CHAPTER 14

Contingent Valuation: Using Surveys to Elicit Information About Costs and Benefits

Economists generally are much more comfortable observing individuals' valuations of goods and services through their behavior in markets than eliciting their valuations through survey questionnaires. They prefer to observe purchasing decisions because these decisions directly reveal preferences, whereas surveys elicit statements about preferences. Nevertheless, for some public goods there are simply no, or very poor, market proxies or other means of inferring preferences from observations. In such circumstances, many analysts have concluded that there is no viable alternative to asking a sample of people about their valuations.

Questionnaires designed to elicit preferences are normally referred to as *contingent valuation* (CV) surveys,[1] or sometimes hypothetical valuation surveys, because respondents are not actually required to pay their valuations of goods.[2] The primary use of CV is to elicit people's willingness-to-pay for changes in quantities or qualities of goods. Many kinds of goods, including water quality at recreation sites, goose hunting, sports stadiums, and outdoor recreation, have been valued with CV surveys.[3] as have health outcomes.[4] Such agencies as the National Park Service and the U.S. Bureau of Reclamation use CV surveys to value recreation and wildlife opportunities. CV surveys are also used to value more complex and abstract goods, such as reductions in the volume of hazardous wastes, the value of archeological artifacts and sites, and the preservation of wetland ecosystems.[5] Valuing the use, or potential use, of goods with CV is relatively uncontroversial. Valuing passive use (nonuse) with CV is more controversial, both for the conceptual reasons discussed in Chapter 9 and the practical reasons related to gathering valid information from surveys discussed later in this chapter.

In spite of the controversy, the use of CV as a method for estimating costs and benefits is now widespread. Indeed, the federal courts have held that surveys of citizens' valuations have "rebuttable presumption" status in cases involving the assessment of damage to natural resources.[6] A blue-ribbon panel of social scientists convened by the National Oceanic and Atmospheric Administration (NOAA) further legitimized the use of CV by concluding that it could be the basis for estimating passive use values for inclusion in natural resource damage assessment cases.[7]

In this chapter we provide an overview of CV methods. We assess the major strengths and weakness of those methods most commonly used, emphasize the importance of informing respondents of how they would pay the costs of a project (the payment vehicle), review the major criticisms and problems related to CV use, discuss

some of the survey issues most relevant to CV, and consider CV's accuracy. Finally, we present some heuristic checklists for analysts preparing or reviewing CV instruments.

OVERVIEW OF CONTINGENT VALUATION METHODS

The general approach of all CV methods is as follows. First, a sample of respondents from the population with standing is identified. Second, respondents are asked questions about their valuations of some good. Third, their responses provide information that enables analysts to estimate the respondents' willingness-to-pay (WTP) for the good. Fourth, the WTP amounts for the sample are extrapolated to the entire population. If, for instance, the respondents comprise a random sample of the population such that each member of the population had an equal chance of being in the sample, then the average WTP for the sample would simply be multiplied by the size of the population to arrive at the aggregate WTP.

As CV surveys are expensive to conduct, analysts may wish to extrapolate the results of existing surveys to different populations. However, the characteristics of these populations usually do not match those of the population previously sampled. For example, the populations may differ in terms of income, access to alternative goods, or other factors that may be relevant to their demands for the good. As discussed in Chapter 15, reasonable extrapolation requires that these differences be controlled for statistically.[8] Therefore, analysts increase the chances that their CV surveys will have use beyond their own CBAs by collecting and reporting information about the characteristics of their samples, including WTP amounts for subsets of the sample, even when such information is unnecessary for their own studies.

The remainder of this section introduces particular methods that are used to elicit WTP amounts from survey respondents. We first briefly sketch three methods that have been used at various times: the *open-ended willingness-to-pay* method, the *closed-ended iterative bidding* method, and the *contingent ranking* method. We then turn to the *dichotomous choice,* or *referendum,* method, which was advocated by the NOAA blue-ribbon panel as the method of choice in most circumstances.[9]

Direct Elicitation (Nonreferendum) Methods

Several CV methods ask questions about preferences directly. The open-ended willingness-to-pay method and the closed-ended iterative bidding method seek to elicit the WPT amounts for each respondent. The contingent ranking method seeks to elicit a preference profile over a set of alternatives for each respondent. These methods contrast with the dichotomous choice method, which is indirect in the sense that it relies on patterns of responses across a large number of respondents to make inferences about the preferences of respondents with particular characteristics.

Open-Ended Willingness-to-Pay Method The earliest method to be used is the *open-ended willingness-to-pay* approach. Here respondents are simply asked to state their maximum WTP for the good, or policy, that is being valued.[10] The question might be formulated as follows: "What is the most that you would be prepared to pay in additional federal income taxes to guarantee that the Wildwood wilderness area will remain closed to development?" This method had dropped out of favor as analysts feared unrealistic responses because respondents needed some initial guidance on

valuations.[11] Concerns that open-ended questions result in excessively large estimates of WTP seem unfounded; indeed, a more serious problem appears to be that respondents who have very low valuations of the good being valued often state a zero value.

Closed-Ended Iterative Bidding Method In the *closed-ended iterative bidding* method, respondents are asked whether they would pay a specified amount for the good or policy that has been described. If respondents answer affirmatively, then the amount is incrementally increased. The procedure continues until the respondent expresses unwillingness to pay the amount specified. Similarly, if respondents answer negatively to the initial amount specified, the interviewer lowers the amount by increments until the respondent expresses a WTP.[12]

The initial question for determining WTP typically starts with something like the following: "Now suppose the costs to clean the Kristiansand Fjord were divided on [sic] all taxpayers in the whole of Norway by an extra tax in 1986. If this extra tax was 200 kronor for an average taxpayer, would you then be willing to support the proposal?"[13] In this CV survey the interviewer set the initial price at 200 kronor. If a respondent indicated a willingness-to-pay this initial price, then the interviewer raised the price by 200 kronor and asked the question again. The interviewer kept going until the respondent gave a negative answer. Similarly, if the initial response was negative, then the interviewer dropped the price by 200 kronor increments until the respondent gave a positive response. Although iterative bidding was at one time the most common method in use, it is rarely used now because of considerable evidence that its results are highly sensitive to the initially presented, or starting, value.

Contingent Ranking Method In the *contingent ranking,* or *ranked choice,* method, respondents are asked to rank specific feasible combinations of quantities of the good being valued and monetary payments. For example, respondents choose on a continuum between a low level of water quality at a low tax-price and a high level of water quality at a high tax-price. The combinations are ranked from most preferred to least preferred.[14] The rankings provide a basis for estimating each respondent's WTP for various increments of quality. Contingent ranking implies an ordinal ranking procedure in contrast to the iterative bidding procedure, which requires cardinal evaluation. Typically, tasks that require only ordinal information processing, that is, a ranking rather than a precise specification of value, are considerably easier for respondents to perform.[15] This is a valuable attribute in the CV context where complex information must often be processed by respondents. Of course, unlike the open-ended WTP method or the closed-ended iterative bidding method, the WTP of interest must be inferred from ordinal rankings rather than directly elicited. Additionally, responses appear to be sensitive to the order in which alternatives are presented to respondents.

Dichotomous Choice (Referendum) Method

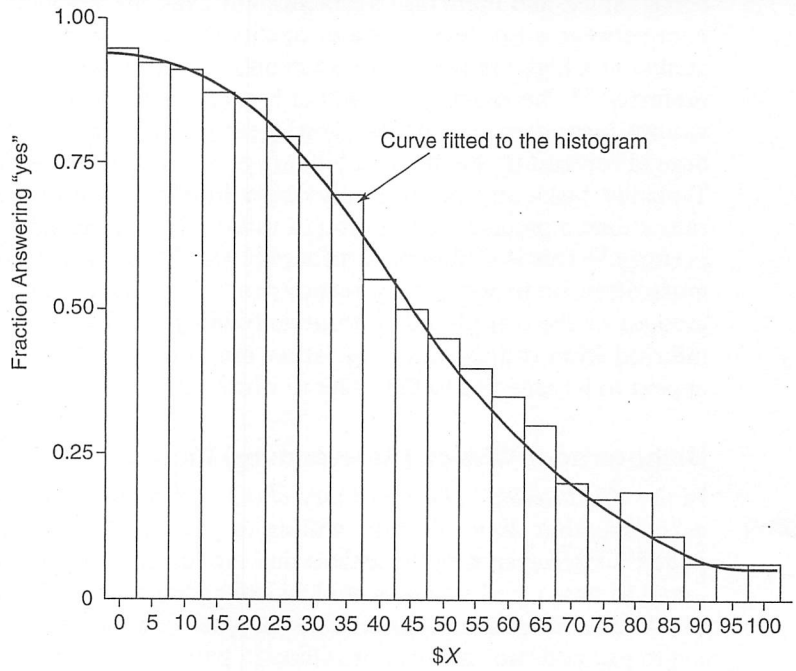
In the *dichotomous choice, binary choice,* or *referendum* method, respondents are asked whether they would be willing to pay a particular price to obtain a good or policy.[16] Each respondent receives one randomly drawn price. Respondents are then asked to state whether they would be willing to pay for the good or policy (e.g., closing the Wildwood wilderness area to development) at the offered price ("yes" means willing to pay and "no" means not willing to pay); in other words, they are made a binary

"take it or leave it" offer of the sort that they face in many markets for private goods. The choice situation is also like that faced by referendum voters—hence the label referendum method. The dollar amounts, often referred to as *bid prices,* that are presented to respondents vary over a range selected by the analyst.[17] The probability of respondents accepting the offer can then be calculated for each bid price.[18]

Figure 14-1 shows the distribution of responses to bid prices in the form of a histogram. Specific bid prices are shown on the horizontal axis ranging from the lowest dollar price offered (X = \$0) to the highest price offered (X = \$100) in \$5 increments. The vertical axis measures the percentage of respondents who answer "yes" to the bid price offered to them. In this example, almost all of the respondents who are offered the specified outcome at X = \$0 state they would accept it. About 75 percent of respondents who are offered the outcome for \$30 indicate that they would accept it at this price. We can interpret the response frequencies as estimates of the probability that a randomly drawn member of the sample of respondents is willing to pay a specific amount. For example, the probability a randomly drawn respondent would pay at least \$30 for the specified outcome is about 0.75.

The fitted curve in Figure 14-1 may be viewed as the demand curve of a randomly drawn (i.e., average) member of the sample, while the histogram may be viewed as a rough approximation of this demand curve. The difference between the demand curve in Figure 14-1 and the more standard demand curve is that instead of the curve indicating the quantity of a good the individual would be willing to purchase at each price, the curve indicates the probability that the individual would be

FIGURE 14-1 Histogram of Dichotomous Choice Responses



willing to pay for the specified outcome at each price.[19] As in the case of a standard demand curve, the area under the curve in Figure 14-1 provides an estimate of the individual's willingness-to-pay.

If the values of X are evenly spread, then the histogram can be readily used to obtain a rough estimate of the average individual's willingness-to-pay by applying a simple formula:

$$\text{WTP} = v\sum_{k=0}^{N}[\text{Probability of acceptance at price } kv] \qquad \textbf{(14.1)}$$

where v is the interval between prices (i.e., the width of the individual bars in the histogram) and N is the number of values of X (i.e., the number of bars). In other words, the area covered by the bars (the approximate WTP for an average member of the sample) can be computed by simply summing the heights of the bars and multiplying by the bar width.

Analysts rarely work directly from the histogram of accepted bids. Better estimates can usually be made by estimating a statistical model for predicting the probability that an individual with specific characteristics will accept a particular bid price. The estimated model, often a logistic regression with bid price as one of the explanatory variables, allows the analyst to sketch the relationship between bid price and the probability of acceptance.[20] The area under this curve approximates the mean willingness-to-pay for these individuals.[21] It can be estimated through numerical methods.[22] The statistical model can also be used to estimate a probability-of-acceptance curve for each identified group within the sample. Therefore, it is possible to estimate a WTP for an average member of each of these groups as the area under each group's probability-of-acceptance curve. *To find the aggregate WTP for the entire population, one multiplies the mean for each group by the size of that group in the population with standing and then sums across groups.*[23]

Because this method only tells one whether a given respondent's valuation is greater or less than the offered amount, sample sizes have to be large in order to achieve reasonable levels of precision. For small samples, the true value of WTP may be considerably different from the observed value of WTP. On the other hand, "take it or leave it" questions are very easy for respondents to answer, and they are the closest approximation to market-like transactions.

Some analysts have now moved to *double dichotomous choice* (sometimes called *double bounded dichotomous choice*) rather than single elicitation questions to reduce the need for large samples. In this version of the method, depending on the answer to the first offer, a follow-up offer is made that is either double (if yes) or half (if no) the first offer. This provides considerably more information than the standard single-offer version. However, there is danger that the respondent's exposure to the first offer may affect the probability that he or she will accept the follow-up offer.[24] Recognizing that follow-up bids provide information to respondents that may change their perceptions of the likely price or quantity of the good being valued, Richard Carson and colleagues suggest several ways that this information might affect their responses to the second bids.[25] First, those who initially rejected bids may take the lower follow-up bid as an indication that the bid price is negotiable and reject bids that they previously

would have accepted. Second, those who initially accepted bids may believe that there is some possibility that the good could be provided at the lower price and reject the higher second bid even though it is below their WTP. Third, in responding to the follow-up bid price, respondents may base their responses not on that price but on some weighted average of the two bid prices they encountered.[26] Whether the greater amount of information generated by the double dichotomous choice method is worth the risk of such response behaviors remains an open research question.

As we discuss in detail in a subsequent section of this chapter [see *the Strategic Response (Honesty) Problem*], a major advantage of a binary choice formulation is that it meets the necessary condition for incentive compatibility—the incentive for respondents to give truthful rather than strategic answers. The other methods we discuss are not incentive compatible. The possibility that the referendum method is incentive compatible is an important reason that it is the most commonly used CV method.

PAYMENT VEHICLE

Almost all CV exercises specify a *payment vehicle,* which describes how the respondent's costs of providing the good will be paid. Payment vehicles include taxes paid into a fund specifically earmarked for the good, increased utility bills, higher income or sales taxes, and higher product prices. Specifying a payment vehicle, along with reminders that payments come at the expense of expenditures on other goods, helps ensure that respondents perceive the questions as real economic choices. In order to increase the realism of CV surveys, researchers generally try to specify a payment vehicle that is as close as possible to the actual one that would be used were supply of the good to proceed.

The choice of payment vehicle can make a difference in the estimated WTP. In a recent study, for example, researchers found that respondents appeared to value watershed management plans differently depending on whether there was a guarantee that taxes imposed to pay for them would not be used for other purposes.[27] The presence or absence of such a guarantee also affected respondents' valuations of specific attributes of watershed plans.

There has been considerable debate in the CV literature as to whether differences in the WTP of respondents due to differences in payment vehicles should be treated as a bias. Kenneth Arrow and others argue that respondents are being asked to value all elements of a project, which include the method of payment; therefore, respondent preferences about payment methods do not imply bias.[28] Other analysts argue that if specific payment vehicles, such as taxes, introduce "protest" valuations, then such outliers should be excluded from the estimation of aggregate WTP. Protest bidders can sometimes be screened out by their answers to specific questions designed to identify them.

GENERIC SURVEY ISSUES

Before turning specifically to CV surveys, it is useful to consider briefly some issues relevant to all survey contexts. One issue concerns the trade-offs among methods for administering surveys. Another issue is the extent to which the procedures for identifying and reaching respondents leads to an appropriate sample from which to estimate the distribution of attitudes within the population with standing.

Survey Administration: In-Person, Telephone, Mail, and Internet

Currently, there are three major technologies for administering surveys: in-person interviews, telephone interviews, and mail questionnaires. Each has strengths and weaknesses, and each raises different methodological issues. None of the three procedures is unequivocally superior to the others. The Internet is now joining the list as one of the major technologies for administering surveys.

Table 14-1 summarizes the characteristics of the survey administration alternatives. In-person interviews involve interviewers meeting face-to-face with respondents. Especially relevant in the CV context, face-to-face interaction facilitates the provision of complex information to respondents through the presentation of maps, diagrams, photographs, and other visual aids by the interviewer. The interviewer can also clarify questions and provide additional information and otherwise interact directly with the respondent. This direct contact, however, also involves a high risk of interviewer bias, as respondents may react to the personal characteristics of the interviewer, perhaps slanting answers to gain the interviewer's approval. Unlike telephone interviews, which typically are conducted from a central location, in-person interviews are difficult to monitor. The greatest disadvantage, however, is that in-person interviews tend to be very expensive, especially for geographically disperse samples, because of the time spent traveling between interviews and the precautions that must be taken to ensure the security of the interviewers. As is the case with mail questionnaires, the capability for identifying a random sample of any population depends on the availability of lists of the individuals, families, or households that make up the population. In many locations, such as cities with guarded apartment buildings or suburbs with restricted-access housing developments, it may be difficult to reach respondents randomly selected to be in the sample.

Telephone interviews have become the most common method of administering CV surveys. Telephone interviews cost substantially less than in-person interviews, even with a large number of call-backs to reach persons who are not at home or who use answering

TABLE 14-1 Survey Administration Alternatives

	Cost per Completed Interview	*Ease of Identifying and Reaching Respondents*	*Risk of Interviewer Bias*	*Maximum Complexity of Provided Information*
In-Person	Very high—depends on questionnaire length and geographic spread	Medium—depends on availability of lists and access	High—personal presence, monitoring difficult	Very high—interactive communication and visual aids possible
Telephone	High—depends on questionnaire length and call-backs	Very high—random digit dialing	Medium—interviewer cues	Low—verbal communication limits complexity of content
Mail	Low—depends on number of follow-ups	High—depends on availability of appropriate lists	Low—uniform presentation	High—visual aids possible
Internet	Low—marginal costs very small	Low—"spamming" restrictions require panels of willing respondents	Low—uniform presentation	Very high—visual aids and interactive questions possible

machines. They also have the advantage of allowing researchers to draw reasonable random samples through random digit dialing, obviating the need for address lists in surveys of households. Unfortunately, verbal communication limits the complexity of information that can be provided. It also opens up the possibility of interviewer bias, as respondents react to voice cues and perceptions of the characteristics of the interviewer.

Several trends are making telephone interviews more costly and difficult to implement effectively. One trend is the greater prevalence of telemarketing, often masquerading as surveys. More people are refusing to participate in surveys to avoid unwanted telemarketing solicitations. Others use answering machines and caller identification to screen out calls from strangers. Another trend is the growing complexity of the telephone system, which makes it increasingly difficult to draw good random samples through random digit dialing. Households more often have multiple telephone lines, some of which are dedicated to facsimile machines or teenagers. Cellular telephones are breaking down area codes, making it more difficult to sample geographically. Dealing with these problems will likely raise the costs of telephone surveys in the future.

Mail questionnaires, which allow the provision of visual aids to respondents, have the advantage of very low cost. The ease of identifying samples depends on the availability of appropriate lists. Accurate lists also make it easy to identify a sample of respondents and contact them by mail, though response rates are typically low, often requiring multiple mailings to reach acceptable levels. As there are no interviewers, all respondents experience the same presentation, and therefore there is no interviewer bias.

As more U.S. adults become Internet users, the plausibility of using the Internet for conducting surveys increases. Internet surveys have several advantages over the other methods. Because interviewers are not needed, Internet surveys have extremely low marginal costs, perhaps even lower than mail surveys. Unlike mail surveys, the actual data collection can often be fully automated and implemented within a very short time frame, an advantage when the CV survey is intended to inform a pending decision. Like mail surveys, they avoid the risk of interviewer bias, and they allow for the provision of complex information. Indeed, one can imagine giving respondents access to large quantities of information through menus and other interactive devices. Unfortunately, drawing random samples of populations remains a barrier. Not all members of most populations of interest are currently Internet users. This is changing, but even when Internet use has become ubiquitous, sampling is complicated by restrictions on spamming, which prevents procedures analogous to random digit dialing in telephone surveys. Firms are beginning to develop databases of willing respondents, which may become sufficiently large and representative of populations of interest to allow scientifically valid CV studies to be administered through the Internet.[29]

Sample and Nonresponse Biases

The essence of survey research is eliciting information about a population from a small sample drawn from that population. As summarized in Table 14-1, survey administration alternatives differ in terms of the ease with which they facilitate identifying and reaching respondents. How to identify individuals to be sampled from target populations is the topic of *sample design.*[30] The extent to which valid responses are obtained from those identified determines whether the sample design will actually produce data adequate for making valid inferences.

In almost all cases, the sample is selected by a probability mechanism, which produces a *random sample.* In a random sample, each individual has a known probability of being drawn from the population. *Simple random samples* give each individual in the target population the same probability of being sampled. *Stratified samples* give members of particular groups within a population the same probability of being sampled. In either case, knowing the selection probabilities allows researchers to base inferences about the characteristics of the population on the characteristics of the sample.

If a given sample is appropriately selected and administered, then sample biases can be avoided. Advances in probability sampling techniques have been such that findings based on samples of approximately a thousand people can be representative of the entire population of the United States. However, approximately the same sample size may also be necessary for much smaller populations—essentially, for large populations the sample size needed to achieve any desired level of precision does not depend on the size of the population being sampled. Additionally, there is considerable evidence that the observed distributions of WTP in CV are skewed toward extreme values. For this reason, CV samples should be larger than samples drawn for many other purposes to obtain reliable estimates of population means. This is especially true for the dichotomous choice method.

For CV purposes, *the relevant target population is usually all individuals with standing who are affected by the policy.* Unfortunately, this heuristic begs the question of who is affected. For many projects it is apparent. But often it is not, especially in addressing environmental issues. In some contexts there is congruence between those who bear the cost of policies and those who are affected by them. In other contexts, however, these groups diverge, as when a state or province provides a wilderness area that is valued by people living in other jurisdictions. The greater such divergence, the more problematic it is to choose the correct population.[31]

Several major issues are involved in assessing who is affected. First, all those "users" directly affected by the project should be included. The term *users* is in quotes because we mean it in a specific way. Those who would directly utilize the good in question are, of course, users. But so are individuals who suffer direct negative impacts that they would pay to avoid. For example, nearby residents of a duck hunting reserve who dislike the noise are as much users as the duck hunters. Potential users should also be included. As discussed in Chapter 8, in situations involving uncertainty, the option price that an individual would be willing to pay for a good may differ from his or her expected surplus. So even people who never actually consume the good may value it.

Second, it is important for survey respondents to understand whether they are being asked to estimate WTP just for themselves or as a representative for their whole household. This distinction is important in extrapolating from the sample to the target population.

Third, an explicit decision should be made concerning the inclusion of passive use benefits. As discussed in Chapter 9, either users or nonusers may derive existence value from a project. Conceptually, existence value should be included as a component of benefits. For environmental goods, CV surveys that either sample nonusers or estimate existence values of users typically yield much higher aggregate WTP estimates than those that include only use benefits. However, there is considerable disagreement on the validity of using CV surveys to estimate nonuse benefits.

Fourth, the geographic *spread,* or *reach,* of the sample should be wide enough to capture all affected individuals. There is increasing recognition that decisions concerning the geographic definition of the relevant market can drive the outcomes of many CBAs, especially if nonusers are included.[32]

An important sampling question relates to the exclusion of responses. It has been suggested that three categories of respondents should be excluded in estimating WTP: first, respondents who reject the whole notion of placing a value on the good in question, or of paying for the good in a certain way (this has already been discussed); second, respondents who refuse to take the exercise seriously; and third, respondents who clearly demonstrate that they are incapable of understanding the survey.[33] In the direct elicitation methods, all three types of respondents are usually assumed to provide either zero valuations or extremely high valuations. Sometimes such respondents can be directly identified by their answers to specific questions intended to screen them from the sample. Respondents who provide extreme values are known as *outliers.* Outliers are normally handled in CV by simply eliminating valuations that are above some specified threshold or that are above a specified percentage of the respondent's gross or discretionary income.

An appropriate sampling design can usually eliminate most sample bias if it is fully executed. Yet, bias can still remain if some individuals do not respond to the survey. Nonresponse bias is a serious problem in almost all survey research. Nonresponse problems have grown over the last 20 years as the public has been asked to give time to more surveys and has become suspicious of the motives of many who claim to be survey researchers. If nonresponse is purely random, then it can be dealt with by increasing the sample size. Unfortunately, nonresponse is often not random.[34]

There are two major types of nonresponse problems: refusal to respond and unavailability to respond. In CV contexts, the primary approaches for dealing with

EXHIBIT 14-1

As part of a court case, the plaintiffs conducted a CV survey to estimate the natural resource damage caused by a mine. They surveyed residents of both the county (Eagle County) and the state (Colorado) in which the mine was located. Based on their surveys, they estimated past damages were $50.8 million and future expected damages would be between $15 and $45 million. The defendants sampled a much smaller group within Eagle County that they believed had been directly affected by pollution from the mine; they assumed that residents in the rest of Colorado did not bear costs from the mine. Although the per-unit values of both sides were similar (for example, on the value of a day's fishing), the defendants' estimate of total past and future expected damage was approximately $240,000, less than 1 percent of the plaintiffs' estimate. "The discrepancies in these respective aggregate estimates arise from the plaintiff's assumption that . . . there would be a much larger number of people experiencing gains with the restoration" (p. 605).

Source: Adapted from Raymond J. Kopp and V. Kerry Smith, "Benefit Estimation Goes to Court: The Case of Natural Resource Damage Assessments," *Journal of Policy Analysis and Management* 8(4) 1989, 593–612.

refusal to respond are to highlight the legitimacy of the exercise (e.g., by stressing government or university affiliations) or to offer various response incentives, such as donations to charities or entries into prize lotteries. Where unavailability biases the sample, researchers typically account for underrepresentation and overrepresentation in the sample when extrapolating to the target population.[35]

CONTINGENT VALUATION PROBLEMS AND ISSUES

This section discusses specific problems relevant to CV surveys. The subsequent section discusses more general survey problems that are relevant to CV. Surveying opinions is not an exact science. The science of CV surveys is even less exact. Some critics argue that CV has such serious weaknesses that it should be used only as a last resort. Other critics contend that it is so seriously flawed when used to value either complex goods or passive use impacts that it should not be used at all for these purposes.[36] It has been pointed out that the summation of average valuations over a broad range of projects to improve the environment would exhaust the budget of average individuals.[37]

Specific CV survey difficulties stem from several sources. First, CV inevitably raises questions that are more novel and complex than those raised in other survey situations. This poses problems of *hypotheticality* (we use this as a catchall word to cover problems of understanding, meaning, context, and familiarity). Hypotheticality appears to be more severe when respondents have not, and will not, consume the good in some way. That is, hypotheticality tends to be most severe in the valuation of passive use. Second, CV raises questions of *neutrality* in the presentation of information to respondents. Third, for certain methods, *judgmental biases* may arise in response to certain kinds of questions (including whether the question is framed as willingness-to-pay or as willingness-to-accept). Not all these problems necessarily create biases (i.e., a systematic tendency to overvalue or undervalue the goods in question), but all of them do raise questions about the validity and reliability of CV as a procedure. Some specific CV methods appear to be more prone to biases than others. Fourth, CV asks questions about WTP, raising the potential for biases related to *strategic behavior* (misstatements intended to influence some outcome) and the specified payment vehicle.

Hypotheticality, Meaning, and Context Problems

A major concern in CV design is whether respondents are truly able to understand and place into context the questions they are being asked and, consequently, whether they can accurately value the good in question. Issues relating to the valuation of the supply of many publicly provided goods are complex and highly contextual. CV questions can be contrasted to many other types of questions for which meaning is not an issue (eg., "For whom do you intend to vote in the next election?").

Questions of hypotheticality and meaning can be thought of as problems of specifying exactly what is the good in question. Understanding the good, or the policy that produces it, is difficult for respondents because they often are not familiar with either. When respondents are presented with questions about goods or projects that they do not understand, attitudes (and responses as expressed in the CV survey) are unlikely to correspond to the behavior that would occur if the project were actually implemented.[38] When a project (or the good itself) has multiple attributes, these all

need to be explained to respondents: "Unless [an attribute is] specified explicitly (and comprehensively), evaluators must guess its value and, hence, what the offer really means. If they guess wrong, then they risk misrepresenting their values."[39]

This problem, however, has to be seen in context. Individuals also differentially value attributes of market goods: Some individuals may value a mountain bike mostly for prestige reasons and others for transportation purposes. The evidence also suggests that people find it somewhat difficult to value the attributes of new and unfamiliar products in market contexts.[40] As Richard Carson and his colleagues point out, "Many new products become available each year creating markets in which consumers regularly make purchase decisions . . . No standard microeconomics text has ever stated that prior experience is a precondition for rational decision making."[41]

Additional problems arise in CV if the perceptions of the good by respondents are not independent of the quality or quantity of the information provided. The possible information that could be provided when describing complex goods are unlimited. The quantity and quality of information, however, are limited in practice by the method of survey administration. Several commentators have emphasized that there is little evidence that hypotheticality per se introduces bias into CV.[42] But in the presence of hypotheticality certain kinds of bias may be more likely.

The potential for hypotheticality varies enormously across different CBA contexts. Unfortunately, CV is likely to be most useful in contexts in which goods are difficult to define, such as projects involving environmental impacts. When it is difficult to specify potential physical impacts, it is also likely to be difficult for respondents to understand what these impacts mean.

Hypotheticality and lack of realism can be reduced in a number of ways. *Clearly specifying the project and its impacts increases the likelihood of correspondence between attitudes and behavior; so too does providing explicit detail about the payment vehicle.* Visual aids such as photographs, maps, and diagrams often assist in understanding. One important class of visual aids useful in reducing hypotheticality is known as quality ladders. An example of a quality ladder is described in Exhibit 14-2. Quality ladders help respondents understand both what quality is under the status quo, and what particular increments of quality mean.

Baruch Fischhoff and Lita Furey have suggested a checklist for evaluating CV instruments in terms of the likelihood that respondents will understand the questions they are being asked.[43] It requires the analyst to assess the comprehensiveness of information with respect to the good, the specification of the payment vehicle, and the social context. In assessing the adequacy of information about the good, they stress the need to provide information on both substantive and formal components. The substantive aspect of the good deals with why someone might value it (basically its attributes), while the formal aspect of the good concerns how much they value it (once they understand its attributes).

In practice, *the only effective way to minimize hypotheticality and meaning problems in CV surveys is to devote extensive effort to developing detailed, clear, informative, and highly contextual materials and to pretest these materials extensively on typical respondents.*

Neutrality

While the previous section indicated that lack of clear meaning does not necessarily pose a bias problem, lack of neutrality is certain to do so. As CBA deals with increasingly

EXHIBIT 14-2

A "water quality ladder" has been used in several CBAs to help respondents understand how differing levels of toxins, dissolved solids, water clarity, and other factors affect water quality. In their CBA of water quality improvements to the Monongahela River, V. Kerry Smith and William Desvousges included a picture of a ladder with a 0-to-10 scale and the following interviewer instructions:

(*Interviewer: Read the following.*) Generally the better the water quality the better suited the water is for recreation activities and the more likely people will take part in outdoor recreation activities on or near the water. Here is a picture of a ladder that shows various levels of water quality. (*Interviewer: Give respondent water quality ladder.*)

The top of the ladder stands for the best possible quality of water. The bottom of the ladder stands for the worst possible water quality. On the ladder you can see the different levels of the quality of the water. For example: (*Interviewer: Point to each level—E, D, C, B, A—as you read the statements that follow*).

Level E (*Interviewer: Point.*) is so polluted that it has oil, raw sewage, and other things like trash in it; it has no plant or animal life and smells bad.

Water at level D is okay for boating but not fishing or swimming.

Level C shows where the water is clean enough so that gamefish like bass can live in it.

Level B shows where the water is clean enough so that people can swim in it safely.

And at level A, the quality of the water is so good that it would be possible to drink directly from it if you wanted to.

(*Interviewer: Now ask the respondent to use the ladder to rate the water quality in the Monongahela River on a scale of 0 to 10 and to indicate whether the ranking was for a particular site, and if so, to name it.*)

Source: Adapted from V. Kerry Smith and William H. Desvousges, *Measuring Water Quality Benefits* (Boston, MA: Kluwer Academic, 1986), 87.

controversial and complex topics, the neutrality of the CV questionnaire becomes an increasingly important issue. Neutrality has come to the fore as litigants in (especially environmental) court cases have conducted their own CV surveys.

Meaning and neutrality issues often intersect in ways that are extremely difficult to disentangle. For example, Daniel Hagen and colleagues surveyed 1,000 U.S. households by mail concerning the value of preserving the spotted owl.[44] Of the total, 409 questionnaires were returned. Some of the information that respondents were given included the following: " . . . a scientific committee concluded that logging should be banned on some forest lands to prevent the extinction of the Northern Spotted Owl . . ." and "a second group of independent scientists examined this study and agreed with these conclusions." The survey also included the comment that: "the well-being of the northern spotted owl reflects the well-being of the entire old-growth forest eco-system."

In a review of spotted owl CV studies in general, and the Hagen et al. study in particular, William McKillop criticized this framing of the issue. He argues that the survey did not include many relevant facts that respondents should have been told.[45] For example, the "committee of scientists" focused almost exclusively on old-growth habitat for spotted owls and largely ignored the fact that many are found in second-growth timber stands. Respondents were also not told that logging was already prohibited on

considerable areas of old-growth timberland, and these acres were likely to increase in the course of normal national forest planning. In sum, McKillop argues that the spotted owl issue was not presented accurately or neutrally to respondents. He further argues that this issue is simply too complicated to be addressed by CV surveys.

There are no simple answers to the neutrality problem. But an inevitable conclusion is that one has to be especially cautious in interpreting the results of CV surveys that have been prepared by either parties to litigation or advocacy groups. At a practical level, *neutrality can best be ensured by pretesting the survey instrument with substantive experts who have "no axe to grind" in terms of the specific project that is being considered.* If neutral experts cannot be found, then pretesting with opposing advocates can be an alternative, perhaps enabling researchers to avoid the most serious challenges from those with positions not supported by the results.

Decision Making and Judgment Biases

Although it is reasonable to assume that individuals can make rational judgments about their valuations of goods in most market situations, evidence suggests that in certain circumstances they may not be able to do so readily. This is even more likely to occur in the context of CV surveys, because judgment rather than decision making is involved, and because there are not opportunities to learn from mistakes (we discuss the evidence on this issue later in the chapter).[46] More formally, in such circumstances, there is a tendency for individuals to behave as if they are not maximizing utility, especially with respect to choices involving uncertainty. In the context of functioning markets, these behaviors appear as decision-making biases that can result in irrational purchases (or lack of purchases). These decision-making errors can be thought of as a type of market failure.[47] In the context of CV, the term *judgment bias* rather than *decision-making bias* is applicable because the respondent is not actually purchasing the good in question.

Both decision-making and judgment biases appear to be most serious for activities or projects that would generate small changes in the probabilities of (already) low-probability events that have "catastrophic" costs if they occur (for example, activities that might cause a marginal change in the probability of a nuclear power plant accident).[48] As WTP depends on how likely respondents perceive such events will be, their perception of changes in probabilities is important in CV studies. Fortunately, researchers and analysts rarely rely solely on CV estimates in such contexts. For example, they can use value-of-life estimates derived from other methods in which these biases are less endemic (see Chapter 15).

Some of the major judgmental biases to which individuals are particularly prone include *availability bias,* whereby individuals estimate the probabilities of events by the ease with which occurrences can be recalled—more salient instances, such as those covered by the media, are more likely to be recalled; *representativeness,* or *conjunction bias,* whereby individuals judge the probabilities of events on the basis of their plausibility—people perceive the probability of an event as being higher as more detail is added, even though the detail is irrelevant; *optimism bias,* whereby people believe that they can beat the objective odds; *anchoring bias,* whereby individuals do not fully update their probability assessments as new information becomes available; *hindsight bias,* whereby individuals believe, after an event occurs, that it was more predictable

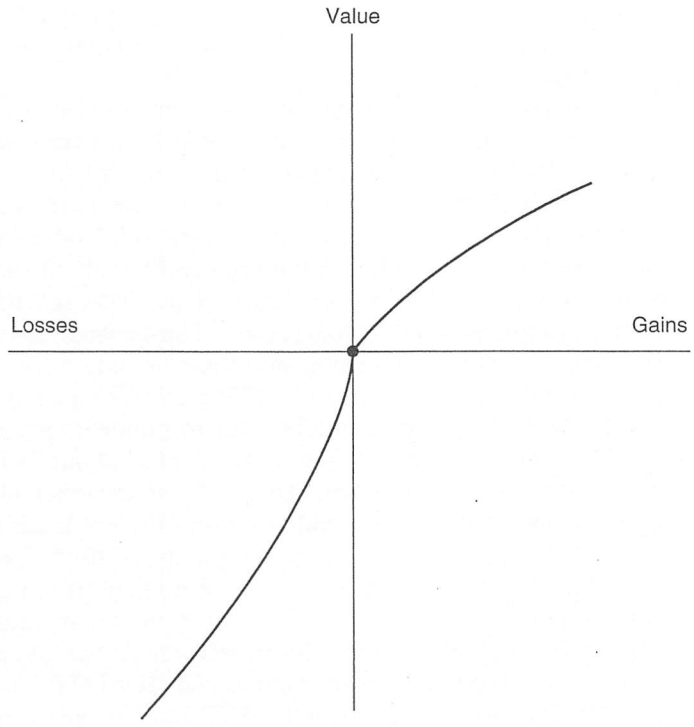
than it actually was; *status quo bias,* whereby individuals stick with the status quo even when it is inexpensive to experiment or when the potential benefits from changing are large; and *probability assessment bias,* whereby people tend to overestimate small probabilities and underestimate large probabilities.[49]

Many of these violations of the expected utility hypothesis can be explained by the fact that, when dealing with complex information, people tend to use simplifying (non-utility-maximizing) heuristics, or "rules of thumb." Essentially, people frame problems consistently, but the framing does not correspond to maximization of expected utility.

One conceptual framework for explaining violations of the expected utility hypothesis is *prospect theory.*[50] Prospect theory, which is particularly relevant to CV issues, suggests that individuals deviate from expected utility maximization in several ways. They value gains and losses from a reference point rather than valuing net wealth. Moreover, people are risk averse toward gains and risk seeking toward losses (known as *loss aversion*)—a loss and a gain of the same size would leave a person with loss aversion worse off. This may stem from an *endowment effect,* whereby individuals have a greater psychological attachment to things they currently possess.[51]

Several of these effects can be summarized in a prospect theory value function, which is illustrated in Figure 14-2. In this figure, the vertical axis measures value and the horizontal axis measures losses and gains. The figure is drawn to show three things. First, people start from a reference point from which changes are measured as losses or gains.

FIGURE 14-2 A "Loss Aversion" Value Function



Second, individuals are risk averse with respect to potential gains (i.e., they prefer a smaller certain gain over a larger probable gain, when the expected values of the two alternatives are the same). Individuals are also risk seeking with respect to potential losses (i.e., they prefer a larger probable loss to a smaller certain loss, when the expected values of the two alternatives are the same). This is represented in the figure by the concave gain function and the convex loss function. Third, losses loom larger than gains of equal size. This is represented in the figure as the loss function being steeper for losses than for gains from the reference point.

The biases suggested by prospect theory are particularly relevant to CV for a number of reasons. Anchoring via reference points is always a potential problem. Even if open-ended questions are used to eliminate starting point bias, payment vehicles and other descriptive detail may introduce anchoring indirectly. Furthermore, detailed descriptions may evoke availability bias. Finally, as we discuss later, CV questions can sometimes be plausibly framed as either involving gains or losses.

Empirical research suggests that the most serious judgment problems actually found in CV results relate to *noncommitment bias, order effects* (sometimes described as sequencing effects), *embedding effects* (sometimes described as whole/part effects or inclusiveness effects), and *starting point bias.*

Noncommitment Bias[52] It is well recognized in the marketing literature that respondents to surveys tend to overstate their willingness to purchase a product that is described to them.[53] This may be a strategic response (an issue discussed later), but can also be thought of as either a form of anchoring bias (e.g., "this product must be valuable because they are asking me about it and describing it in such detail") in a context in which potential consumers do not engage in any learning. It is likely to be quite unconscious. The bias can flourish, of course, because the respondent does not actually have to commit money.

One set of experiments has attempted to test for noncommitment bias. The researchers conclude that "hypothetical WTP is consistently and significantly higher than the WTP that reflects real economic commitments."[54]

It is difficult to test for noncommitment bias when dealing with passive use values. One indirect way of testing for the bias is to introduce elements to the survey that encourage respondents to think more carefully about their income and their budget constraints. Michael Kemp and Christopher Maxwell have developed a "top-down disaggregation method" to do just this.[55] Top-down disaggregation attempts to mitigate noncommitment bias by raising awareness of budget constraints. After asking respondents to state initially their total WTP, they were questioned specifically about comparative valuations. For example, after respondents were asked about their WTP to avoid a specified oil spill, they were then asked about their valuations of environmental protection versus reduction in crime, homelessness, and other social problems. They were also asked about their valuations for different kinds of environmental protection (wilderness areas versus groundwater quality, rainforest protection, and other environmental goals). At the next level, they were asked to evaluate various kinds of wilderness area protection (reduction in harm from human-caused problems versus natural degradation and other destructive processes). At the end of this top-down disaggregation process, respondents were again asked their WTP. The result was WTP values several hundred times smaller than the WTP values from the initial open-ended questions.

In the more commonly used CV methods, it may be desirable to ask questions that require respondents to think more carefully about the budget constraints that they face.[56] For example, asking questions that encourage respondents to think about how much of their income is discretionary may help avoid noncommitment bias. The answers may also be useful in helping to assess whether the WTP values provided are subject to noncommitment bias.

Order Effects George Tolley and Alan Randall found that survey respondents' statements of the value of improved visibility in the Grand Canyon were greatly affected by the order in which the issue was raised.[57] Other studies have also found important order, or sequence, effects. Consider, for example, a study that asked some respondents to value preserving seals and then whales, while others were asked to value preserving whales and then seals. Seal values were considerably lower when the seal question was asked after the whale question.[58]

These findings could be explainable in terms of either an income effect, a substitution effect, or a combination of both. The rationale for an income effect is as follows: If someone has expressed a positive WTP to pay for the first good in a sequence of goods, then that person has less income to spend on the second good in the sequence. Critics of CV, however, have observed that the steep declines in WTP as a good moves down in order cannot be fully explained by income effects because these should be relatively small.[59]

Substitution effects, such as between seals and whales, could be quite large, and consequently they can be important in CBA, especially in terms of assessing the aggregate impacts of projects. (That is, respondents may view one environmental improvement as a substitute for another quite different environmental improvement.) If people engage in extensive substitution, then the net aggregate benefits from a project may be smaller than predicted. For example, if a resident of Chicago agrees to contribute to a project that cleans the air in Chicago and is then offered a project that preserves visibility in the Grand Canyon, the value of the Grand Canyon project may be decreased because of the substitution effect.

The issue of whether substitution effects could account for much of the order inconsistency in CV surveys of passive use values is still undecided.[60] Critics argue that the phenomenon is explained neither by income nor by substitution effects, but it instead demonstrates that respondents cannot really understand these kinds of questions. Hence, they inevitably engage in judgment heuristics that usually cause them to overstate valuations. Thus, it is uncertain that income and substitution effects provide a complete explanation for order effects.[61]

Embedding Effects A fundamental axiom in the standard economic description of preferences is that individuals value more of a good more highly than less of it. This is sometimes referred to as a *scope test.* If CV respondents' valuations are only slightly higher for large changes in the amount of the good offered than for small changes, then the validity of their responses becomes a concern. But research indicates that individuals often do not readily distinguish between small and large quantities in their valuations of a good when the different quantities are *embedded* in one another. For example, William H. Desvousges and colleagues found that different samples of respondents value 2,000 migratory birds approximately the same as 200,000 birds, and small oil spills much the same as much larger oil spills (given that these are different

samples, it does not directly test whether given individuals prefer more to less).[62] Two additional examples of embedding are described in Exhibit 14-3. It is unlikely that declining marginal utility can explain all or even most of the absence of different valuations for different quantities of goods.

When dealing with passive use values, embedding is probably the most worrisome problem identified by critics of CV. It goes to the very heart of welfare economics. Critics argue that the empirical evidence suggests that in these contexts respondents are not actually expressing their valuations but instead are expressing broad moral attitudes to environmental issues—a "warm glow" or "moral satisfaction."[63] However, Carson and colleagues have argued that most studies that manifest embedding problems, or scope insensitivity, are poorly designed and executed, and they further argue that well-designed CV instruments do not manifest the problem.[64]

Starting Point Bias Prospect theory identifies anchoring as a common behavioral response to being asked to make complex judgments. A problem arises in CV when starting values are presented to respondents. The iterative bidding method is particularly prone to this problem because this method provides respondents with a specific initial starting "price."[65] Consider, for example, the Kristiansand Fjord study mentioned earlier in the chapter. It was found that a range starting at 200 kronor and progressing to 2,000 kronor produced different valuations than a range starting at 0 kronor and progressing to 2,000 kronor, even when there were no bids between 0 and 200 kronor.[66]

The dichotomous choice question format is intended to eliminate starting point bias. It has been argued, however, that "responses to dichotomous choice questions are

EXHIBIT 14-3

Daniel Kahneman and Jack L. Knetsch report that residents of Toronto expressed a willingness to pay increased taxes to prevent a drop in fish stocks in all Ontario lakes that was only slightly larger than their expressed WTP to preserve fish stocks in a small area of the province. This is implausible.

The same researchers studied the impact of embedding. They did so by specifying a good: (1) very broadly to one sample of respondents (environmental services that included "preserving wilderness areas, protecting wildlife, providing parks, preparing for disasters, controlling air pollution, insuring water quality, and routine treatment and disposal of industrial wastes"; (2) considerably more narrowly to a second sample of respondents ("to improve preparedness for disasters" with a subsequent allocation to go to "the availability of equipment and trained personnel for rescue operations"), and (3) more narrowly still to a third sample ("improve the availability of equipment and trained personnel for rescue operations"). Respondents in each sample were asked to express their willingness to pay for their "good." The differences in WTP among the three samples were not large.

Source: Adapted from Daniel Kahneman and Jack L. Knetsch, "Valuing Public Goods: The Purchase of Moral Satisfaction," *Journal of Environmental Economics and Management* 22(1) 1992, 57–70. (For a critique of this study, see V. Kerry Smith, "Comment: Arbitrary Values, Good Causes, and Premature Verdicts," *Journal of Environmental Economics and Management* 22[1] 1992, 71–89.)

very strongly influenced by starting-point bias, because respondents are likely to take initial cues of resource value from the solicited contribution amount (e.g., assuming it to be his/her share of the needed contribution)."[67] Moreover, there is at least some empirical evidence to support the contention that the dichotomous choice method is subject to starting point bias,[68] although WTP estimated from dichotomous choice responses appears to be less affected by the provision of information about the total cost of providing a good or the size of the group receiving it than does WTP estimated from open-ended surveys.[69]

Hypotheticality Bias versus Judgment Bias In the last resort it is almost impossible to disentangle hypotheticality issues from various forms of judgment bias in nonuse contexts because they are essentially two different sides of the same coin. Hence, the same approach is required to minimize both: an effort to present the project or good in question as concretely as possible, at the lowest level of disaggregation as possible, and with as much realism concerning budget constraints as possible. Neutrality problems also appear to be generic to CV. But this problem is less worrisome because it can be made reasonably transparent with good research protocols.

WTP versus WTA

Economic theory implies that if individuals behave rationally, and if markets are working efficiently, then in most circumstances it should make little difference whether respondents to a CV survey are asked their WTP for receiving a good or their willingness-to-accept (WTA) the loss of a comparable good.[70] Similarly, it should make little difference whether they are asked their WTP to prevent a loss or their WTA for a comparable loss.

As we have already discussed, considerable evidence suggests that individuals demand greater monetary compensation to give up things that they already possess than they are willing to pay to acquire the same exact item. In experiments that actually require people to trade goods for money, as well as in other contexts, it has been found that required WTA amounts range from a ratio of four to fifteen times greater than WTP amounts.[71] Similar differences have also been found in CV studies, even with a well-specified market good. There is some evidence that this difference between WTP and WTA is attributable to loss aversion.[72]

There is also considerable evidence, however, that as the subjects in experiments become more experienced, differences between WTP and WTA shrink considerably, usually as a result of decreases in WTA amounts.[73] Other experimental evidence suggests that for nonmarket goods with imperfect substitutes the divergence between WTP and WTA persists even with experience.[74] In any event, CV survey contexts, which are typically one-shot events for respondents, do not present opportunities for learning. Moreover, they often involve nonmarket goods with no close substitutes.

Some commentators have argued that stated preferences are preferences and that, if respondents are actually being asked to give up something, then the relevant formulation is WTA. But even in a context in which something is being given up, analysts still have to decide whether to treat WTA as involving a judgment bias problem that requires adjustments. In view of the fact that experiments show learning effects in some contexts and, as discussed later, WTP amounts are much closer to estimates derived from methods based on revealed preferences, the usual procedure is to use

WTP estimates.[75] This procedure is typically followed even in cases in which WTA questions fit the facts better than WTP questions—for example, as when the project involves the respondent giving up a good, such as a scenic view, that he or she currently consumes. The heuristic is, therefore, *that WTP question formats rather than WTA question formats should be used in CV in almost all cases.*

In thinking about WTP and WTA, one additional point should be kept in mind: We are concerned only about differences between the two *for the same individual.* The fact that different individuals have differing WTPs and WTAs is what makes markets work and presents no problem to CV.

The Strategic Response (Honesty) Problem

Will respondents answer honestly when asked about their WTP? It is frequently argued that respondents in CV surveys have incentives to behave *strategically,* that is, to misrepresent their true preferences in order to achieve a more desired outcome than would result if they honestly revealed their preferences. An analogy is often drawn between strategic behavior in CV studies and free riding in the provision of public goods.[76] The potential for strategic behavior in CV, however, is actually more varied than the free-riding characterization suggests.[77]

The Carson, Groves, and Machina Framework Richard Carson, Theodore Groves, and Mark Machina assess the nature of strategic responses to CV questions likely to be encountered in use of the major CV methods.[78] They begin by making the point that we should not expect respondents' answers necessarily to be consistent with economic theory, and hence appropriate for inclusion in CBA, unless respondents believe that the survey is *consequential* in the sense that it could potentially influence some outcomes about which they care. In contrast, if respondents believe that the survey will have absolutely no influence on outcomes about which they care, then it is *inconsequential* and economic theory offers no predictions about the nature of responses.

They next note that the design of consequential CV surveys falls under the theory of *mechanism design,* which deals with the problem of creating rules for collective choice based on signals sent by individuals. Mechanisms that provide incentives for individuals to reveal their preferences truthfully are called *incentive compatible.*

One of the central theorems of mechanism design is that, in the absence of restrictions on the domain of preferences,[79] mechanisms involving more than binary signals will always be incentive compatible.[80] Mechanisms employing binary signals will not always be incentive compatible, but depending on the specific circumstances of choice, they may be (in other words, a binary signal is a necessary, but not necessarily sufficient, condition). More complex signals provide an opportunity for individuals to misstate their preferences in order to obtain a more desirable outcome.

As a brief illustration, return to the choice situation presented Chapter 2. Table 2-2 shows how the mechanism, pair-wise majority rule voting, could result in an intransitive social ordering. Consider a case in which the mechanism is being implemented by first putting X against Y in round one and then, in round two, Z against the winner of round one. Note that the signal sent by the voters in this situation is not binary—it consists of *two* binary signals, one for each round. If the voters send truthful signals about their preferences, then X beats Y in round one and Z beats X in round two, so that the

mechanism selects Z as the social choice, which is voter 1's least preferred outcome. Anticipating this, voter 1 has an incentive to misrepresent his preferences by voting for Y in round one, so that Y would win and be put against Z in round 2 and win. This misrepresentation of preferences in round 1 would result in Y, an outcome more desirable to voter 1 than the one that resulted from sending a truthful signal about his preferences in round one. Once the voters reach round 2, it is as if they are facing a new choice situation in which they send only a binary signal. In this last round, there is no incentive to misrepresent preferences.

One immediate implication of this theorem is that continuous response formats, as used in the open-ended WTP method, or multiple response formats, as used in the contingent ranking method, will always be vulnerable to strategic responses that misstate true preferences.

Consider the open-ended WTP method. We can imagine two types of misrepresentation of preferences. First, assume that respondents perceive having to make payments that do not depend on their stated WTPs, that they have true WTPs above their anticipated payments, and that they anticipate that the likelihood of the good being provided depends on the aggregate of stated WTPs. They would then have an incentive to overstate their WTPs, so that the estimated aggregate WTP would be too high. The possibility of such overstatement was widely anticipated by economists. Yet at least in comparison with the dichotomous choice method, which, as we will discuss, can be incentive compatible, it does not appear that such overstatement is a major problem in the open-ended WTP method.[81] One explanation is that the assumptions that make overstatement a desirable strategy for respondents may not often hold. For example, respondents may fear that their shares of the cost of providing the good will be a positive function of their stated WTPs and, hence, state WTPs that are as small as possible. However, they may also believe that the likelihood of provision depends on the fraction of respondents who state WTPs at or above the average cost of provision, rather than on aggregate WTP. If both of these assumptions hold, then a strategic respondent with a true WTP above the anticipated cost would state a WTP that is only slightly above the anticipated cost—above anticipated cost to increase the chances of provision but as little above as possible to minimize cost share.

Second, now consider respondents with true WTPs that are less than the costs that they anticipate would be imposed on them if the good were to be provided. Assume that they believe that the likelihood of provision depends on aggregate WTP. As they do not want the good provided, they have an incentive to state the smallest possible WTP, usually set to zero for goods researchers believe to be desirable. These strategic responses in the open-ended WTP method should lead to many zeros, and few WTP amounts under the respondents' perceived costs of provision. The larger the fraction of respondents following this strategy, the greater will be the underestimation of aggregate WTP.

As it involves a binary response, the dichotomous choice method *may* be incentive compatible. Carson, Groves, and Machina note that incentive compatibility requires that payment for the good can be compelled and that the question deals only with a single issue. Table 14-2 summarizes the incentive properties of several circumstances in which the dichotomous choice method has been used.

Using the dichotomous choice method to elicit WTP for a new public (nonexcludeable) good that will be funded with coercive payments by respondents meets the

TABLE 14-2 Incentive Properties of Dichotomous Choice Questions Applied to Types of Goods

Type of Good	*Incentive Property*
New public good with coercive payment	Incentive compatible
New public good with voluntary payment	Not incentive compatible
Introduction of new private good	Not incentive compatible
Choice between two new public goods	Incentive compatible
Choice between an existing and an alternative private good	Incentive compatible, but choice does not reveal information about quantity demanded

Source: Adapted from Richard T. Carson, Theodore Groves, and Mark J. Machina, "Incentive and Informational Properties of Preference Questions," Plenary Address, European Association of Resource and Environmental Economists, Oslo, Norway, June 1999. Reprinted with authors' permission.

requirements for incentive compatibility. Respondents have no incentive to vote against their true preferences. Because CV is often the only way analysts can estimate WTP for public goods, this is reassuring. Incentive compatibility also holds for comparison between two mutually exclusive public goods with the same cost. As long as respondents place a positive value on both goods, they have no incentive to misstate their preferred good.

Incentive compatibility is lost in cases in which payment is voluntary because it introduces a second issue, whether to donate or to purchase, into the choice situation. In the case of a public good to be funded by contributions, respondents who place any positive value on the good have an incentive to accept bids above their WTPs to increase the chances that the good is provided because they do not actually have to make a donation in that amount.[82] Similarly, respondents who have any probability of actually wanting a new private (excludeable) good have an incentive to accept bids above their WTP to increase the chances that it will actually be provided because they can decline to purchase it if they decide that they do not want it. The problem is so severe for new private goods that dichotomous choice CV has largely been abandoned for this purpose.

Questions that ask respondents to choose between an existing private good and an alternative are incentive compatible about that choice as long as only potential users are surveyed. (The question will not be consequential for those who are not potential users.) Answers to the questions, however, do not reveal information about how much of the selected private good will actually be demanded. For example, a town may ask a sample of residents to choose between the existing skating rink, which is available only during daylight hours and has a small entrance fee, and a new skating rink with electric lighting, which would be available evenings and would have a higher entrance fee. The respondents have no incentive to misstate their true preferences over the choices, but the answers do not tell the town about how often residents will use the new facility.

Conclusion About the Importance of Strategic Responses The danger that strategic responses to CV questions will bias estimates of aggregate WTP cannot be assessed without considering both the elicitation method and the nature of the good being valued. The dichotomous choice method applied to the valuation of new public goods to

be funded by taxes or other coercive payments can be designed to avoid strategic response bias. It is subject to an upward bias, however, if payments are voluntary. The open-ended WTP method will generally be subject to strategic response bias, most likely leading to aggregate WTP estimates that are too small if payments are coerced and too large if payments are voluntary.

HOW ACCURATE IS CONTINGENT VALUATION?

It is possible to test the accuracy of CV WTP estimates in a number of ways.[83] The first method is to compare CV values to those generated by other indirect methods. CV values have been found to be approximately the same as those derived from travel cost studies.[84] They have also been found to be reasonably similar to prices derived from hedonic price regressions[85] and to the market prices of substitutes.[86] Wesley Magat and W. Kip Viscusi tested whether respondents' WTP for risk reductions associated with an insecticide and toilet bowl cleaner were consistent with standard economic theory or subject to the kinds of judgment biases described earlier. In general, they did not find strong evidence of bias.[87]

The second type of comparison, one that is more common and more appropriate, is between respondents' CV statements and their actual behavior when they participate in an experiment that utilizes a simulated or constructed market for the good in question.[88] Results from studies that have used experimental techniques to examine CV accuracy typically suggest that CV valuations of WTP relying on open-ended and dichotomous choice methods approximate actual market transactions, although there is some tendency for overvaluation.[89] In assessing these experiments, it is useful to keep in mind that the simulated market only approximates the workings of a real market; for example, there is only one opportunity to buy the good. In addition, the experiments have only been conducted in contexts in which respondents clearly derive use value from the good. Richard Bishop and Thomas Heberlein have conducted a number of experiments along these lines. The example presented in Exhibit 14-4 is typical of these experiments.

John Loomis has investigated how consistent household CV valuations are over time. Though not a direct test of accuracy, his investigation is relevant because consistency is a prerequisite for accuracy. He surveyed both visitors and the general public's WTP for recreational, option, existence, and bequest values derived from Mono Lake in California. Identical surveys were administered to the same individuals eight or nine months apart. The results were virtually identical.[90]

Though the evidence tends to suggest that CV is plausible in use contexts, the jury is still out in terms of nonuse values, given the ordering, embedding, noncommitment, and starting point problems discussed previously. Obviously, given the nature of nonuse values, it is much more difficult to elicit WTP from observed behavior. Nonetheless, with sufficient cleverness researchers may be able to find ways to do so. For example, as noted in Chapter 9, voluntary contributions to environmental causes might serve as the basis for estimation. Market-like experiments might also be used.[91] For example, individuals who have expressed WTP for nonuse could be given the option of returning none, part, or all of the checks sent to them by experimenters (as described in Exhibit 14-4), thus testing noncommitment bias. Successfully implementing such experiments would be extremely difficult with current levels of knowledge and typically available resources, however.

EXHIBIT 14-4

A wildlife area in Wisconsin is periodically opened to hunting. Hunters require a special permit to participate in these hunts. The permits are issued free to the winners of a lottery. For the 1984 lottery, a large number of applications were received. As a result of the lottery, 150 permits to hunt were issued.

To measure WTA, half of these hunters also received a letter explaining the study and a check made out in their name. To cash these checks, the hunters had to relinquish their permits. The denominations of the checks, which ranged randomly from $18 to $518, corresponded to the dichotomous choice method described earlier. The other half of the hunters received a similar letter with a hypothetical payment offer drawn from the same range as the first half. They were asked if they would have been willing to give up their permits for the hypothetical amounts stated in their letters.

To measure WTP, 150 unsuccessful lottery participants were selected at random. Again, 75 received a letter explaining that a hunting permit was available if they paid the amount specified in the letter. The amounts covered the same range as described previously. The other 75 were asked the same question hypothetically. That is, they were asked the CV question rather than given the opportunity to reveal their preferences through purchases.

Looking first at WTP, the results indicated that the CV valuations ($35) were only slightly higher than those measured through revealed preference ($31), a difference that was not statistically significant. These results suggest that noncommitment bias was not a major problem in this context. But, consistent with the findings discussed earlier, WTA valuations were much higher than WTP valuations. Furthermore, the CV valuations ($420) were considerably higher than those actually revealed by check cashing ($153).

Source: Adapted from Richard C. Bishop and Thomas A. Heberlein, "The Contingent Valuation Method," in Rebecca L. Johnson and Gary V. Johnson, eds., *Economic Valuation of Natural Resources: Issues, Theory, and Application* (Boulder, CO: Westview Press, 1990), 81–104.

The Potential for Calibration

It has long been recognized in marketing research that it may be necessary to *calibrate* respondents' valuations of goods in various ways to obtain more accurate estimates of their WTP.[92] As noted earlier, noncommitment bias, which may also be interpreted as strategic response bias, is a well-recognized problem in marketing research when individuals are asked to express their valuations of market goods. Kenneth Arrow has concluded, "A hypothesis to be explored is that, if the results of CVs for nonuse values were suitably calibrated, they would provide useful and reliable estimates."[93] Testing Arrow's hypothesis may be one of the most valuable areas of future research on CV.

The Accuracy of Different CV Methods and Proneness to Biases

Richard Bishop and Thomas Heberlein have conducted experiments comparing the size of valuations from different elicitation formats (again, this is in a use context). They conclude that:

> These experiments indicate that contingent values for willingness to pay may be somewhat high, but for open-ended and dichotomous choice questions the difference was not large enough to be statistically significant. . . . Bidding seems to introduce a substantial upward bias. . . . Contingent compensation demanded tended to produce excessive values when an open-ended question was asked and values that were biased substantially upward compared to values obtained from actual cash transactions when dichotomous choice questions were used.[94]

Other researchers have also addressed this issue. V. Kerry Smith and William Desvousges compared the valuations from several CV methods in their study of improvements in the quality of water in the Monongahela River. Their analysis suggests that starting point bias is important with iterative bidding: A low starting point ($25) generated lower valuations than other methods, while a high starting point ($125) produced valuations higher than other methods. Open-ended questions produced fairly similar valuations. Overall, mean valuations from different CV methods ranged from $7 to $36 for one water quality change, from $4 to $31 for another change, and from $11 to $51 for a third change.[95]

William Desvousges and his colleagues have compared open-ended question responses with dichotomous choice responses to the same questions. They found that in some contexts the differences in the mean WTP valuations were not statistically significant, but in other contexts they were. They also found the dichotomous choice method generated a considerable number of very high valuations.[96] This is consistent with the likely relative importance of strategic response bias discussed previously, as well as with the relatively lower precision of the dichotomous choice method relative to direct elicitation methods for given sample sizes.

HEURISTICS FOR THE DESIGN AND USE OF CV SURVEYS

Several experts in CV have suggested overall criteria for evaluating CV instruments. Ronald Cummings, David Brookshire, and William Schulze, for example, suggest five criteria for evaluating instruments. First, respondents should understand and be familiar with the good that is being valued. Second, respondents should have, or be given, experience in both valuation and the choice procedure. Third, there should be as little uncertainty as possible about the details of the project. All three of these concerns can best be addressed by attempting to reduce hypotheticality—for example, by employing quality ladders, by stressing realistically and concretely the substitution possibilities, and by presenting gains (benefits) in percentage terms as well as in absolute terms. Fourth, WTP rather than WTA should be used for valuation purposes. Fifth, attempts should be made to avoid anchoring and starting point bias.[97]

Heuristics for Using Estimates from Previous CV Studies

Because CV surveys are inevitably complex and expensive, many analysts are more concerned with using values derived from existing CV studies than they are in going out and doing CV surveys themselves. In an examination of alternative environmental regulatory policies for dealing with water pollution, Ralph Luken illustrates how

existing CV estimates can be "plugged in" to a CBA. Specifically, he extrapolates previous CV valuations of benefits from the Monongahela River studies discussed earlier in this chapter. His analysis provides a useful example of how to use plug-in values, and is described more fully in Chapter 15.[98] He is careful to specify which studies and which specific estimates he is using, the specific assumptions he makes in the extrapolations, the quality changes that are involved, the distinction between use and any nonuse components, and any potential remaining biases. He also performs sensitivity analysis.

Environment Canada, in cooperation with several international organizations, has developed a database of CV and other studies that value environmental impacts.[99] In order to facilitate the use of findings from these studies in CBA, it provides an abstract for each study describing the study area and its population characteristics, the environmental asset being valued, the research methods employed, and the monetary values that were estimated. Access to the online database is by subscription.[100]

CONCLUSION

Contingent valuation is now relatively uncontroversial in contexts involving use values, although it may overestimate them. Its accuracy in nonuse contexts is more controversial, yet it is in this context where its potential usefulness is likely to be the greatest. Doubts about the accuracy of CV in nonuse contexts stem from the problems of hypotheticality and the attendant judgment biases that appear to flourish in the nonuse context. Strategic response bias, on the other hand, does not appear to be a major problem in the use of the dichotomous choice method to value public goods with coercive payments, and neutrality bias can be minimized by use of appropriate survey techniques. Probably the most important topics for future CV research are the directional biases in particular methods and the development of techniques that allow reliable calibration.

EXERCISES FOR CHAPTER 14

1. The construction of a dam that would provide hydroelectric power would result in the loss of two streams: one that is now used for sport fishing, and another that does not support game fish but is part of a wilderness area.
   a. Imagine that a contingent valuation method is used to estimate the social cost of the loss of each of these streams. Would you be equally confident in the two sets of estimates?
   b. Consider two general approaches to asking contingent valuation questions about the streams. The first approach attempts to elicit how much compensation people would require to give up the streams. The second approach attempts to elicit how much people would be willing to pay to keep the streams. Which approach would you recommend? Why?
2. A number of residents of Dullsville have complained to the mayor that the center of town looks shabby compared to the centers of many other nearby towns. At the mayor's request, the Parks Department has put together a proposal for converting the town square parking lot into a sitting park with flower displays—it modeled the design on a similar park in the neighboring town of Flowerville. The annualized cost of installing and maintaining the park,

and relocating parking to nearby Short Street, would be about $120,000. With about 40,000 households paying property taxes, the project would cost an average household about $3 per year.

You have been asked to give advice about conducting a survey to measure the benefits of the project.

a. The Parks Department proposes conducting a telephone survey. Does this seem like an appropriate survey vehicle?
b. How might a random sample be drawn for a telephone survey?
c. Write a statement that could be read by the interviewer to describe the project.
d. Write questions to implement the open-ended WTP method.
e. Propose a procedure for implementing the dichotomous choice method.

3. Consider a project that would involve purchasing marginal farmland that would then be allowed to return to wetlands capable of supporting migrant birds. Researchers designed a survey to implement the dichotomous choice method. They reported the following data.

Stated Price (annual payment in dollars)	*Fraction of Respondents Accepting Stated Price (percent)*
0	98
5	91
10	82
15	66
20	48
25	32
30	20
35	12
40	6
45	4
50	2

What is the mean willingness-to-pay for the sampled population?

NOTES

1. Most observers attribute the origins of CV to Robert K. David, "Recreation Planning as an Economic Problem," *Natural Resources Journal* 3(2) 1963, 239–249. Early applications include Ronald G. Ridker, *Economic Costs of Air Pollution: Studies in Measurement* (New York: Praeger, 1967); and Charles J. Cicchetti and V. Kerry Smith, "Congestion, Quality Deterioration, and Optimal Use: Wilderness Recreation in the Spanish Peaks Primitive Area," *Social Science Research* 2(1) 1973, 15–30.
2. The two "classic" CV overviews are Ronald G. Cummings, David S. Brookshire, and William D. Schulze, *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method* (Totowa, NJ: Rowman & Allanheld, 1986); and Robert C. Mitchell and Richard T. Carson, *Using Surveys to Value Public Goods: The Contingent Valuation Method* (Washington, D.C.: Resources for the Future, 1989). Some overviews include Richard C. Bishop and Thomas A. Heberlein, "The Contingent Valuation Method," in Rebecca L. Johnson and Gary V. Johnson, eds., *Economic Valuation of Natural Resources: Issues, Theory, and Application* (Boulder, CO: Westview Press, 1990), 81–104; Ian J. Bateman and Kenneth

G. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the U.S., EU, and Developing Countries* (Oxford and New York: Oxford University Press, 1999); Timothy C. Haab and Kenneth E. McConnell, *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation* (Cheltenham, UK: Edward Elgar Publishers, 2003).

3. William H. Desvousges, V. Kerry Smith, and Ann Fisher, "Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River," *Journal of Environmental Economics and Management* 14(3) 1987, 248–267; and Richard C. Bishop and Thomas A. Heberlein, "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61(5) 1979, 926–930.
4. Alan Diener, Bernie O'Brien, and Amiram Gafni, "Health Care Contingent Valuation Studies: A Review and Classification of the Literature," *Health Economics* 7(4) 1998, 313–326.
5. For example, see Gretchen J. Burger, Application and Assessment of the Contingent Valuation Method for Federal Hazardous Waste Policy in the Washington, D.C. Area, Ph.D. dissertation, University of New Mexico, Albuquerque, 1984; David Maddison and S. Mourato, "Valuing Different Road Options for Stonehenge," *Conservation and Management of Archeological Sites* 4(4) 2001, 203–212; Peter C. Boxall, Jeffrey Englin, and Wiktor L. Adamowicz, "Valuing Aboriginal Artifacts: A Combined Revealed-Stated Preference Approach," *Journal of Environmental Economics and Management* 45(2) 2003, 213–230; Richard T. Woodward and Yong-Suhk Wui, "The Economic Value of Wetland Services: A Meta-Analysis," *Ecological Economics* 37(2) 2001, 257–270; and Bruce K. Johnson and John C. Whitehead, "Value of Public Goods from Sports Stadiums: The CVM Approach," *Contemporary Economic Policy* 18(1) 2000, 48–58.
6. Raymond J. Kopp, Paul R. Portney, and V. Kerry Smith, "The Economics of Natural Resource Damages after *Ohio v. U.S. Dept. of the Interior*," *Environmental Law Reporter* 20(4) 1990, 10,127–10,131.
7. Kenneth Arrow, Robert Solow, Paul Portney, Edward Leamer, Roy Radner, and Howard Schuman, "Report of the NOAA Panel on Contingent Valuation," *Federal Register* 58(10) January 15, 1993, 4601–4614.
8. On this point, see Daniel McFadden and Gregory Leonard, "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis," in J. A. Hausman, ed., *Contingent Valuation: A Critical Assessment* (New York: North-Holland, 1993), 165–208.
9. Arrow et al., "Report of the NOAA Panel on Continent Valuation," p. 4608.
10. For an example of this approach, see J. C. Horvath, *Southeastern Economic Survey of Wildlife Recreation* (Atlanta, GA: Environmental Research Group, Georgia State University, 1974).
11. One compromise between the open-ended format and the provision of "guidance" that has been used in in-person interviews is the *payment card* with a range of prices. For example, in the Nestucca oil spill CV survey, respondents were asked to specify the maximum amount their households would pay for particular programs to prevent oil spills. They were then shown a payment card with dollar amounts ranging from $0 to $5,000 and asked to circle the amount closest to their evaluation. Robert Rowe, William D. Schulze, W. Douglas Shaw, David Schenk and Lauraine G. Chestnut, "Contingent Valuation of Natural Resource Damage Due to Nestucca Oil Spill," prepared for the Department of Wildlife, Ministry of the Environment, BC, and Environment Canada, by RCG/Hagler, Bailly Inc., June 1991.
12. For an example, see David Brookshire, Berry Ives, and William D. Schulze, "The Valuation of Aesthetic Preference," *Journal of Environmental Economics and Management* 3(4) 1976, 325–346.
13. James L. Regens, "Measuring Environmental Benefits with Contingent Markets," *Public Administration Review* 51(4) 1991, 345–352.
14. For an example, see V. Kerry Smith and William H. Desvousges, *Measuring Water Quality Benefits* (Boston, MA: Kluwer Nijhoff Publishing, 1986), Chapter 6.
15. Baruch Fischhoff and Louis A. Cox Jr., "Conceptual Framework for Regulatory Benefits Assessment," in Judith D. Bentkover, Vincent T. Covello, and Jeryl Mumpower, eds., *Benefits Assessment: The State of the Art* (Boston, MA: D. Reidel Publishing Co., 1986), 51–84.

16. For an example using this method, see John Loomis, *Integrated Public Lands Management* (New York: Columbia University Press, 1993), 276–279.
17. Selecting an appropriate set of bid prices is one of the most difficult tasks analysts face in applying the dichotomous choice method. The bid prices must be spaced sufficiently far apart so that they show a range of acceptance rates, ideally from about .85 to about .15. Simply spreading bid prices over a very wide range, however, is unsatisfactory because, for a given sample size, statistical efficiency in estimating willingness-to-pay from the pattern of acceptances is greater for fewer bid prices. These considerations suggest the importance of conducting pre-test surveys to determine an appropriate set of bid prices. (Of course, in any survey study, pre-tests are important for ensuring that respondents understand questions.) See Hans Nyquist, "Optimal Design of Discrete Response Experiments in Contingent Valuation," *Review of Economics and Statistics* 74(3) 1992, 559–563; and Barbara J. Kanninen, "Bias in Discrete Response Contingent Valuation," *Journal of Environmental Economics and Management* 28(1) 1995, 114–125.
18. For more detail on this method, see Bishop and Heberlein, "Measuring Values of Extra-Market Goods," 89–91. On the technical issues relating to interpreting the function as utility and how to deal with truncation, see W. Michael Hanemann, "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American Journal of Agricultural Economics*, 66(3) 1984, 332–341.
19. The formula for computing the WTP of the average member of the sample can be derived more formally if some additional notation is introduced. (Note that this derivation involves approximating the expected value of WTP.) Assume that the offer prices are $0, v, 2v, 3v, \ldots Nv$, where v is the interval between prices, and Nv is the maximum price offered. For example, if the offer prices were 0, \$10, \$20, \$30, . . . , \$200, then v = \$10 and N = 20. Let $\mathbf{F}[X]$ be the fraction of respondents offered price X who accept the offer—this is the height of the bar in the histogram that sits above X. In order to calculate the expected value of WTP, we require the probability that each offer price is the *maximum* amount that a respondent would be willing to pay. We can approximate this for offer amount $\$kv$ with the expression: $\mathbf{F}[kv] - \mathbf{F}[(k+1)v]$ for $k = 0, 1, \ldots, N-1$ which is roughly the probability of a respondent accepting $\$kv$ minus the probability of accepting a larger offer price, $\$(k+1)v$. The mean WTP (that is, the WTP of a randomly selected respondent) is then approximately the sum of the products of maximum payments times their probabilities:

$$\mathbf{E}[\text{WTP}] = Nv\mathbf{F}[Nv] + (N-1)v\{\mathbf{F}[(N-1)v] - \mathbf{F}[Nv]\} + (N-2)v\{\mathbf{F}[(N-2)v] - \mathbf{F}[(N-1)]\} + \ldots + 0\{\mathbf{F}[0] - \mathbf{F}[v]\}$$

Collecting terms yields the simple expression:

$$\mathbf{E}[\text{WPT}] = v\sum_{k=0}^{N}\mathbf{F}[kv]$$

20. For a thorough treatment of estimation issues, see W. Michael Hanemann and Barbara Kanninen, "The Statistical Analysis of Discrete Response CV Data," in Ian J. Bateman and Ken G. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the U.S., EC, and Developing Countries* (Oxford and New York: Oxford University Press, 1999), 302–441.
21. A complication arises because the statistical model usually assumes an infinite range for x. In practice, surveys use offer prices with a finite range, usually bounded from below at zero. Averaging over this finite range will miss some of the probability assumed by the statistical model. A correction to help account for this truncation bias involves rescaling the distribution to account for the missing areas in the tails of the distribution. See Kevin J. Boyle, Michael P. Welsh, and Richard C. Bishop, "Validation of Empirical Measures of Welfare Change: Comment," *Land Economics* 64(1) 1988, 94–98.
22. The area under the curve can be estimated through numerical integration. The procedure involves approximating the area with rectangles through the following steps. First, divide the range of X into equal segments of width n. Second, calculate the probability of acceptance at each of these points. Third, find the average acceptance value for adjacent pairs of points. Fourth, multiply each of these averages by n.

Fifth, sum all these products to get the estimate of the area. Any desired degree of accuracy can be obtained by choosing a sufficiently small value of *n*.

23. For an overview of the statistical issues in estimating aggregate WTP from such models, see Timothy Park, John B. Loomis, and Michael Creel, "Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies," *Land Economics* 67(1) 1991, 64–73.
24. Joseph A. Herriges and Jason F. Shogren, "Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning," *Journal of Environmental Economics and Management* 30(1) 1996, 112–131.
25. Richard T. Carson, Theodore Groves, and Mark J. Machina, "Incentive and Informational Properties of Preference Questions," Plenary Address, European Association of Resource and Environmental Economists, Oslo, Norway, June 1999.
26. Empirical work provides support for this weighted averaging explanation; see Anthony C. Burton, Katherine S. Carson, Susan M. Chilton, and W. George Hutchinson, "An Experimental Investigation of Explanations for Inconsistencies in Responses to Second Offers in Double Referenda," *Journal of Environmental Economics and Management* 46(3) 1996, 472–489.
27. Robert J. Johnston, Stephen K. Swallow, and Thomas F. Weaver, "Estimating Willingness to Pay and Resource Tradeoffs with Different Payment Mechanisms: An Evaluation of a Funding Guarantee for Watershed Management," *Journal of Environmental Economics and Management* 38(1) 1999, 97–120.
28. See "The Review Panel's Assessment," 180–204 and Kenneth Arrow, "Comments," 180–185, in Cummings, Brookshire, and Schulze, *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.*
29. See Robert P. Berrens, Alok K. Bohara, Hank Jenkins-Smith, Carol Silva, and David L. Weimer, "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples," *Political Analysis* 11(1) 2003, 1–22; and Robert P. Berrens, Alok K. Bohara, Hank C. Jenkins-Smith, Carol L. Silva, and David L. Weimer, "Information and Effort in Contingent Valuation Surveys: Application to Global Climate Change Using National Internet Samples," *Journal of Environmental Economics and Management* 47(2) 2004, 331–363.
30. For overviews, see Martin Frankel, "Sampling Theory," 21–67, and Seymour Sudman, "Applied Sampling," 145–194, in Peter H. Rossi, James D. Wright, and Andy B. Anderson, *Handbook of Survey Research.* For issues specifically relevant to CV, see Steven F. Edwards and Glen D. Anderson, "Overlooked Biases in Contingent Valuation Surveys: Some Considerations," *Land Economics* 63(2) 1987, 168–178.
31. Richard T. Carson, "Constructed Markets," in John B. Braden and Charles D. Kolstad, eds., *Measuring the Demand for Environmental Quality* (New York: Elsevier Science Publishers, 1991), 152.
32. For an example of an attempt to estimate the extent of a recreational market, see V. Kerry Smith and Raymond J. Kopp, "The Spatial Limits of the Travel Cost Recreational Demand Model," *Land Economics* 56(1) 1980, 64–72.
33. Desvousges, Smith, and Fisher, "Option Price Estimates for Water Quality Improvements."
34. F. L. Filion, "Estimating Bias Due to Nonresponse in Mail Surveys," *Public Opinion Quarterly* 39(4) 1975–1976, 482–492.
35. For a review, see John B. Loomis, "Expanding Contingent Value Sample Estimates to Aggregate Benefit Estimates: Current Practices and Proposed Solutions," *Land Economics* 63(4) 1987, 396–402.
36. The most comprehensive criticisms of CV can be found in Hausman, ed., *Contingent Valuation: A Critical Assessment.*
37. William D. Schulze, Ronald G. Cummings, and David S. Brookshire, *Methods Development in Measuring Benefits of Environmental Improvements, Vol. II* (Washington, D.C.: Report to U.S. EPA, 1983).
38. For an extensive discussion of the attitude-behavior nexus, see Mitchell and Carson, *Using Surveys to Value Public Goods: The Contingent Valuation Method,* 175–187.
39. Baruch Fischhoff and Lita Furey, "Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility," *Journal of Risk and Uncertainty* 1(2) 1988, 147–184, at 179–180.
40. In marketing, the technique called *conjoint analysis* is used to elicit consumer valuations for

new, unfamiliar goods. For a review, see Daniel McFadden, "The Choice Theory Approach to Market Research," *Marketing Science* 5(4) 1986, 275–297.

41. Richard T. Carson, Nicholas E. Flores, and Norman F. Meade, "Contingent Valuation: Controversies and Evidence," *Environmental and Resource Economics* 19(2) 2001, 173–210, at 178.
42. For a discussion of this point, see Cummings, Brookshire, and Schulze, *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method,* at 43, 253.
43. Fischhoff and Furey, "Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility," Table 1, 156.
44. Daniel A. Hagen, James W. Vincent, and Patrick G. Welle, "The Benefits of Preserving Old-Growth Forests and the Northern Spotted Owl," *Contemporary Policy Issues* 10(2) 1992, 13–16. A second CV study of the spotted owl that can be critiqued on neutrality grounds is Jonathan Rubin, Gloria Heland, and John Loomis, "A Benefit-Cost Analysis of the Northern Spotted Owl: Results from a Contingent Valuation Survey," *Journal of Forestry* 89(12) 1991, 25–30.
45. William McKillop, "Use of Contingent Valuation in Northern Spotted Owls Studies: A Critique," *Journal of Forestry* 90(8) 1992, 36–37.
46. Ed Bukszar, "Does Overconfidence Lead to Poor Decisions? A Comparison of Decision Making and Judgment Under Uncertainty," paper presented at the Annual Conference of Judgment and Decision Making, Washington, D.C., November 7, 1993.
47. See David L. Weimer and Aidan R. Vining, *Policy Analysis: Concepts and Practice,* 4th ed. (Upper Saddle River, NJ: Pearson Prentice Hall, 2005), 121–122.
48. For a review of this issue, see Colin F. Camerer and Howard Kunreuther, "Decision Processes for Low Probability Events: Policy Implications," *Journal of Policy Analysis and Management* 8(4) 1989, 565–592.
49. For an extensive review of biases and the various theories behind them, see Camerer and Kunreuther, "Decision Processes for Low Probability Events: Policy Implications."
50. Daniel Kahneman and Amos Tversky, "Prospect Theory," *Econometrica* 47(2) 1979, 263–292. Many of the violations of expected utility hypothesis that can be explained by prospect theory can also be explained by assuming people maximize expected utility but employ Bayesian updating of probabilities; see W. Kip Viscusi, "Prospective Reference Theory: Toward an Explanation of the Paradoxes," *Journal of Risk and Uncertainty* 2(3) 1989, 235–263.
51. Richard H. Thaler, "Towards a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization* 1(1) 1980, 39–60. For some empirical evidence on loss aversion in housing markets, see David Genesove and Christopher Mayer, "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics* 116(4) 2001, 1233–1260.
52. It may seem somewhat artificial to separate hypotheticality from noncommitment bias. However, as emphasized earlier, hypotheticality does not appear to generate upward or downward bias, whereas, as we will see, noncommitment does appear to bias valuations upward.
53. See, for example, Linda F. Jamieson and Frank M. Bass, "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods," *Journal of Marketing Research* 26(3) 1989, 336–345; and F. Thomas Juster, *Anticipations and Purchases* (Princeton, NJ: Princeton University Press, 1964).
54. Helen R. Neill, Ronald G. Cummings, Philip T. Ganderton, Glenn W. Harrison, and Thomas McGuckin, "Hypothetical Surveys and Economic Commitments," *Land Economics* 70(2) 1994, 145–154; and Kalle Seip and Jon Strand, "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment," *Environmental and Resource Economics* 2(1) 1992, 91–106.
55. For more detail, see Michael A. Kemp and Christopher Maxwell, "Exploring a Budget Context for Contingent Valuation Estimates," 217–265, in Hausman, ed., *Contingent Valuation: A Critical Assessment.*
56. I. J. Bateman and I. H. Langford, "Budget-Constrained, Temporal, and Question-Ordering Effects in Contingent Valuation Studies," *Environment and Planning A* 29(7) 1997, 1215–1228.
57. George Tolley and Alan Randall, "Establishing and Valuing the Effects of Improved Visibility in the Eastern United States," Report to the U.S. EPA, Washington, D.C., 1983.

58. Karl C. Samples and James R. Hollyer, "Contingent Valuation of Wildlife Resources in the Presence of Substitutes and Complements," in Rebecca L. Johnson and Gary V. Johnson, eds., *Economic Valuation of Natural Resources: Issues, Theory, and Application,* (Boulder, CO: Westview Press, 1990), 177–192.
59. See Peter A. Diamond, Jerry A. Hausman, Gregory K. Leonard, and Mike Denning, "Does Contingent Valuation Measure Preferences? Experimental Evidence," in Hausman, ed., *Contingent Valuation: A Critical Assessment,* 43–85.
60. See the heated argument between Peter Diamond and Richard Carson, in Hausman, ed., *Contingent Valuation: A Critical Assessment,* 87–89.
61. The case that these effects can adequately explain the empirical findings is made in Carson, Flores, and Meade, "Contingent Valuation: Controversies and Evidence," 186–189; see also Richard T. Carson, Nicholas E. Flores, and Michael Hanemann, "Sequencing and Valuing Public Goods," *Journal of Environmental Economics and Management* 36(3) 1998, 314–323.
62. William H. Desvousges, F. Reed Johnson, Richard W. Dunford, Sara P. Hudson, and K. Nicole Wilson, "Measuring Natural Resources Damages with Contingent Valuation: Tests of Validity and Reliability," in Hausman, ed., *Contingent Valuation: A Critical Assessment,* 91–159.
63. Daniel Kahneman and Jack L. Knetsch, "Valuing Public Goods: The Purchase of Moral Satisfaction," *Journal of Environmental Economics and Management* 22(1) 1992, 57–70, at 64–68.
64. Carson, Flores, and Meade, "Contingent Valuation: Controversies and Evidence," *Environmental and Resource Economics.*
65. Kevin J. Boyle, Richard C. Bishop, and Michael P. Walsh, "Starting Point Bias in Contingent Valuation Bidding Games," *Land Economics* 16(2) 1985, 188–194.
66. Regens, "Measuring Environmental Benefits with Contingent Markets," 347–348.
67. Walter J. Mead, "Review and Analysis of State-of-the-Art Contingent Valuation Studies," in Hausman, ed., *Contingent Valuation: A Critical Assessment,* 305–332.
68. See Trudy A. Cameron and D. D. Huppert, "Referendum Contingent Valuation Estimates: Sensitivity to the Assignment of Offered Values," *Journal of the American Statistical Association* 86(416) 1991, 910–918.
69. Alok K. Bohara, Michael McKee, Robert P. Berrens, Hank Jenkins-Smith, Carol L. Silva, and David S. Brookshire, "Effects of Total Cost and Group-Size Information on Willingness to Pay Responses: Open Ended vs. Dichotomous Choice," *Journal of Environmental Economics and Management* 35(2) 1998, 142–163.
70. But Michael Hanemann ("Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 81[3] 1991, 635–647) points out that there is no reason why WTA should not be considerably larger than WTP in certain circumstances. For example, one cannot pay more than one's income year after year to save one's own life, but there may be no finite offer that could make one willingly give up one's life.
71. Jack Knetsch and J. S. Sinden, "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value," *Quarterly Journal of Economics* 99(3) 1984, 507–521; Shelby Gerking, Menno De Haan, and William Schulze, "The Marginal Value of Job Safety: A Contingent Valuation Study," *Journal of Risk and Uncertainty,* 1(2) 1988, 185–199; and Wesley Magat and W. Kip Viscusi, *Informational Approaches to Regulation* (Cambridge, MA.: MIT Press, 1992).
72. Timothy McDaniels, "Reference Points, Loss Aversion, and Contingent Values for Auto Safety," *Journal of Risk and Uncertainty* 5(2) 1992, 187–200.
73. Don L. Coursey, John J. Hovis, and William D. Schulze, "The Disparity between Willingness to Accept and Willingness to Pay Measures of Value," *Quarterly Journal of Economics,* 102(3) 1987, 679–690; Jason F. Shogren, Seung Y. Shin, Dermot J. Hayes, and James B. Kliebenstein, "Resolving the Differences in Willingness to Pay and Willingness to Accept," *American Economic Review* 84(1) 1994, 255–270; and Wiktor L. Adamovicz, Vinay Bhardwaj, and Bruce McNab, "Experiments on the Difference Between Willingness to Pay and Willingness to Accept," *Land Economics* 64(4) 1993, 416–427.

74. Shogren, Shin, Hayes, and Kliebenstein, "Resolving the Differences in Willingness to Pay and Willingness to Accept."
75. Arrow et al., "Report of the NOAA Panel on Continent Valuation," 4608.
76. On free riding in the provision of public goods, see R. Mark Isaac, James Walker, and Susan Thomas, "Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations," *Public Choice* 43(2) 1984, 113–149; Oliver Kim and Mark Walker, "The Free Rider Problem: Experimental Evidence," *Public Choice* 43(1) 1984, 3–24; G. Marwell and R. Ames, "Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods," *Journal of Public Economics* 15(3) 1981, 295–310; and Linda Goetz, T. F. Glover, and B. Biswas, "The Effects of Group Size and Income on Contributions to the Corporation for Public Broadcasting," *Public Choice* 77(2) 1993, 407–414.
77. Mitchell and Carson, *Using Surveys to Value Public Goods: The Contingent Valuation Method.*
78. Carson, Groves, and Machina, "Incentive and Informational Properties of Preference Questions."
79. On the domain of preferences, see the discussion of Arrow's Possibility Theorem in Chapter 2.
80. Alan Gibbard, "Manipulation of Voting Schemes: A General Result," *Econometrica* 41(4) 1973, 587–601; and Mark A. Satterthwaite, "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory* 10(2) 1975, 187–217.
81. See Kevin J. Boyle, F. Reed Johnson, Daniel W. McCollum, William H. Desvousges, Richard W. Dunford, and Sara P. Hudson, "Valuing Public Goods: Discrete versus Continuous Contingent Valuation Responses," *Land Economics* 72(3) 1996, 381–396.
82. A number of studies have found a large divergence between WTP as expressed in contingent choice CV and actual donations. For a review, including a discussion of using follow-up questions on certainty of donation to adjust contingent donations, see Patricia A. Champ, Richard C. Bishop, Thomas C. Brown, and Daniel W. McCollum, "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods," *Journal of Environmental Economics and Management* 33(2) 1997, 151–162.
83. For an overview of this issue and a listing of relevant studies, see V. K. Smith, "Nonmarket Valuation of Environmental Resources: An Interpretive Appraisal," *Land Economics* 69(1) 1993, 1–26, at 8–14, and Table 1.
84. See William K. Desvousges, V. Kerry Smith, and Matthew P. McGivney, "A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements," Report to U.S. EPA, Research Triangle Institute, 1983; and Christine Sellar, John R. Stoll, and Jean-Paul Chavas, "Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques," *Land Economics* 61(2) 1985, 156–175. Smith and Desvousges, *Measuring Water Quality Benefits,* compare CV to "simple" travel cost methods.
85. David S. Brookshire, Mark A. Thayer, William D. Schulze, and Ralph C. d'Arge, "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches," *American Economic Review* 72(1) 1982, 165–177.
86. Mark A. Thayer, "Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence," *Journal of Environmental Economics and Management* 8(1) 1981, 27–44.
87. Magat and Viscusi, *Informational Approaches to Regulation,* Chapter 7.
88. For a listing of such studies, see Carson, "Constructed Markets," at 121–126.
89. Coursey, Hovis, and Schulze, "The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value"; Mark Dickie, Ann Fisher, and Shelby Gerking, "Market Transactions and Hypothetical Demand Data: A Comparative Study," *Journal of the American Statistical Society* 82(398) 1987, 69–75.
90. John Loomis, "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses," *American Journal of Agricultural Economics* 71(1) 1989, 76–84.
91. For a discussion of this possibility, see Douglas M. Larson, "On Measuring Existence Value," *Land Economics* 69(3) 1993, 377–388.
92. An example of such calibration is Donald G. Morrison, "Purchase Intentions and Purchase Behavior," *Journal of Marketing,* 43(2), 1979,

65–74. "It is possible to improve prediction accuracy by measuring and using perceptions that affect and modify the relationship between stated intentions and trial purchase for new products," in Jamieson and Bass, "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products," 344.

93. Kenneth Arrow, "Contingent Valuation of Nonuse Values: Observations and Questions," in Hausman, ed., *Contingent Valuation: A Critical Assessment,* 479–483.
94. Bishop and Heberlein, "The Contingent Valuation Method," 97.
95. Smith and Desvousges, *Measuring Water Quality Benefits,* Table 105, 271.
96. Desvousges, Johnson, Dunford, Hudson, and Wilson, "Measuring Natural Resources Damages with Contingent Valuation: Tests of Validity and Reliability."
97. Summarized from "Comparison Studies: What Is Accuracy," 71–109, and "reference operating conditions," Table 13-1, p. 230, in "Summary and Conclusions," 205–236 in Cummings, Brookshire, and Schulze, *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.*
98. Ralph A. Luken, *Efficiency in Environmental Regulation: A Benefit-Cost Analysis of Alternative Approaches* (Boston, MA: Kluwer Academic Publishers, 1990), 45–53.
99. Richard T. Carson, "Contingent Valuation: A User's Guide," *Environmental Science and Technology* 34(8) 2000, 1413–1418.
100. The database address is *www.evri.ec.gc.ca/evri/.*