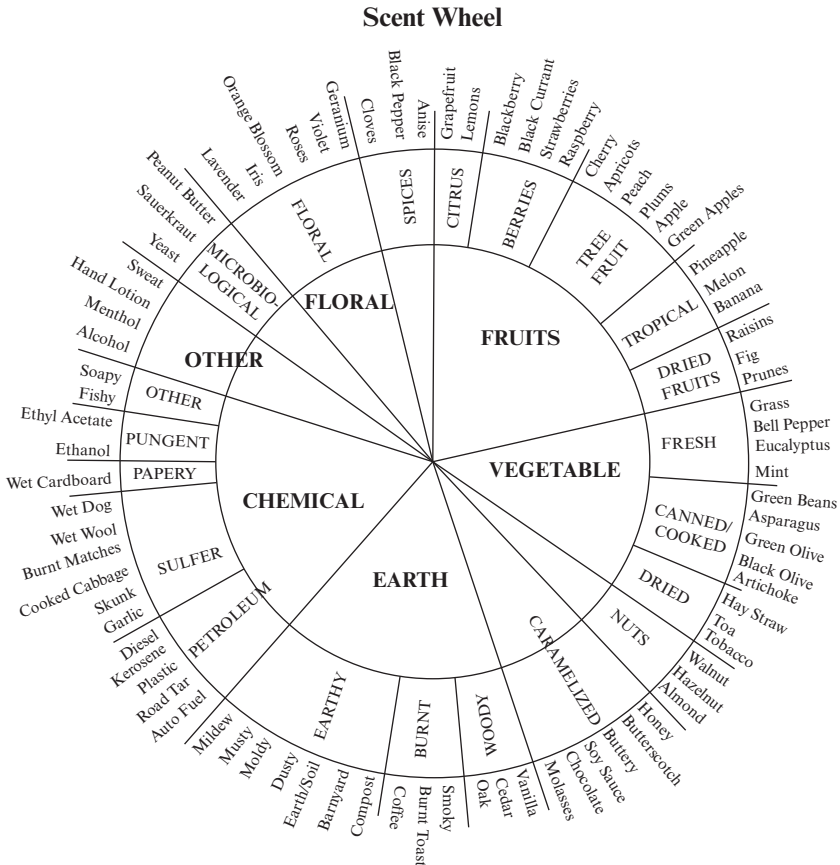
commuter modes, state lottery products, personal computers, and fork-lift trucks are close to actual results, and conclude that “[t]hese results provide strong support for the validity of self-explicated conjoint models in predicting marketplace choice behavior.” At a disaggregate level, Wittink and Montgomery (1979) use CBC data to estimate individual weights on eight attributes of jobs open to MBA graduates, and four months later observe actual job choices and the actual attributes of jobs offered. They report 63% accuracy (percentage of hits) in predicting the jobs students chose out of those offered, compared to a 26% expected hit rate if the students had chosen randomly. They attribute the failure to achieve higher accuracy to noise in estimates of weights for individuals, and to the influence of job attributes not included in the CBC study. They conclude: “On balance, the published results of forecast accuracy are very supportive of the value of conjoint results.” They do caution that “[o]ne should keep in mind that positive results (conjoint analysis providing accurate forecasts) are favored over negative results for publication. Nevertheless, the evidence suggests that marketplace forecasts have validity.”

Based on my own experience with CBC and review of the marketing literature, I conclude that CBC-based market demand forecasts for the population or for large consumer segments are reasonably predictive for familiar products, or products whose attributes are easily extrapolated from past experience. Further, I conclude that uneven performance in some studies is likely due to failure to follow the design principles given in this section, and failures at the stages of modeling, estimation, and forecasting, rather than any intrinsic inability of subjects to state preferences accurately. On the other hand, the evidence is that when respondents have not had market experience with products similar to the ones being studied, or the products in the study have attributes or attribute levels that are unfamiliar, CBC responses are often hypersensitive to framing and context effects, making it very difficult to design elicitations that will give reliable market demand forecasts.

### **Making CBC Menus Realistic**

The CBC design discussion has emphasized product familiarity and menu realism as key ingredients in successful demand forecasting studies. To explore these problems at a concrete level, consider the problem of estimating consumer preferences for red table wines in order to guide blending and pricing decisions. Suppose that in preliminary focus groups, it is found that in addition to *price* per 750ml bottle the attributes that consumers mention as important are *appearance* (brilliant, clear, hazy, cloudy), *bouquet* (outstandingly complex and balanced, distinguished, pleasant,

flat, offensive), *aroma/scent* (Figure 1), *taste and texture* (smooth and full-bodied, good balance with some imperfections, undistinguished, noticeable off-flavors, objectionable flavors), *aftertaste* (outstanding, pleasant, undistinguished, unpleasant), *dryness* (crisp, sensation of residual sugar, sweet), *acidity* (sour, bright, soft), *alcohol content* (percent by volume), *Wine Spectator* magazine rating (60–100), and whether the grapes used to make the wine come from an *organic* vineyard. In a typical CBC study, subjects intercepted in a supermarket would be asked to make choices from eight menus, with each menu containing a no-purchase alternative and



Source: Accessed December 5, 2016 at [http://www.nathankramer.com/wine/wine\\_lists/smelling.htm](http://www.nathankramer.com/wine/wine_lists/smelling.htm).

Figure 1 Aromas/scent attributes of wine

three alternative bottles of wine, described in pictures and in words giving price and attributes, and/or offered in tastes. To motivate participation and for incentive alignment, subjects might be instructed that at the end of the experiment, they will receive one of their eight menu choices plus a Visa cash card for the balance after paying for the product they choose (or the full balance if they chose the no-purchase option on this menu).

A first question in the design of the CBC experiment is how to describe attributes, and set price and attribute levels. Three (sometimes conflicting) criteria are realism, inclusion of existing products (to aid physical tasting, fulfillment of the incentive scheme, and calibration to market data), and sufficient independent variation to allow estimation of the distribution of consumer trade-offs across attributes. Realism requires that prices be in the general range of the subject's experience, and that menus exclude obviously dominated products; for example, the same wine at two different prices. A second question is how to map attributes consumers can understand and relate to, or sensations that consumers experience and consider in making choices, into technical attributes of wines that can be controlled in the production process, such as pH (unbuffered acidity), total acidity (g/L), volatile acidity (g/L), alcohol content (volume percentage), residual sugar (Brix), levels (mg/L) of compounds that influence aroma and taste (e.g., total monomers, total tannins, malic and lactic acid balance, pigmented polymers, catechin, sulfates, free sulfur), and levels of undesirable compounds (ethyl acetate, 2,4,6-trichloroanisole). Most consumers will not be familiar with these technical attributes, and would be unable to incorporate them consistently into stated tastes for different products. One solution is to train subjects to evaluate the products on the basis of these dimensions. For example, with a great deal of wine consumption, subjects may learn to map wine tastes into the scent wheel in Figure 1, and conversely to anticipate accurately the taste of a wine characterized by adjectives in this figure. A panel of experts may be able to map attribute levels that consumers have learned to consistently evaluate and report into technical attributes controlled in the process of making the product; thus, the mapping between chemical scents reported by consumers and flaws in the production process. However, these complex steps may still fail to forecast correctly consumer choices in real market settings in response to the selective information these markets provide, and it may be beyond the capacities of CBC survey design to describe attributes such as the scents in Figure 1 in menus that subjects can understand.

The incentive alignment described above, offering a chosen alternative from one of the menus each subject faces, will be feasible only if one of the stated choices is an existing product. Otherwise, it may not be possible to fulfill a promised transaction. Incentives can be aligned through more

general promises to deliver an available product (or cash) consistent with the subject's stated preferences, but it is likely to be more difficult for the subject to recognize that this makes it in their interest to be truthful.

Whether subjects view the offered prices as reasonable, or a purchase as tempting, will depend on their unknown shopping histories and wine inventories, their expectations regarding availability and prices of wine elsewhere, particularly awareness of and response to real market promotions and sales, and their anticipations of how they will feel if they receive wine from the experiment. In particular, it is important how a "no purchase" option is framed and interpreted. Without prompts, subjects will probably think of the offered menus in the context of their past consumption and current stock of wine at home, and of their options for purchasing wine when they next go to the market. These factors also enter real market purchase decisions, so this context may be realistic, but without measurement or control it is risky to assume that the real environment of the CBC experiment will also prevail in future real markets. Alternately, subjects could be prompted to "think only in terms of what combination of money and wine you leave the experiment with today," but clearly this prompt may elicit different behavior than a prompt such as "Suppose you are on your way to get-together with friends, wine is one of the possible things you could bring, and the menus you see in this experiment are your only opportunity to buy wine" or a prompt such as "You can always take your cash card and buy wine from the regular supermarket shelves if that is more appealing to you than the wines available through this experiment." Some benchmarking to real market penetration rates for wine purchases is likely to be needed to correct for distortion in wine-buying propensity induced by the experience of participating in the experiment and the prompts it contains. Overall, the usefulness of having the "no purchase" option will depend on its being given a sufficiently specific description and context so that it corresponds realistically to the options and attention the consumer will have in the forecast market.

### **Data Analysis**

The CBC elicitation format produces data on choices from hypothetical market experiments that must then be analyzed to model preferences. Simulations from these preferences can then be used to predict market demands for products in the future with different attribute profiles or prices, and if stated consumer behavior is judged to be consistent with (random) utility maximization, used to measure the impact on consumer welfare of changes in product prices and attributes. In marketing, the most widely used model for this analysis is a mixed (or random coefficients)



multinomial logit form, with hierarchical Bayes estimation; see McFadden (1986), Allenby and Rossi (1999, 2006), McFadden and Train (2000), Train and Winston (2007), and Ben-Akiva et al. (2016). The population preference heterogeneity in this model seems to be necessary to reliably predict demands. Further, a number of details seem to be important. It is important to allow correlation in random parameters attached to different product attributes; see Haaijer et al. (1998). It is often important to allow the possibility of market segmentation in which some attributes or products are of no value for some segments of the population. For stable and reliable estimates of WTP for various product attributes, it is useful to model consumer utility in what is termed “money metric” or “WTP” form; see Train and Weeks (2005) and Ben-Akiva et al. (2016). Finally, it is important to be careful in translating changes in attributes of products into changes in consumer well-being when tastes are heterogeneous so that selection becomes an issue, and to define welfare changes consistent with policy alternatives considered and transfers made prior to choice. Hierarchical Bayes estimation methods combined with realistically flexible distributions of model parameters across people then allow simulation-based prediction of market demand. It is then relatively straightforward to infer the impact on utility, or consumer surplus, from adding or altering a product in the market; McFadden (2016) discusses the details of the welfare calculus. The most serious defect in this program comes when consumer choice behavior is so inconsistent and context dependent that the logical connection between consumer choice and well-being breaks down. This is a problem for all of welfare economics predicated on the neoclassical assumptions of maximization of predetermined preferences, but it is particularly acute in SP studies where market discipline of deviations from self-interest breaks down.

### **CBC Failures**

There are a number of things that can go wrong with a CBC experiment and render it unreliable, even when the “necessary” experimental conditions described in this section are met. Mostly, these come from inconsistencies in consumer choice behavior, and in failures of consumers to attend to or understand the task, the alternatives, or the offered incentives. For example, McFadden et al. (1988) find a strong “status quo bias” in a CBC study of electricity reliability choices – there are consistent trade-offs between reliability and cost for alternatives other than the subject’s status quo, but a strong distaste for moving away from the status quo, no matter what its initial level. McFadden (1994) finds “extension bias” in CBC responses, a phenomenon related to preferences over alternatives of

different scope – subjects discount the size dimension of product comparisons. Green et al. (1998) find strong “anchoring biases” that are similar in both estimation and valuation tasks, suggesting that subjects think about and estimate uncertain facts and uncertain preferences in similar ways. Morikawa et al. (2002) merge stated and revealed preference data, and find evidence that subjects make *similar but not identical trade-offs* in the two circumstances. Ariely (2009) and McFadden (1998, 1999, 2014a, 2014b) give a broader list of cognitive effects on consumer behavior. Researchers collecting stated preference data need to be keenly aware of how cognitive effects can influence subject’s responses, develop elicitation methods that minimize the distortions in apparent preferences these effects can produce, and find testing and calibration methods that detect and correct for the presence of these effects.

There is a qualitative difference between the impact of inconsistencies in stated or revealed consumer choice behavior on the reliability of demand forecasts and on the reliability of measures of consumer well-being. Market demand forecasting has the relatively robust feature that so long as the distribution of decision-making rules in the population is stationary and a CBC experiment is realistic, mimicking well the information, manipulations, incentives, and social context of choices in real markets, then models fitted to the CBC data will usually forecast successfully even if consumers systematically deviate from neoclassical utility maximization so that the forecasting model predicated on utility maximization is only an approximation. However, the tight neoclassical link between choice behavior and consumer welfare that holds under utility maximization, allowing changes in well-being to be inferred from consumer surplus calculations, breaks down when choices are inconsistent with utility maximization. In this case, there is no foundation for a supposition that neoclassical consumer surplus measures are reliable indicators of well-being.

The question remains whether stated WTP or well-being might provide satisfactory measures of consumer welfare even if choice behavior is inconsistent with neoclassical utility maximization. The answer in short is that there are currently no accepted scientific principles that support such an inference. In particular, the economic theory of markets, and the methods of neoclassical economics, provide no support for direct elicitation of well-being. While it is possible that psychological scales of well-being, such as those promoted by Kahneman and Krueger (2013), may develop to the point where they can pass reasonable tests for reliability and plausibility, their current implementations are far too sensitive to context and framing to be useful now; see Deaton (2012). It is also possible that in the future the tools of neuroeconomics can provide a physiological basis for measures of well-being that can be tied to stated choices and WTP. Khaw et al. (2015) find that

elicitations of values for ordinary market goods and for use values of environmental goods provoke similar brain activity in MRI scans, but elicitations for non-use values are processed differently. However, this is an early paper, and a full investigation will require substantial progress in brain science. Pending such developments, there is no scientific basis for a claim that in situations where choice behavior cannot pass tests for consistency with utility maximization, stated WTP for changes in attributes of alternatives is reliable.

## LESSONS FOR VALUATION OF ENVIRONMENTAL PUBLIC GOODS

Applications of SP methods to environmental use values raise similar questions to applications of CBC to conventional products and services. Carson (2012), Hausman (2012), and Kling et al. (2012) discuss the particular challenges of using CVM for natural resource valuation. To the extent that consumers are unfamiliar with market transactions for environmental goods, the challenges are similar to those of forecasting demand for unfamiliar consumer products. Extensive training is likely to be necessary to get subjects to think of their CBC offerings in the same way they do personal market purchases, and the framing and context provided by this training itself can manipulate stated preferences. Issues are (1) that preferences may not be well formed, and consequently may be particularly susceptible to manipulation coming from the framing of attributes and ranges of attribute levels, and order and emphasis that influence prominence of different attributes, (2) that personal preferences are sensitive to social judgments that are difficult to frame and control in an experiment, and (3) subjects may not be persuaded that the incentive alignment they are offered is real. For these reasons, CBC or CVM elicitation of WTP for use of environmental goods face major hurdles to the achievement of consistency and reproducibility.

For valuation of lost use, such as lost opportunities for anglers when streams are closed due to hazardous waste, the services involved may be sufficiently familiar, and the elicitations may be sufficiently similar to ordinary market opportunities, so that the CBC methods may prove reliable. However, all the cautions that apply to conventional marketing applications of CBC also apply here, and these environmental good applications will often be low on the reliability gradient, involving public good and social aspects that can interfere with individual utility assessments, and requiring unfamiliar stated purchase choices. At best, CBC elicitations of use values are likely to require extensive validation through merger of stated and revealed preference data. For example, Myer et al. (2010) summarize findings from CVM studies of

the recreational use value of birdwatching, and show a great range of valuations. Some of this variation is due to differences in bird populations and birdwatching opportunities, but the greatest part of the variation seems to come from apparently modest differences in CVM study design and analysis methods. A question for future research is whether moving from CVM and econometric methods appropriate for such data (e.g., Cameron and James, 1987; Cameron, 1988; Cameron and Quiggin, 1994) to CBC designs and hierarchical Bayes methods would reduce the variations in stated use values for environmental services like birdwatching.

The greatest need for stated preference data is in application to environmental non-use values, but here all the circumstances that make market research-oriented CBC unreliable are reinforced. My judgment is that to this point no one has been able to develop and demonstrate stated preference methods that are reliable for valuation of non-use claims. For example, consider the demonstrated ability of market researchers to nail down predictions of demand for products such as smartphones with specific attributes and prices. I am not aware of parallel success stories for environmental non-use values. Seemingly innocuous and inconsequential changes in stated preference study designs that should have little influence on neoclassical consumers with well-formed preferences induce major changes in valuations, and there are no good yardsticks for either "correct" study design or results. The combination of consumers who are hypersensitive to context and susceptible to manipulation, and SP analysts who lack market benchmarks to constrain their experimental designs, is toxic for reliability. As noted in the last section for marketing applications, and with more emphasis here for non-use values, I believe that a great deal of scientific work, and some major breakthroughs, will be needed before stated WTP or well-being measures have a sufficient foundation in neuroscience so that they can be measured reliably without close links to revealed market behavior. In my judgment, environmental economists should be more scientifically cautious in weighing the evidence for and against the reliability of stated preference methods, and more prudent in the claims they make for these methods in environmental non-use valuations.

## REFERENCES

- Acito, F. and A. Jain (1980), "Evaluation of conjoint analysis results: A comparison of methods," *Journal of Marketing Research*, **17**(1), 106–12.
- Akaah, I. and P. Korgaonkar (1983), "An empirical comparison of the predictive validity of self-explicated, Huber-hybrid, traditional conjoint, and hybrid conjoint models," *Journal of Marketing Research*, **20**(2), 187–97.

- Allenby, G. and P. Rossi (1999), "Marketing models of consumer heterogeneity," *Journal of Econometrics*, **89**(1–2), 57–78.
- Allenby, G. and P. Rossi (2006), "Hierarchical Bayes models: A practitioner's guide," in R. Grover and M. Vriens (eds), *The Handbook of Marketing Research*, Thousand Oaks, CA: Sage Publications.
- Ariely, D. (2009), *Predictably Irrational*, New York: Harper-Collins.
- Bateson, J., D. Reibstein, and W. Boulding (1987), "Conjoint analysis reliability and validity: A framework for future research," in M. Houston (ed.), *Review of Marketing*, Chicago, IL: American Marketing Association, pp. 451–81.
- Becker, G., M. DeGroot, and J. Marschak (1964), "Measuring utility by a single-response sequential method," *Behavioral Science*, **9**(3), 226–32.
- Beggs, S., S. Cardell, and J. Hausman (1981), "Assessing the potential demand for electric cars," *Journal of Econometrics*, **17**(1), 1–19.
- Ben-Akiva, M., D. McFadden, and K. Train (2016), "Foundations of stated preference elicitation," *UC Berkeley Working Paper*, accessed December 4, 2016 at <http://eml.berkeley.edu/~train/foundations.pdf>.
- Ben-Akiva, M. and T. Morikawa (1990), "Estimation of switching models from revealed preferences and stated intentions," *Transportation Research A*, **24**(6), 485–95.
- Bishop, R. and T. Heberlein (1979), "Measuring values of extramarket goods: Are indirect measures biased?," *American Journal of Agricultural Economics*, **61**(5), 926–30.
- Bishop, R. and T. Heberlein (1986), "Does contingent valuation work?," in R.G. Cummings, D.S. Brookshire, R.C. Bishop, and K.J. Arrow (eds), *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, Totowa, NJ: Rowman & Allanheld, pp. 123–47.
- Bliemer, M. and J. Rose (2009), "Designing stated choice experiments: The state of the art," in R. Kitamura, T. Yoshi, and T. Yamamoto (eds), *The Expanding Sphere of Travel Behaviour Research*, Kyoto: International Association for Travel Behaviour Research, pp. 499–538.
- Bohm, P. (1972), "Estimating the demand for public goods: An experiment," *European Economic Review*, **3**(2), 111–30.
- Bohm, P., J. Linden, and J. Sonnegard (1997), "Eliciting reservation prices: Becker-DeGroot-Marschak mechanisms vs. markets," *The Economic Journal*, **107**(443), 1079–89.
- Brookshire, D., B. Ives, and W. Schultze (1976), "The valuation of aesthetic preferences," *Journal of Environmental Economics and Management*, **3**(4), 325–46.
- Brookshire, D., A. Randall, and J. Stoll (1980), "Valuing increments and decrements in natural resource service flows," *American Journal of Agricultural Economics*, **62**(3), 478–88.
- Brownstone, D. and K. Train (1999), "Forecasting new product penetration with flexible substitution patterns," *Journal of Econometrics*, **89**(1–2), 109–29.
- Brownstone, D., D. Bunch, and K. Train (2000), "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles," *Transportation Research B*, **34**, 315–38.
- Camerer, C. and R. Hogarth (1999), "The effects of financial incentives in experiments: A review and capital-labor-production framework," *Journal of Risk and Uncertainty*, **19**(1–3), 7–42.
- Camerer, C., G. Lowenstein, and M. Rabin (2004), *Advances in Behavioral Economics*, Princeton, NJ: Princeton University Press.

- Cameron, L., M. Cragg, and D. McFadden (2013), "Determining reasonable royalty damages: the role of conjoint surveys," *Law360*, October 16, accessed December 4, 2016 at <http://www.law360.com/articles/475390/the-role-of-conjoint-surveys-in-reasonable-royalty-cases>.
- Cameron, T. (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**(3), 355–79.
- Cameron, T. and M. James (1987), "Efficient estimation methods for closed-ended contingent valuation survey data," *Review of Economics and Statistics*, **69**(2), 269–76.
- Cameron, T. and J. Quiggin (1994), "Estimation using contingent valuation data from a 'dichotomous choice with follow-up' questionnaire," *Journal of Environmental Economics and Management*, **27**(3), 218–34.
- Carmone, F.J., P.E. Green, and A.K. Jain (1978), "Robustness of conjoint analysis: Some Monte Carlo results," *Journal of Marketing Research*, **15**(2), 300–303.
- Caro, F., T. Ho, D. McFadden, A. Gottlieb, C. Yee, and T. Chan et al. (2012), "Using the Internet to administer more realistic vignette experiments," *Social Science Computer Review*, **30**(2), 184–201.
- Carson, R. (2012), "Contingent valuation: A practical alternative when prices aren't available," *Journal of Economic Perspectives*, **26**(4), 27–42.
- Carson, R., N. Flores, and N. Meade (2001), "Contingent valuation: Controversies and evidence," *Environmental and Resource Economics*, **19**(2), 173–210.
- Chapman, R. and R. Staelin (1982), "Exploiting rank ordered choice set data within the stochastic utility model," *Journal of Marketing Research*, **19**(3), 288–301.
- Ciriacy-Wanthurp, S.V. (1947), "Capital returns from soil-conservation practices," *Journal of Farm Economics*, **29**(4), 1181–96.
- Cummings, R., D. Brookshire, and W. Schulze (1986), *Valuing Environmental Goods*, Totowa, NJ: Rowman & Allanheld.
- Davis, R. (1963a), "Recreation planning as an economic problem," *Natural Resources*, **3**, 239–49.
- Davis, R. (1963b), "The value of outdoor recreation: An economic study of the Maine woods," PhD dissertation, Harvard University.
- Deaton, A. (2012), "The financial crisis and the well-being of America," in D. Wise (ed.), *Investigations in the Economics of Aging*, Chicago, IL: University of Chicago Press, pp.343–76.
- Diamond, P.A. and J.A. Hausman (1993), "On contingent valuation measurement of non-use values," in J. Hausman (ed.), *Contingent Valuation: A Critical Assessment*, Amsterdam: North Holland, pp.3–38.
- Dong, S., M. Ding, and J. Huber (2010), "A simple mechanism to incentive-align conjoint experiments," *International Journal of Research in Marketing*, **27**(1), 25–32.
- Elrod, T., J. Louviere, and K. Davey (1992), "An empirical comparison of ratings-based and choice-based conjoint models," *Journal of Marketing Research*, **29**(3), 368–77.
- Fischhoff, B. and C. Manski (1999), *Elicitation of Preferences*, Boston, MA: Kluwer Academic Publishers.
- Freeman, R. (1986), "On assessing the state of the arts of the contingent valuation method of valuing environmental change," in R. Cummings, D. Brookshire,

- and W. Schulze (eds), *Valuing Environmental Goods*, Totowa, NJ: Rowman & Allanheld, pp. 148–61.
- Frisch, R. (1926), “Sur un problème d’économie pur,” *Norsk Matematisk Forenings Skrifter*, **16**, 1–40; translation in J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein (eds), *Preferences, Utility, and Demand*, New York: Harcourt.
- Green, D., K. Jacowitz, D. Kahneman, and D. McFadden (1998), “Referendum contingent valuation, anchoring, and willingness to pay for public goods,” *Resource and Energy Economics*, **20**(2), 85–116.
- Green, P. and V. Rao (1971), “Conjoint measurement for quantifying judgmental data,” *Journal of Marketing Research*, **8**(3), 355–63.
- Green, P. and V. Srinivasan (1978), “Conjoint analysis in consumer research: Issues and outlook,” *Journal of Consumer Research*, **5**(2), 103–23.
- Green, P., D. Carroll, and S. Goldberg (1981), “A general approach to product design optimization via conjoint analysis,” *Journal of Marketing*, **45**(3), 17–37.
- Green, P., A. Krieger, and Y. Wind (2001), “Thirty years of conjoint analysis: Reflections and prospects,” *Interfaces*, **31**(Supplement), S56–S73.
- Groves, T. and M. Loeb (1975), “Incentives and public inputs,” *Journal of Public Economics*, **4**(3), 211–26.
- Haaiker, R., M. Wedel, M. Vriens, and T. Wansbeek (1998), “Utility covariances and context effects in conjoint MNP models,” *Marketing Science*, **17**(3), 236–52.
- Hammack, J. and G. Brown (1974), *Waterfowl and Wetlands: Toward Bioeconomic Analysis*, Baltimore, MD: Johns Hopkins Press.
- Hanemann, M., J. Loomis, and B. Kanninen (1991), “Statistical efficiency of double-bounded dichotomous choice contingent valuation,” *American Journal of Agricultural Economics*, **73**(4), 1255–63.
- Hauser, J. and F. Koppelman (1979), “Alternative perceptual mapping techniques: Relative accuracy and usefulness,” *Journal of Marketing Research*, **16**(4), 495–506.
- Hauser, J. and G. Urban (1979), “Assessment of attribute importances and consumer utility functions: Von Neumann-Morgenstern theory applied to consumer behavior,” *Journal of Consumer Research*, **5**(4), 251–62.
- Hauser, J. and V. Rao (2002), “Conjoint analysis, related modelling, and applications,” in Y. Wind and P.E. Green (eds), *Market Research and Modeling: Progress and Prospects*, New York: Springer, pp. 141–68.
- Hausman, J. (ed.) (1993), *Contingent Valuation: A Critical Assessment*, Amsterdam: North-Holland.
- Hausman, J. (2012), “Contingent valuation: From dubious to hopeless,” *Journal of Economic Perspectives*, **26**(4), 43–56.
- Hensher, D. and M. Bradley (1993), “Using stated response data to enrich revealed preference discrete choice models,” *Marketing Letters*, **4**(2), 139–52.
- Hensher, D. and J. Louviere (1983), “Identifying individual preferences for international air travel: An application of functional measurement theory,” *Journal of Transport Economics and Policy*, **17**(3), 225–45.
- Hensher, D., J. Louviere, and J. Swait (1999), “Combining sources of preference data,” *Journal of Econometrics*, **89**(1–2), 197–221.
- Hicks, J. (1939), *Value and Capital*, Oxford: Clarendon Press.
- Huber, J. (1975), “Predicting preferences on experimental bundles of attributes: A comparison of models,” *Journal of Marketing Research*, **12**, 290–97.
- Huber, J. (1987), “Conjoint analysis: How we got here and where we are,” in

- Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, pp.237–52.
- Huber, J. and K. Train (2001), “On the similarity of classical and Bayesian estimates of individual mean partworths,” *Marketing Letters*, **12**(3), 259–69.
- Huber, J. and K. Zwerina (1996), “The importance of utility balance in efficient choice designs,” *Journal of Marketing Research*, **33**(3), 307–17.
- Huber, J., D. Wittink, J. Fiedler, and R. Miller (1993), “The effectiveness of alternative preference elicitation procedures in predicting choice,” *Journal of Marketing Research*, **30**(1), 105–14.
- Hurd, M. and D. McFadden (1998), “Consumption and savings balances of the elderly: Experimental evidence on survey response bias,” in D. Wise (ed.), *Frontiers in the Economics of Aging*, Chicago, IL: University of Chicago Press, pp. 353–87.
- Jain, A., F. Acito, N. Malhotra, and V. Mahajan (1979), “A comparison of the internal validity of alternative parameter estimation methods in decompositional multiattribute preference models,” *Journal of Marketing Research*, **16**(3), 313–22.
- Johnson, R. (1974), “Trade-off analysis of consumer values,” *Journal of Marketing Research*, **11**(2), 121–7.
- Johnson, R. (1999), “The joys and sorrows of implementing HB methods for conjoint analysis,” *Working Paper*, Sequim, WA, Sawtooth Software.
- Johnson, R. and B. Orme (1996), “How many questions should you ask in choice-based conjoint studies?,” *Research Paper Series*, Sequim, WA: Sawtooth Software, Inc.
- Kahneman, D. and A. Krueger (2013), *Developments in the Measurement of Subjective Well-Being*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Kahneman, D. and A. Tversky (eds) (2000), *Choices, Values, and Frames*, New York: Sage.
- Kahneman, D., E. Diener, and N. Schwarz (eds) (1999), *Well-Being: The Foundations of Hedonic Psychology*, New York: Sage.
- Khaw, M., D. Grab, M. Livermore, C. Vossler, and P. Glimcher (2015), “The measurement of subjective value and its relation to contingent valuation and environmental public goods,” *PLOS ONE*, **10**(7).
- King, G., C. Murray, J. Salomon, and A. Tandon (2004), “Enhancing the validity and cross-cultural comparability of measurement in survey research,” *American Political Science Review*, **98**(1), 191–207.
- Kling, C., D. Phaneuf, and J. Zhao (2012), “From Exxon to BP: Has some number become better than no number?,” *Journal of Economic Perspectives*, **26**(4), 3–26.
- Kuhfield, W., R. Tobias, and M. Garratt (1994), “Efficient experimental design with marketing research applications,” *Journal of Marketing Research*, **31**(4), 545–57.
- List, J. and C. Gallet (2001), “What experimental protocol influences disparities between actual and hypothetical stated values?,” *Environmental and Resource Economics*, **20**(3), 241–54.
- Louviere, J. (1988), *Analyzing Decision Making: Metric Conjoint Analysis*, Newbury Park, CA: Sage.
- Louviere, J.J. and G.G. Woodworth (1983), “Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data,” *Journal of Marketing Research*, **20**(4), 350–67.



- Louviere, J., D. Hensher and J. Swait (2000), *Stated Choice Methods: Analytics and Applications*, New York: Cambridge University Press.
- Luce, D. and P. Suppes (1965), 'Preferences, utility and subjective probability', in R. Luce, R. Bush, and E. Galanter (eds), *Handbook of Mathematical Psychology*, New York: John Wiley & Sons, pp. 249–410.
- Luce, D. and J. Tukey (1964), "Simultaneous conjoint measurement: A new type of fundamental measurement," *Journal of Mathematical Psychology*, **1**(1), 1–27.
- MacCrimmon, K. and M. Toda (1969), "The empirical determination of indifference curves," *Review of Economic Studies*, **36**(4), 433–51.
- McFadden, D. (1981), "Econometric models of probabilistic choice," in C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data*, Cambridge, MA: MIT Press, pp. 198–272.
- McFadden, D. (1986), "The choice theory approach to market research," *Marketing Science*, **5**(4), 275–97.
- McFadden, D. (1994), "Contingent valuation and social choice," *American Journal of Agricultural Economics*, **76**(4), 689–708.
- McFadden, D. (1998), "Measuring willingness-to-pay for transportation improvements," in T. Garling, T. Laitila, and K. Westin (eds), *Theoretical Foundations of Travel Choice Modeling*, New York: Elsevier.
- McFadden, D. (1999), "Rationality for economists?" in B. Fischhoff and C. Manski (eds), *Elicitation of Preferences*, Boston, MA: Kluwer, pp. 73–106.
- McFadden, D. (2012), "Economic juries and public project provision," *Journal of Econometrics*, **166**(1), 116–26.
- McFadden, D. (2014a), "The new science of pleasure: Consumer behavior and the measurement of well-being," in S. Hess and A. Daly (eds), *Handbook of Choice Modelling*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing, pp. 7–48.
- McFadden, D. (2014b), "An economist's perspective on environmental damages," paper at The Eighth Annual Advanced Conference on Litigating Natural Resource Damages, Law Seminars International.
- McFadden, D. (2016), "Welfare economics with partially-observed consumers," *UC Berkeley Working Paper*.
- McFadden, D. and K. Train (2000), "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, **15**(5), 447–70.
- McFadden, D., A. Goett, and C.-K. Woo (1988), "Estimating household value of electric service reliability with market research data," *Energy Journal: Special Electricity Reliability Issue*, **9**, 105–20.
- McFadden, D., M. Ben-Akiva, A. Goett, and D. Bolduc (1986), *The Choice Theory Approach to Analyze the Preferences and Choices of Electric Utility Consumers*, EPRI Final Report RP2671-1.
- Miller, K., R. Hofstetter, H. Krohmer, and Z.J. Zhang (2011), "How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches," *Journal of Marketing Research*, **48**(1), 172–84.
- Mitchell, R. and R. Carson (1989), *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Washington, DC: Resources for the Future.
- Moore, W. (2004), "A cross-validity comparison of rating-based and choice-based conjoint analysis models," *International Journal of Research in Marketing*, **21**(3), 299–312.
- Morikawa, T., M. Ben-Akiva, and D. McFadden (2002), "Discrete choice models incorporating revealed preferences and psychometric data," in P. Franses and

- A. Montgomery (eds), *Econometric Models in Marketing*, Amsterdam, London and New York: Elsevier Science, pp. 29–55.
- Moscati, I. (2007), “Early experiments in consumer demand theory: 1930–1970,” *History of Political Economy*, **39**(3), 359–401.
- Myer, K., G. Parsons, and P. Edwards (2010), “Measuring the recreational use value of migratory shorebirds on the Delaware Bay,” *Marine Resources Economics*, **25**(3), 247–64.
- Natter, M. and M. Feurstein (2002), “Real world performance of choice-based conjoint models,” *European Journal of Operational Research*, **117**, 448–58.
- Neslin, S. (1981), “Linking product features to perceptions: Self-stated versus statistically revealed importance weights,” *Journal of Marketing Research*, **18**(1), 80–86.
- Orme, B. (1999), “ACA, CBC, or both? Effective strategies for conjoint research,” *Working Paper*, Sequim, WA: Sawtooth Software.
- Raghavarao, D., J. Wiley, and P. Chitturi (2010), *Choice-Based Conjoint Analysis: Models and Designs*, Boca Raton, FL: CRC Press.
- Randall, A. (1986), “The possibility of satisfactory benefit estimates with contingent markets,” in R. Cummings, D. Brookshire, and W. Schulze (eds), *Valuing Environmental Goods*, Totowa, NJ: Rowman & Allanheld, pp. 114–22.
- Randall, A., B.C. Ives, and E. Eastman (1974), “Bidding games for valuation of aesthetic environmental improvements,” *Journal of Environmental Economics and Management*, **1**(2), 132–49.
- Rao, V. (1977), “Conjoint measurement in marketing analysis,” in J. Sheth (ed.), *Multivariate Methods for Market and Survey Research*, Chicago, IL: American Marketing Association, pp. 257–86.
- Rao, V. (2014), *Applied Conjoint Analysis*, Berlin: Springer.
- Reibstein, D., J. Bateson, and W. Boulding (1988), “Conjoint analysis reliability: Empirical findings,” *Marketing Science*, **7**(3), 271–86.
- Rose, J. and M. Bleimer (2009), “Constructing efficient stated choice experimental designs,” *Transport Reviews*, **29**(5), 587–617.
- Rossi, P. (1979), “Vignette analysis: Uncovering the normative structure of complex judgments,” in R.K. Merton, J.S. Coleman, and P.H. Rossi (eds), *Qualitative and Quantitative Social Research: Papers in Honor of Paul Lazarsfeld*, New York: Macmillan, pp. 175–88.
- Rossi, R., G. Allenby, and R. McCulloch (2005), *Bayesian Statistics and Marketing*, New York: John Wiley & Sons Ltd.
- Samuelson, P. (1947 [1983]), *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.
- Schkade, D. and J. Payne (1993), “Where do the numbers come from? How people respond to contingent valuation questions,” in J. Hausman (ed.), *Contingent Valuation: A Critical Assessment*, Amsterdam: North-Holland, pp. 271–93.
- Schultz, H. (1925), “The statistical law of demand as illustrated by the demand for sugar,” *Journal of Political Economy*, **33**(5), 481–504 and 577–637.
- Schultz, H. (1928), *Statistical Laws of Demand and Supply with Special Application to Sugar*, Chicago, IL: University of Chicago Press.
- Segal, M. (1982), “Reliability of conjoint analysis: Contrasting data collection procedures,” *Journal of Marketing Research*, **19**, 139–43.
- Shocker, A. and V. Srinivasan (1974), “A consumer-based methodology for the identification of new product ideas,” *Management Science*, **20**(6), 921–37.
- Smith, K., W. Desvousges, and A. Fisher (1986), “A comparison of direct and

- indirect methods for estimating environmental benefits,” *American Journal of Agricultural Economics*, **68**(2), 280–90.
- Srinivasan, V. (1988), “A conjunctive-compensatory approach to the self-explication of multiattributed preferences,” *Decision Sciences*, **19**(2), 295–305.
- Thurstone, L.L. (1931), “The indifference function,” *Journal of Social Psychology*, **2**(2), 139–67.
- Train, K. and C. Winston (2007), “Vehicle choice behavior and the declining market share of U.S. automakers,” *International Economic Review*, **48**(4), 1469–98.
- Train, K., D. McFadden, and A. Goett (1987), “Consumer attitudes and voluntary rate schedules for public utilities,” *Review of Economics and Statistics*, **LXIX**, 383–91.
- 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, pp. 1–16.
- Urban, G., J. Hauser, and J. Roberts (1990), “Prelaunch forecasting of new automobiles: Models and implementation,” *Management Science*, **36**(4), 401–21.
- Urban, G., J. Hauser, W. Qualls, B. Weinberg, J. Bohlmann, and R. Chicos (1997), “Validation and lessons from the field: Applications of information acceleration,” *Journal of Marketing Research*, **34**(1), 143–53.
- Wallis, A. and M. Friedman (1942), “The empirical derivation of indifference functions,” in O. Lange, F. McIntyre, and T.O. Yntema (eds), *Studies in Mathematical Economics and Econometrics*, Chicago, IL: University of Chicago Press, pp. 175–89.
- Wittink, D. and T. Bergestuen (2001), “Forecasting with conjoint analysis,” in J. Armstrong (ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, New York: Springer, pp. 147–67.
- Wittink, D. and D. Montgomery (1979), “Predicting validity of trade-off analysis for alternative segmentation schemes,” paper at the American Marketing Association Educator’s Conference, pp. 69–73.
- Wright, P. and M. Kriewall (1980), “State-of-mind effects on accuracy with which utility functions predict marketplace utility,” *Journal of Marketing Research*, **17**(3), 277–93.
- Yadav, L. (2007), “Hypothetical bias in contingent valuation,” dissertation, University of Massachusetts Amherst.