CHAPTER 3

LITERATURE REVIEW: PART 2 - VALUATION METHODOLOGY

How does a society determine which public goods to provide and the appropriate level at which to provide them? Once a public good has been provided, one cannot exclude others from using it. For example, improvements in environmental quality and the health benefits stemming from a cleaner environment are essentially public goods; the benefits of cleaner water or air in a particular area typically accrue not only to the individuals that pay the cost of the environmental improvements, but also to all of the individuals that spend a significant amount of time in that area. In a society dominated by the economic system of capitalism, the quantity produced and sold of a market or private good, and its price, are determined by the interaction of supply and demand; by the interactions between those who wish to enjoy the good and those who wish to profit from producing it. Market equilibrium price and quantity are reached in part because the benefits derived from a market good are enjoyed solely by the consumer that paid for the good. The owner of the good has the right to exclude others from using and enjoying it; there are no external benefits provided to others.

Public goods pose a unique problem because of the external benefits associated with their provision and mankind's proclivity to "free-ride". A "free-rider" is an individual who pays less for an additional unit of a good than the true value he places on the additional unit; the free-rider pays less for the good than his marginal willingness-to-pay. The "free-rider" is keenly aware that once the public good has been provided, he can enjoy its benefits. Thus, he is reluctant to contribute to the cost of its provision out of a desire that others will provide the good and the knowledge that, if they do so, he will be

able to enjoy its external benefits at little or no cost to himself. Consequently, private firms would be unable to collect enough revenue to cover their costs if they were to supply public goods. Thus, the task of providing the public goods that society desires, such as environmental quality and national defense, usually falls to a collective body, such as a government, which has the power to collect revenue through taxation (Field, 1994).

Benefit-cost analysis is a tool used by decision makers in the public sector; it seeks to determine whether or not the benefits of a proposed project exceed its costs. The Reagan Administration issued an Executive Order in 1981, mandating that benefit-cost analyses be performed for all major federal regulatory programs. Benefit-cost analysis is also now required by many states. As economists seek to refine its use, two questions must be answered: "what is the appropriate methodology for measuring the benefits associated with non-market goods?" and "how should benefits be discounted?" These questions lie at the heart of this study.

3.1 The Valuation Methodology Debate

The valuation of health benefits is a crucial aspect of designing environmental policy. Without knowledge of the value members of our society place on alternative health outcomes, it is impossible for regulatory agencies to design effective policies that accurately reflect the desires of society. For example, should policy focus on the problem of ozone pollution exposure (thereby decreasing the risk of chronic bronchitis) or concentrate its efforts on safer drinking water (thereby decreasing the risk of neurological problems due to lead poisoning)? In a world of limited resources, it is only by being

aware of the relative values individuals place on these health benefits that we may hope to answer these questions successfully.

When confronted with the task of placing a monetary value on a non-market good, such as environmental quality or human health, economists have two types of tools to call upon: indirect or revealed preference methods, and direct or stated preference methods. Revealed preference methods derive the individual's monetary value for a particular non-market good by observing his economic behavior in actual markets related to the non-market good in question. These methods include hedonic pricing, travel cost and averting cost analyses. For example, to arrive at an estimate of a particular individual's willingness-to-pay (WTP) for groundwater protection, one might employ averting cost analysis and examine the amount of money, if any, the individual has spent on bottled water and home filtration devices in attempt to avoid potential contamination. Stated preference methods differ from revealed preference methods in that no actual market behavior is observed in the estimation of the individual's economic value for a specific non-market good. The individual is asked, in a hypothetical situation such as survey or experimental context, to indicate her WTP for a specific good. For example, individuals may be asked what they would be willing to pay for a given reduction in a specific health risk, as if the market for such reduction in risk actually existed. The contingent valuation method and conjoint analysis are the most commonly used stated preference methods. Revealed preference and stated preference methods each have their strengths and weaknesses. Let us compare the applicability and drawbacks to each type of valuation tool, beginning with the indirect techniques.

3.2 Indirect or Revealed Preference Methods

3.2.1 The Hedonic Pricing Technique

Estimation of the value of a non-market good through the hedonic pricing method is based upon the assumption that the price of a related market good is a function of its attributes. For example, consider market good X, which has n attributes, $Q_1 \ldots Q_n$. The price of good X, P_X, may be thought of as a function of both the presence of good X's attributes and the levels of those attributes:

$$P_x = P(Q_1, \ldots, Q_n) \tag{3.1}$$

The implicit price of a particular attribute, Q_i, is derived by differentiating the price of good X with respect to attribute of interest, i.e. $\partial P_X / \partial Q_i$. This technique is often used in property value studies, where the price of a particular unit of housing is expected to include the value of environmental amenities, such as an ocean view or exceptionally clean air. The other common application of the hedonic pricing technique is wage studies, where the wage that an employee earns is expected to reflect, in part, the risk levels that she faces. Estimates of the implicit value of life using labor market data range from $580,000 (Thaler and Rosen, 1976) to $7-$11 million (Viscusi, 1981). The largest problem that hedonic analysis faces is the requirement that the researcher be able to control for all relevant attributes. This will likely be a difficult, if not impossible, task. For example, in a property value study one would need to control for the characteristics of the structure itself, the safety of the neighborhood, environmental amenities, the quality of the local public school system, and the proximity to highways, to name a few. Other problems with hedonic property value and wage studies include, but are not limited to, the need for the individuals in question to be aware of the differences in attributes

between potential housing units and jobs, insufficient market data, and unknown functional forms of the pricing equations (Mitchell and Carson, 1989).

3.2.2 Travel Cost Analysis

Travel cost analysis is typically used to value benefits associated with access to recreation sites. For example, to value the benefits associated with Yellowstone National Park, one could examine the average number of per capita visits by U.S. residents. One could then group such residents into zones based upon distance from the park, and assign a monetary value to each zone, representing the amount of money one would need to spend to travel from the zone to the park (e.g. expenditures on plane tickets, gasoline, tolls, etc). Criticisms of the travel cost method include the possibility that individuals can substitute one recreation site for another, the difficulty of defining relevant costs, the difficulty in incorporating environmental quality more explicitly into the model, and the question of how to value the time one may have spent in travel and multipurpose trips (Mitchell and Carson, 1989). Travel cost estimation has served as a useful tool for measuring the benefits associated with recreational and scenic areas, but it is more difficult to use travel cost analysis to value many important classes of public goods, such as benefits stemming from regulation directed at reducing air and water pollution.

3.2.3 Averting Cost Analysis

Averting behavior analysis stems from the ability of consumer to purchase goods in traditional economic markets that can be used to mitigate adverse health and environmental effects. However, averting behavior analysis suffers from the possibility that averting expenditures may represent spending not only on the specific good in

question, but also on some other good; this is referred to as the "joint products" problem. For example, an individual may purchase bottled water not only in an effort to protect himself from contamination, but also because he believes that bottled water tastes better. The asthmatic who purchases an air conditioner likely purchases it not only to alleviate symptoms of her disease that are aggravated by air pollution, but also because she enjoys cooler air in her home during the summer months. In addition, averting expenditures should not include sunk costs associated with the purchase of durable goods, since typically the full price of a durable good cannot be characterized as an expenditure for a change in environmental quality (Bartik, 1988). There is strong evidence that averting behavior expenditures are a lower bound to compensating and equivalent variation (Roach, 1990; Shortle and Roach, 1989), and will significantly underestimate true willingness-to-pay even if the value of time that may be lost due to averting behavior is included in the analysis (for example, see Akerman, Johnson, and Bergman, 1991).

3.2.4 The Importance of Non-Use Value: Why Revealed Preference Methods Sometimes Fall Short

A major flaw that is common to all revealed preference methods is that these methods cannot measure non-use values. The term "non-use value" refers to the value that an individual may place on a good independent of his personal use of the good. The recognition that non-use values may be a significant component of total economic value stems from a 1967 article by John Krutilla. Non-use values include social use, existence, bequest, and option values. Social use value is value that an individual derives from knowing that persons other than himself will be able to use the good in question once it has been provided. For example, one may be willing to pay something to protect the

drinking water of another household, even if his household does not benefit directly from that protection. The value that one may derive simply from knowing that a good exists, independent of its use to humans, is referred to as existence value. For example, one may be willing to make a monetary contribution to an African wildlife preserve even if she knows she is highly unlikely to ever personally visit the preserve. One may also be willing to pay for provision of a non-market good if he wishes to provide the good not for his own enjoyment, but for the enjoyment of his loved ones. For example, an eighty-five year old grandfather may be willing to pay for some form of groundwater protection not so much out of concern for his health in the remaining years of his life, but out of deep concern for his children and grandchildren's health. Finally, willingness-to-pay for a non-market good may stem from a desire to preserve the option to enjoy the good in the future. An individual might contribute to preservation of the African wildlife preserve because she hopes to visit the preserve at some future date. Such non-use values may make up a significant portion of total economic value. In this case, revealed preference methods may significantly underestimate true willingness-to-pay for the non-market good.

3.3 Direct or Stated Preference Methods

While we have seen that revealed preference methods suffer from several problems, their appeal is that they examine consumer behavior in actual economic markets related to the non-market good in question, rather than simply trying to establish a hypothetical market for that good. Let us now turn to an examination of the stated preference techniques and the difficulties associated with these methods.

3.3.1 The Contingent Valuation Method

The contingent valuation method (CVM) is the most common form of stated preference method used in the valuation of non-market goods. Its goal is to establish a hypothetical market for a particular non-market good. A sample of individuals is presented with a hypothetical situation, in which they have the opportunity to purchase a particular good (Mitchell and Carson, 1989). A detailed description of the good the economist seeks to value is presented, along with the means for provision of the good, and the quality and quantity of the good. The description must specify the payment vehicle by which the consumer would pay for provision of the good. For example, the payment vehicle may take the form of a donation, a user fee, or an increased tax payment. Valuation of the good in question is achieved by asking the selected sample of individuals to indicate their maximum willingness-to-pay for the hypothetical good. Subjects may be asked to value a hypothetical non-market good in a telephone or mail survey, personal interview, or experimental setting.

The contingent valuation method began to be used in the early 1960s; Robert Davis (1963, 1964) used questionnaires in an attempt to measure the benefits associated with outdoor recreation in Maine. Acton (1973) conducted the first contingent valuation study examining health risks; he estimated willingness-to-pay for improved ambulance service to heart attack sufferers. The method was used throughout the 1960s and 1970s to estimate benefits in areas as diverse as air pollution (Ridker, 1967), waterfowl (Hammack and Brown, 1974), public parks (Darling, 1973) and improved water quality at Boston area beaches (Hanemann, 1978; Binkley and Hanemann, 1978). Evidence of the contingent valuation method's growing acceptance was its inclusion as one of the

three recommended methods for estimating project benefits in the Water Resource Council's 1979 "Principles and Standards for Water and Related Land Resources Planning", as well the Department of the Interior's 1986 approval of the method for benefit and damage assessment under the Comprehensive Environmental Response, Compensation and Liability Act of 1980, the Superfund Amendments and Reauthorization Act of 1986, the Clean Water Act of 1977, and the Oil Pollution Act of 1990 (Hausman, 1993). The growth of the contingent valuation method as a viable alternative for estimating the benefits of public goods was in part fed by funding from the United States Environmental Protection Agency, beginning in the mid 1970s. The need for accurate benefit estimation began to be acutely felt once the agency had a mandate, under the Reagan Administration's Executive Order 12291, to subject its regulations to benefit-cost analysis. In July 1984, a review panel of experts, including economists and psychologists, actively involved with contingent valuation research concluded that although the contingent valuation method held promise, further exploration of the technique was necessary (Mitchell and Carson, 1989). However, in 1993 a panel of experts appointed by the National Oceanic and Atmospheric Administration concluded that the contingent valuation method provides information that is "reliable enough to be the starting point of a judicial process of damage assessment" (Arrow et al., 1993). The panel's tentative endorsement of the CVM did not silence the controversy surrounding the method; the very same year a collection of research papers critical of the CVM was published (see Hausman, 1993). Among other things, some suggested that the contingent valuation method did not measure economic value, but rather ethical or emotional values, and general public opinion. Diamond and Hausman (1993) concluded that the contingent

valuation method "does not measure an economic value that conforms with economic preference concepts" and stated that CVM values should not be included in either benefit cost analyses or in compensatory damage measurements. Interestingly, the collection of studies was funded by the Exxon oil company, which had recently been embattled with legal troubles stemming from the Exxon Valdez oil spill in Alaska; nonetheless, Hausman maintains that the research presented persisted long after Exxon had settled all natural resource claims with the United States Government and the State of Alaska.

Contingent valuation methodology remains a prominent research area. While scholars of the technique have perhaps learned more about the CVM, new questions have arisen, and consensus regarding the technique is unlikely to be achieved in the near future, if at all.

3.3.1.1 CV Elicitation Methods

Several formats may be used to elicit the subject's willingness-to-pay; each method has its advantages and disadvantages. Mitchell and Carson (1989) state that the oldest and most commonly used CV elicitation method for some time was the bidding game, which operates on the same premise as an auction. Some have argued that this method has the potential to elicit the respondent's maximum willingness-to-pay (Cummings, Brookshire, and Schulze, 1986) and that the iterative process aids respondents in fully considering the value of the good in question (Hoehn and Randall, 1987). However, recently iterative bidding has fallen out of favor because, despite some of its attractive features, many feel that the bidding process is unduly influenced by the starting bid (Cummings, Brookshire, and Schulze, 1986). Starting point bias occurs if the elicitation method or the payment vehicle influences a respondent's stated willingness-to-

pay. Research has demonstrated that even if the initial bid is rejected, starting bids that significantly exceed the respondent's true WTP will usually increase the WTP value elicited, while initial bids much lower than true WTP will likely result in lower WTP estimates (Roberts, Thompson and Pawlyk, 1985).

Mitchell and Carson (1981, 1984) developed the payment card approach as an alterative to iterative bidding. The payment card approach asks a respondent to select the amount that represents his maximum WTP from a list of dollar amounts presented on a card. Mitchell and Carson (1989) note that this approach is subject to implied value cue bias, based upon the range of bids presented and any references to benchmarks that may be made. The term "implied value cue bias" is used to describe biases that may occur if the survey design causes a respondent to deviate from her true WTP value. This type of bias can occur for a number of reasons, including a very complex or long survey (Mitchell and Carson, 1989). Implied value cue bias will be discussed in detail later in this chapter.

Many investigators prefer the dichotomous choice CVM, sometimes referred to as the "take it or leave it" approach. This approach presents the subject with the description of the good, and a single bid, or monetary value, which the consumer may choose to accept or reject. A closely related method is the referendum approach, where the respondent chooses to vote 'yes' or 'no', reflecting whether or not he supports the provision of a specific good at the stated cost to himself and to the other parties involved. In each of these cases, the respondent is faced with a choice between two options. In the status quo scenario, the respondent's income remains the same, and the good to be valued is set at some baseline level. The alternative scenario in a willingness-to-pay study

lowers the income of the respondent in exchange for an increased level of the good. The respondent theoretically compares his utility under the status quo scenario to his utility under the alternative scenario; a 'yes' response is taken to indicate that the utility of the alternative exceeds the utility of the status quo. These methods are viewed as producing a hypothetical situation that is quite similar to actual market situations or actual voting situations; therefore, they should be familiar to most individuals. The NOAA Panel on the CVM recommended the dichotomous choice referendum format an appropriate elicitation method in most circumstances (Arrow et al., 1993). Mitchell and Carson (1989) suggest that these methods may be especially appropriate for mail surveys, and telephone surveys (provided that the scenarios presented do not require the use of visual aids). However, the "take it or leave it" approach and the simple referendum provide less information than methods such as open-ended CV questions or iterative bidding; an indicator of the respondent's WTP is obtained, but his actual WTP amount remains unknown. Another serious drawback of the traditional DC and referendum formats is that they require the researcher to make strong assumptions about the statistical distribution of values among respondents, or about the indirect utility function, to obtain mean willingness-to-pay (Mitchell and Carson, 1989; Ready, Whitehead, and Blomquist, 1995). However, it has been shown that if the bid amount is chosen carefully and the sample size is large, median willingness-to-pay can be calculated accurately and is less sensitive to assumptions regarding valuation and the indirect utility function (Mitchell and Carson, 1989). In addition, Ready, Whitehead, and Blomquist (1995) have noted that in status quo and alternative scenarios presented by these methods, neither choice will

dominate the other, and respondents may find it difficult or impossible to answer such questions if they are ambivalent between the two scenarios.

A modified form of the often problematic "take it or leave it" approach was proposed by Carson, Hanemann, and Mitchell (1986) in an effort to overcome some of the flaws of the traditional approach; this modified method is called the "take it or leave it with follow-up". The follow-up question depends upon the respondent's decision regarding the initial bid. If she chooses to undertake the alternative scenario at the initial bid, she is asked a follow-up WTP question that uses a higher price. On the other hand, the respondent may decline to undertake the alternative scenario at the initial bid, in which case the follow-up WTP question uses a lower price. Carson, Hanemann, and Mitchell (1986) argue that this method increases the amount of information the researcher elicits from the respondent, but caution that only one or two follow-up questions should be asked and that the follow-up price should be significantly different from the original price. They also note that all of the other difficulties associated with the traditional "take it or leave it" approach remain pertinent to the modified approach.

Finally, an "open-ended" or "direct question" CVM format would ask the subject to state his or her maximum willingness-to-pay for the hypothetical good described. In theory, this amount of money indicates the individual's point of indifference; he gains the same utility from having the described good and from having the monetary value that is his stated maximum WTP. This method is problematic because respondents may find it very difficult to simply come up with and state their true WTP. Consequently, this CVM format is associated with a large number of respondent non-responses and protest zero bids. The term "protest zero bids" refers to a situation where a respondent indicates that

his willingness-to-pay is zero, not because he has no value for the good in question, but because he objects to some aspect of the survey. For example, one may indicate a zero willingness-to-pay if he believes that no monetary value can be placed on the good in question or if he experiences a general sense of frustration with the survey.

3.3.1.2 Criticisms of the Contingent Valuation Method

The list of criticisms leveled at the contingent valuation method is long. While survey methods have been widely used in many areas of social science, economists have been particularly reluctant to sanction the use of such methods. Mitchell and Carson (1989) have observed that although many studies of an economic nature use survey data of some form, this data is typically collected by some entity other than the investigators themselves, such as the US Census Bureau. Mitchell and Carson (1989) also discuss the problems associated with the technique in depth. It is difficult to determine the extent of bias the elicited CVM value may have, since typically a traditional economic market does not exist for the good in question.

3.3.1.2.1 Informational Bias

Early work on the validity of the contingent valuation method suggested that only the type, quality and quantity of the good being valued should influence willingness-to-pay for that good; the influence of other factors on WTP was termed “informational bias”. Informational bias refers to the common sense notion that the willingness-to-pay estimate elicited via the contingent valuation method will depend upon how much and what type of information is presented in the description of the hypothetical non-market good. Recently, the danger of informational bias has been questioned; Arrow (1986), Kahneman (1986) and Randall (1986) have argued that elements of a hypothetical

scenario such as the payment vehicle and the amount of information provided to the respondents should influence willingness-to-pay values, since respondents are not valuing an abstract amenity, but a policy package. Hence, Mitchell and Carson (1989) reject the concept of informational bias, focusing instead on other forms of bias.

3.3.1.2.2 Implied Value Cue Bias

Mitchell and Carson (1989), in their extensive discussion of the CVM, describe biases that may result from "implied value cues". These biases exist because of the individual's proclivity to try and lighten the burden the survey task places upon him, particularly in the case where he is unfamiliar with the questions or the amenity that the survey is concerned with. Survey respondents may be quite uncomfortable with placing a dollar value on many public goods (e.g. wildlife preservation). Critics of the CVM argue that respondents may indicate their willingness-to-pay by simply pulling numbers out of the air, rather than truly grappling with the difficult trade-offs they are being asked to make. Kahneman, Slovic and Tversky (1982) maintain that when individuals are faced with an unfamiliar situation, they tend to make adjustments to some initial starting value to arrive at a final response. We like to take short cuts when a task seems difficult or foreign to us. This tendency is known as the "anchoring effect", possibly creating problems for the CVM because the adjustments to the initial value made by the respondents are often inadequate.

Implied value cue bias has several sources, which necessitates its further classification into the following forms of bias: starting point bias, range bias, relational bias, importance bias, and position bias. Starting point bias, which was described briefly above in the section on elicitation methods, is said to occur if the initial starting value

presented by the survey's hypothetical scenario influences respondents' stated WTP values. As noted above, iterative bidding and "take-it-or-leave-it" elicitation methods are believed to be particularly likely to result in starting point bias. A related form of bias is "yea-saying", the tendency of respondents to agree to the provision of a particular amenity at the specified bid, regardless of their own economic value for that amenity. Individuals may find it difficult to say 'no' to the provision of public goods, as these types of goods are usually viewed as "good causes". Mitchell and Carson (1989) describe "yea-saying" as "the discrete choice analogue of starting point bias".

Range bias refers to a situation in which the information presented on a payment card influences the respondent's WTP; respondents may consider the monetary amounts listed on the payment card as indicating the "correct" values one should place on the good being valued. If a particular respondent has a higher WTP, or a lower WTP, than any of the amounts listed on the card, she may be reluctant to reveal her true willingness-to-pay, fearing that it may be considered inappropriate in some way. Mitchell and Carson (1989) suggest that the possibility of range bias can be mitigated by listing a sufficiently high WTP upper bound and by allowing the respondent to choose from a large number of monetary values.

Relational bias may occur if the specific good being valued is linked in some way to another public good, which is called a "reference good" (Mitchell and Carson, 1989). If such a link exists and if an implied or explicit price is associated with the reference good, the respondent may be influenced by the price of the reference good when he states his WTP for the good actually being valued. This is yet another case where the respondent may lighten the burden of the valuation task by indicating a value similar to

the price of the reference good, avoiding the taxing problem of considering the true value he has for the amenity in question. Mitchell and Carson (1989) argue that the use of benchmarks in CV surveys, which is intended to remind respondents of public goods they currently provide and the prices of their provision, should be restricted to goods that are unrelated to the specific amenity being valued. They also note that relational bias is not restricted to payment cards; it can also occur when risk ladders are used in an attempt to convey certain levels of risk. Relational bias may also stem from respondents' knowledge of the prices of goods similar to the good they are being asked to value. Mitchell and Carson (1989) state that this type of relational bias is often seen in cases where quasi-private goods are provided by some form of government, and the government charges a user fee of some type (e.g. hunting and fishing license fees, parking lot fees at public beaches). Mitchell and Carson observe that such fees are not determined by market forces, but rather by political forces, and caution that individuals may view an established fee as representing a "fair" price for the resource. Consequently, when asked to indicate their value for similar resources, individuals may look to the range of current fees and restrict their WTP values to similar amounts. Several studies support this hypothesis (see Sorg and Brookshire, 1984; Nachtman, 1983).

Importance bias occurs if a respondent feels that if a CV study seeks to value a particular good, the good must be of value in spite of his or her own personal value for the good. Mitchell and Carson (1981, 1984) identified this problem during the pre-testing process for their national water quality studies. To avoid the presence of importance bias and overestimated benefits, efforts must be made to ensure that

respondents feel comfortable expressing their true values for the amenity in question, even if their true WTP is zero dollars. Mitchell and Carson (1989) also caution researchers against the use of evocative symbols in survey artwork, overly emotional wording, wording that portrays "villains" such as corporations, and one-sided presentations that ignore or downplay the difficult trade-offs that must be made. A related form of bias, position bias, may occur if respondents interpret a sequence of valuation questions that are concerned with the valuation of different resources as implying something about the relative importance or value of the resources. For example, the respondents may believe that only the first or last amenity has any value. If a sequence of questions involving more than one amenity is used, researchers should take care to inform respondents that this sequence of amenities is random (Mitchell and Carson, 1989).

3.3.1.2.3 Scenario Misspecification Biases

Mitchell and Carson (1989) also describe another class of related biases, which they term "scenario misspecification" biases. Such biases can be a problem if the respondent perceives some aspect of the contingent market incorrectly. However, the respondent's so-called "incorrect" perceptions may be due to an oversight on the part of the researcher. Scenario misspecification bias may stem from either theoretical or methodological problems. Theoretical misspecification bias occurs if the scenario described by the researcher is incorrect from a theoretical standpoint or is contrary to established facts. In such a case, although the respondent may understand the presented scenario perfectly, his indicated WTP values are not valid (Mitchell and Carson, 1989).

Methodological misspecification occurs when some aspect of the contingent market is inadequately described, and thus, the respondent perceived the scenario in a way unintended by the researcher. For example, respondents may interpret commonly used words in different ways. Elements of the hypothetical scenario that may be misinterpreted include descriptions of risk, the payment vehicle, and the implied budget constraint (Mitchell and Carson, 1989).

One form of methodological misspecification is "amenity misspecification bias", a term used to describe biases that may result when some aspect of the amenity description presents a problem. Amenity misspecification bias may be further classified into symbolic, part-whole, metric, and probability of provision biases. Symbolic bias refers to a situation where respondents focus on the general symbolism of the amenity being described, rather than the specific levels of provision. Some empirical studies have shown that symbolic bias can be a significant problem. For example, Kahneman et al. (1986) found that WTP values for a water quality program to preserve fishing were very similar for two programs that differed only in the extent of geographical coverage: one program benefited a single region in Oregon, while the other benefited a much larger geographical area. However, a comparison of WTP values for local water quality improvements (Desvousges, Smith, and McGivney, 1983) versus those for national water quality improvements (Mitchell and Carson, 1984) indicates that WTP values for national improvements are four times greater than values for local improvements, suggesting that symbolic bias may not be a problem in many CV surveys. Mitchell and Carson (1989) state that the chances of symbolic bias may be minimized by the avoidance of overly simplistic hypothetical good descriptions, special care when valuing a good that is likely

to be controversial (e.g. nuclear power risks) or a that may evoke strong emotional reactions (e.g. risk reductions in the risk of contracting AIDS from contaminated blood). Part-whole bias, which is closely related to symbolic bias, occurs if the respondent views the amenity being valued as a universal good, failing to weigh the specific description put forth by the CVM survey. Part-whole bias is a more specific problem than symbolic bias, as it focuses on three aspects of the good in question: its geographic distribution, benefit composition, and the policies that may accompany the good. Geographical part-whole bias occurs if the respondent cannot separate the part of the amenity that is provided from other parts of the amenity. Mitchell and Carson (1989) provide us with an example of geographical part-whole bias: a respondent, when asked to value water quality improvements in a local basin, may not be able to separate the basin from the other rivers. Benefit part-whole bias refers to a situation where the respondent has difficulty distinguishing between benefit sub-components. Finally, part-whole bias can occur if the respondent treats the policy package that accompanies the hypothetical good as part of a broader policy package. For example, they may believe that paying for a water quality program will also provide air quality benefits. To avoid part-whole biases, Mitchell and Carson (1989) suggest that respondents must be cautioned not to confuse the specific amenity being valued with a broader good, and including maps in the description of the good, if necessary. Researchers may also ask individuals to indicate their WTP for the complete amenity, followed by a request that respondents allocate their WTP to the various subcomponents of the amenity. Metric bias can occur if the scale that respondents use to value the resource is different from the one the researcher intended them to use. For example, respondents may ignore the cardinal nature of a

scale, and view the scale as ordinal. The problem of metric bias is especially relevant for valuations of risk reductions; respondents may not be able to process very small changes in risk. A final type of amenity specification bias is probability of provision bias, which can result if respondents' estimation of the probability that the good will be provided is different from the one intended by the researcher. Although surveys are usually designed to indicate that the good will indeed be provided if enough money is collected, some respondents may view this assurance with skepticism. This skepticism may be especially strong if the amenity description indicates that the government will provide the amenity being valued (Mitchell and Carson, 1989).

Context misspecification bias is another form of methodological misspecification, and this type of bias may be further classified into payment vehicle, property right, method of provision, budget constraint, elicitation question, instrument context, and question order biases (Mitchell and Carson, 1989). Payment vehicle bias occurs if the means of payment for the hypothetical good (e.g. a one-time fee, higher prices or utility bills, or an increase in taxes) influences the stated willingness-to-pay of the respondent. A number of studies provide evidence that the payment vehicle influences WTP for public goods (see, for example, Rowe et al., 1980; Greenley et al., 1981; Brookshire et al., 1980). However, Mitchell and Carson (1989) state that no misspecification or bias exists if the researcher understands the influences the chosen payment vehicle may have, and accepts these influences; they suggest that higher taxes and prices are relatively neutral payment vehicles and should be used when possible. Property rights bias may occur if the respondent feels uneasy about the notion that she "owns" the amenity being valued and questions the act of "paying" for its access or provision. This problem is tied

to the dilemma between using willingness-to-accept and willingness-to-pay as measures of value; this issue will be discussed in detail later in this chapter. Mitchell and Carson (1989) suggest that WTP may be the appropriate measure for avoiding property right bias. The means by which the amenity is provided can also affect WTP values; this influence is known as method of provision bias. WTP can be biased upward or downward if the provider is the government, and provision by charities may result in WTP values that are biased upward (Mitchell and Carson, 1989).

Many have argued that respondents approach CVM surveys in a causal way, failing to take into account fully their monetary constraints when asked to consider a hypothetical good. This problem is known as budget constraint bias. Mitchell and Carson (1989) argue that when valuation task involves small monetary amounts, the risk of such bias is small. However, when large, costly programs are valued, budget constraint bias can result lead to serious bias in the estimation of the benefits from such programs. Budget constraint bias may result from respondent confusion about whether the constraint is personal or household income. It can also occur when the respondent is asked to provide a good at a higher level than respondent he is currently providing it via higher taxes or prices; respondents may not know if the amount of money involved is in addition to what he is already spending for the good's provision, or the amount that he would need to provide if he were currently spending no money for provision of this good. Mitchell and Carson (1989) advocate the use of respondent pretests to avoid problems with budget constraint bias.

Elicitation question bias is likely to be present when the elicitation question used fails to elicit the respondent's maximum WTP and a firm commitment from him to pay

this amount. The strengths and weaknesses of CV elicitation methods were discussed in considerable detail above. A contributing factor to this bias is a belief on the part of the respondent that any amount he commits to is probably greater than the actual amount needed for provision of the good. Thus, the respondent fears that his money may be squandered. Mitchell and Carson (1981, 1984) found evidence of such a belief in their national drinking water quality studies. They suggest that the risk of elicitation question bias can be lowered by careful attention to the survey's wording and pre-testing.

Instrument context bias may occur if the survey information that precedes the hypothetical scenario affects WTP; the words the respondent reads or hears prior to the valuation task may "prime" her to give a certain response since the respondent will refer to the most easily accessible information in her memory. The work of Kahneman, Slovic, and Tversky (1982) supports the existence of this form of bias, terming this tendency the "availability heuristic". Evidence of this effect is seen in a study by Bishop (1985) on the level of interest in political affairs. Context effects likely do occur frequently in survey research, but consensus has not been reached regarding the importance of such effects; some authors have argued that such context effects are likely to be small and occur primarily when the issue at hand is particularly complex or obscure (Tourangeau et al., 1985).

A final form of context misspecification bias, question order bias, may occur if the sequence in which the questions are asked affects individuals' responses in a manner unintended by the researcher. This problem may be especially important in surveys that ask respondents to value several different levels of the same good, or several different goods. Mitchell and Carson (1989) suggest that independent valuation questions should

be asked in a random order; it has now become common practice among survey researchers to test for order effects.

3.3.1.2.4 Embedding, Scope, and Warm Glow Effects

Embedding effects present additional problems for the validity of the contingent valuation method. Kahneman and Knetsch (1992) defined "perfect embedding" as a case where the respondent's stated WTP remained the same for preservation of one, two, or three wilderness areas. Kahneman and Knetsch believed that such embedding was observed because individuals view CVM questions not as an opportunity to express preferences over resources, but as an opportunity to "purchase" moral satisfaction. The desire to support general environmental quality is fulfilled whether one contributes to the preservation of one or several wilderness areas; thus WTP does not increase with the extent of protection as economic theory suggests it should.

The concept of "scope" is closely related to embedding effects. Many feel that in order for CVM WTP estimates to be valid measures, they must be sensitive to the amount, or scope, of the environmental amenity being offered. The NOAA panel on the CVM proposed the "scope test" in its 1993 recommendations (Arrow et al., 1993). The panel stated in comments one year after its recommendations that "a survey instrument is judged unreliable if it yields estimates which are implausibly unresponsive to the scope of the insult" but noted that "this, of course, is a judgment call, and cannot be tested in a context-free manner" (Arrow et al., 1994).

Smith and Osborne (1996) conducted a meta-analysis of WTP estimates for visibility improvements at US National Parks from five different CVM studies (Rowe et al., 1980; MacFarland et al., 1983; Schulze et al., 1983; Chestnut and Rowe, 1990; and

Balson et al., 1990). These five studies were chosen for the meta analysis from thirteen studies of air pollution changes and their influence on visibility and other local conditions because they employed similar methods. The dependent variable of the model was the estimates of mean WTP for a visibility change, as originally reported in each study (for a total of 116 WTP estimates). All estimates either originally expressed mean WTP as a monthly payment in 1990 dollars, or were converted to do so using the Consumer Price Index. The model's independent variables included commodity-related and survey design variables. Six different model specifications were estimated. A positive and significant relationship between mean WTP and proportionate improvements in visibility range was found, and this relationship held regardless of the model specification used. The relationship was also robust to restrictions of the sample set, changes in independent variables, and the treatment of the intercept term. Smith and Osborne conclude that a consistent economic relationship exists between mean WTP and changes in visibility relationship, suggesting that the CVM mean WTP estimates do indeed pass the "scope" test.

The issues of embedding and the sensitivity of WTP values to scope are related to what has been termed the "warm glow" effect; respondents to CVM surveys indicate a positive willingness-to-pay for a specific resource, not because their stated WTP indicates the amount of economic value on the particular resource in question, but because they enjoy contributing to "good" causes and derive a general sense of happiness from such contributions. Thus, critics of the contingent valuation method maintain that CVM does not measure individual's underlying preferences (Hausman, 1993).

3.3.1.2.5 Lack of Familiarity with Valuing Non-Market Goods

Much of the critique of the contingent valuation method focuses on the strong possibility that individuals are unfamiliar with the good being valued and thus may take "shortcuts"; individuals do not have experience with purchasing "goods" such as health risk reductions or environmental amenities. However, Krupnick and Cropper (1992) have pointed out that citizens often vote on issues about which they do not have perfect knowledge. Thus, one could contend that valuing an unfamiliar good is not so different from decisions that arise out of voting behavior. Viscusi, Magat, and Huber (1991) conducted a computer interactive survey to estimate willingness-to-pay for a decreased risk of developing chronic bronchitis. The sample used was 389 shoppers at a mall in North Carolina. Krupnick and Cropper (1992) later administered the same survey to individuals that had a relative afflicted with a chronic lung disease. They found that while familiarity with chronic lung disease had no influence on mean responses for risk-risk trade-offs, those subjects who had a relative with chronic lung disease were willing to pay more, on average, for a reduction in the risk of chronic bronchitis.

3.3.1.2.6 Divergence between Willingness-to-Pay versus Willingness-to-Accept Measures of Value

A well-documented discrepancy often exists between an individual's willingness-to-pay (WTP) and willingness-to-accept (WTA) values (see Schulze et al.,1983; Tolley et al., 1983; Knetsch and Sinden, 1984; Cummings et al., 1986). Empirical comparisons of WTP and WTA commonly yield WTA estimates that are four to fifteen times as large as WTP estimates for the same amenity, in spite of standard economic theory's prediction that these measures should be approximately equal (Hausman, 1993). Mitchell and

Carson (1989) argue that there are two questions raised by this discrepancy: the theoretical issue of how far apart WTP and WTA Hicksian consumer surplus measures should be, and the methodological issue of whether or not WTA elicitation questions are more or less valid than WTP measures. They also argue that the choice between a WTP or WTA elicitation format may be viewed as a question of property rights--that is, who has the right to sell the non-market good? In the case of public goods, property rights are not well defined; thus we must consider if the polluter should "buy" the right to pollute from the citizens or if the citizens should pay for cleanup of pollution.

In the first study that elicited both WTP and WTA measures, Hammack and Brown (1974) found in their waterfowl benefits study that WTA values were over four times larger then WTP values for the same good. Expanding of the work of Willig (1973, 1976), Randall and Stoll (1980) were able to show that Willig's findings for changes in price (that the divergence between WTP and WTA is relatively narrow for expenditures that are small relative to one's income, and that Marshallian consumer surplus is bounded by WTP and WTA) also held for changes in the quantity of a good (Mitchell and Carson, 1989). Mitchell and Carson (1989) state that during the same time period as these theoretical developments, large discrepancies between WTA and WTP continued to characterize CVM studies. These studies also found that protest bids and infinite values were more prevalent when respondents were asked to value hypothetical goods in the WTA format.

The WTP vs. WTA question continued to be heavily researched throughout the late 1970s and 1980s, and four possible explanations began to emerge: rejection of the WTA property right, consumer cautiousness, prospect theory, and modifications of

established economic theory (Mitchell and Carson, 1989). Many individuals may balk at the property rights implied by the WTA format, and thus their WTA values are exaggerated, or they may outright refuse to provide a value. The work of Hoehn and Randall (1983, 1985, 1987) gave rise to the consumer cautiousness hypothesis, suggesting that consumers who are uncertain or risk adverse, or who may not have adequate time to perform the optimization problem required of them, may provide lower WTP and higher WTA values than they would under a more favorable set of conditions (certainty, risk neutrality and unlimited time to consider the optimization problem at hand). The solution proposed by Hoehn and Randall was to allow respondents the opportunity to become comfortable with valuation in both the WTP and WTA formats. Several studies put this solution to the test (Brookshire, Coursey, and Schulze, 1986; Brookshire, Coursey, and Radosevich, 1986; Coursey, Hovis, and Schulze, 1987), finding that the discrepancy between WTP and WTA values tended to narrow as respondents repeated the valuation process. WTA values declined, while WTP values remained relatively constant, as the learning trials progressed. Kahneman and Tversky's prospect theory (1979) defined a value function that is defined on deviations from a reference point, is generally concave for gains and convex for losses, and steeper for losses than for gains. This suggests that, contrary to the assumptions of standard expected utility theory, individuals will value losses from their status quo situation more heavily than gains of the same magnitude. Hanemann (1982, 1983a, 1984a, 1984b) argued that the WTP vs. WTA discrepancy was due, in part, to the discrete choice format then prevalent in CVM surveys. His work showed that, in certain cases, the gap between WTP and WTA could be substantial, yet consistent with established economic theory. Subsequent theoretical

work by Hanemann (1986a) further explored the work of Randall and Stoll (1980), demonstrating that the size of the WTP-WTA gap depends upon availability of substitutes for the hypothetical public good in question; the gap narrows as the number of available substitutes increases because the elasticity of substitution is large relative to the elasticity of income. The ratio of income elasticity to elasticity of substitution, which determines the size of the gap, is not readily obtainable from the standard valuation function regression results; one must specify the utility function (Mitchell and Carson, 1989). Mitchell and Carson (1989), after extensive review of the literature surrounding the WTP-WTA discrepancy, reasonably conclude that the discrepancy likely persists because of a combination of the factors discussed above.

Advocates for the continued use of the contingent valuation method have addressed the WTP-WTA dilemma by pointing to theoretical results such as Hanemann's and arguing that one can use WTP as a conservative, lower-bound estimate of economic value. Skeptics of the CVM use the WTP-WTA disparity as further evidence that the contingent valuation method cannot validly measure underlying preferences. Some critics of the CVM contend that respondents answer valuation questions from an emotional or ethical, rather than an economic, perspective and point to the WTP-WTA gap as a natural consequence of the ethical mindset of individuals when they respond to valuation tasks. They argue that respondents may view WTP as merely the amount of money they are willing to sacrifice for the pursuit of their ethical beliefs, while WTA is viewed as essentially a bribe (Hausman, 1993). The observation that individuals tend to value gains differently from losses will again be made when the issues surrounding the discounting of future benefits are discussed in Chapter Four.

3.3.1.2.7 Hypothetical, Strategic, and Compliance Biases

The hypothetical nature of stated preference methods is perhaps its biggest downfall; after years of research, stated preference methodology continues to draw the attention of critics. Among the most serious difficulties for practitioners of the contingent valuation method are the problems of hypothetical, strategic, and compliance biases. Hypothetical bias refers to a significant difference between an individual's stated, hypothetical willingness-to-pay, and his true willingness-to-pay (Cummings and Taylor, 1999). Harrison and Rutström (1999) suggest that the presence of hypothetical bias is highly likely because it is rational for respondents to overstate their WTP values in a hypothetical scenario: "As a matter of logic, if you do not have to pay for the good, but a higher verbal willingness-to-pay increases the chance of its provision, then verbalize away to increase your expected utility!" It is rather easy to indicate that one *would* be willing to pay some positive amount of money for a particular good; it is something else to *actually* give up that money and the other goods it has the power to purchase. Cropper and Oates (1992) caution that the subject must be reminded of his budget constraint; otherwise, he may exaggerate the value that he places on the hypothetical good. Mitchell and Carson (1989) argue that there is a class of biases that stem from incentives survey respondents may have to "misrepresent responses". Strategic bias occurs if the respondent intentionally tries to influence the future payment or provision of the specific amenity in question. The free-rider problem is the form of strategic bias that draws the greatest attention. Suggestion for mitigating strategic bias include designing the survey such that the payment vehicle is plausible and taking care not to over-emphasize the importance of the individual's response to policy makers (Mitchell and Carson, 1989).

Compliance bias occurs if the respondent tries, whether unconsciously or consciously, to fulfill his perceived expectations of the interviewer or the purpose of the study. Mitchell and Carson (1989) argue that compliance bias can be avoided by taking care not to provide survey participants with any additional instructions beyond those that are carefully prepared in advance and given to all survey participants.

Comparisons between responses to hypothetical and real valuation questions, as well as comparisons between hypothetical values and actual charitable giving, have yielded considerable empirical evidence that hypothetical bias is a significant problem for the CVM (Seip and Strand, 1991; Duffield and Patterson, 1992; Loomis et al., 1994; Neill, 1995; Cummings et al., 1995; Brown et al., 1996; Fox et al., 1998). However, a few studies have failed to find evidence of significant hypothetical bias (Brookshire and Coursey, 1987; Dickie et al. 1987; Smith and Mansfield, 1997). Researchers may assume that differences between hypothetical WTP and actual WTP must represent hypothetical bias. However, one cannot necessarily attribute such differences to hypothetical bias alone, as there are several confounding factors that must be addressed. Most stated preference studies are concerned with the valuation of public goods; these studies cannot ignore the proclivity of respondents to "yea-say" and "free-ride" when asked to value public good. Willingness-to-pay estimates may be biased upward by closely associated problems of hypothetical bias and "yea-saying", but may be biased downward by the tendency to free-ride. Thus, it is difficult to attribute discrepancies between responses to hypothetical and real valuation questions to hypothetical bias alone; a combination of factors may be at work.

The extent of these types of biases will depend in part on whether or not the method for eliciting individuals' true preferences is "incentive-compatible". A demand revelation method is incentive-compatible if it creates a situation where it is in the individual's best interests to reveal his true preferences for the good, and he will not be tempted to engage in free-riding behavior (Mitchell and Carson, 1989). Paul Samuelson (1954, 1955, 1958, 1969) believed that free-riding behavior would always be the individual's best response, regardless of the questioning format. This assumption contributed to arguments against the CVM. However, the assumption that individuals will always choose to free-ride has been challenged by several authors including Malinvaud (1971), Dreze and Vallee Poussin (1971), Clarke (1971), Groves (1973), Groves and Loeb (1975), and Groves and Leydard (1977). Malinvaud (1971), and Dreze and Vallee Poussin (1971) showed that a benevolent planner may acheive a Lindahl-Pareto equilibrium using individuals' marginal WTP in a continuous incremental dynamic Walrasian tatonnement process. Clarke (1971), Groves (1973), Groves and Loeb (1975), and Groves and Leydard (1977) developed a set of incentive-compatible demand revelation methods using taxes, subsidies, or side-payments; these methods create a hypothetical scenario where it is in the individual's best interest to reveal his true WTP. Akerlof (1983), and Akerlof and Dickens (1982), challenged the assumption that strategic behavior is costless beyond the immediate choice before the individual, building a model of honesty and cooperative behavior in which honesty leads to long-run economic gains that may be lost if dishonesty and non-cooperative behavior occurs.

The theoretical challenges to the assumption that "free-riding" is the individual's rational, dominant strategy were later buffered by the rise of experimental economics and

the empirical measures it provided. Early experiments by Babb and Scherr (1975) compared three elicitation methods for WTP for a concert series and additional books for a college library: the Clarke tax, another incentive-compatible demand revelation method, and a voluntary revelation method. They found that the voluntary method resulted in larger real payments than the two incentive-compatible demand revelation methods. While many colleagues dismissed this result as occurring due to the novelty of the experiment, induced altruism, insufficient incentives, and respondent lack of experience with incentive-compatible methods, Johansen (1977) maintained that there was little empirical evidence to support free-riding behavior and suggested that economists may have overestimated the importance of such behavior. The experiments of Vernon Smith and his colleagues (1977, 1979, 1980, 1982), regarded as more realistic, failed to refute the earlier findings of Babb and Scherr (1975). Smith's work suggests that both incentive-compatible and voluntary demand revelation methods eventually achieve cooperative behavior, and cooperative behavior is often reached more quickly with voluntary mechanisms. Mitchell and Carson (1989) suggest that theoretical results supporting incentive-compatible methods rests on several assumptions that do not seem to hold: zero costs to determining one's optimal strategy, zero costs to the act of responding dishonestly, and that the individual perceives no risk that his actions may prevent an optimal competitive or cooperative strategy.

Research exploring the incentive compatibility of various elicitation methods suggests that a commonly used CVM format can induce the individual to reveal her true preferences. Zeckhauser (1973), and Hoehn and Randall (1987), demonstrate that the discrete choice "take-it-or-leave-it approach" for the provision of a public good at a set

price is an incentive-compatible mechanism; the respondent's best interest to answer 'yes' if her willingness-to-pay is greater than or equal to the stated price. However, the belief that the dichotomous choice (DC) CVM format is incentive-compatible has been challenged by more recent literature (Cummings and Taylor, 1999; Taylor, 1998) that suggests that in order for a DC referenda to be incentive-compatible, it must be a closed referenda; this literature will be discussed in detail later in this section.

Research examining the problem of hypothetical bias has benefited greatly from the rise of experimental economics. Two types of approaches have been used: calibration of CVM estimates and survey designs that encourage the respondent to provide hypothetical values that accurately reflect his actual willingness-to-pay (Cummings and Taylor, 1999). The goal of calibration techniques is to quantify the differences between individuals' hypothetical WTP values and true (non-hypothetical) WTP values. Shogren (1993) proposed the CVM-X method, which has four steps: elicitation of hypothetical values for the good in question via a survey, elicitation of real values for the good from a subset of the survey respondents in a laboratory setting via an incentive-compatible auction, estimation of a calibration function relating the hypothetical values of the entire sample (the survey respondents) to the real values of the sample subset (those that participated in the auction), and use of the estimated calibration function to correct the values of the survey respondents that did not participate in the auction.

Fox et al. (1998) argue that the calibration method CVM-X is cost-effective method that combines the strengths of CVM surveys and experimental economics. Their study focused on the value of reducing the risk of illness from the parasite *Trichinella* by

food irradiation. The authors argue that this "good" is similar to many non-market goods because most Americans are unfamiliar with irradiation. Two hundred and fifty randomly selected households were asked to indicate whether they would prefer to eat irradiated pork sandwich (with no risk of trichinosis) or a non-irradiated pork meat sandwich (with a 1 in 2.6 million chance of trichinosis); respondents were then asked, in an open-ended format, to indicate their maximum WTP for their preferred sandwich. After completing the telephone interview, each respondent was given the opportunity to participate in an experiment in exchange for a $30 payment. Each experiment had two stages: in the first stage respondents gain experience with auctions through auctioning candy bars, and in the second stage respondents are informed about irradiation and told that they will have to eat either their less preferred sandwich (the endowed sandwich) or their preferred sandwich (the sandwich they may bid on). A reminder of the individual's hypothetical bid (obtained through the telephone interview) is given to him or her at the beginning of the second stage; this reminder may introduce an "anchoring" effect since respondents may be reluctant to significantly alter their initial bid. The Vickery auctions had ten trials to enable respondents to gain repeated market experience. Additional information about the irradiation process was provided to respondents after trial five. The average bid for non-irradiated pork fell $0.18, from $0.44 in the survey to $0.26 in trial 2 of the auction, while the average bid for irradiated pork fell $0.19, from $0.58 in the survey to $0.39 in trial 2 of the auction. The final bid for non-irradiated pork was $0.27, while the final bid for irradiated pork was $0.48.

Simple calibration factors were calculated based on the ratio of the trial 2 bid to the survey bid, and on the ratio of the final bid to the survey bid; respondents'

socioeconomic characteristics and bid elicitation factors were not considered in the simple calibration calculations. For non-irradiated pork, the calibration factor between the survey and trial 2 bids was 0.59 and the calibration factor between the survey and final bids was 0.61. For irradiated pork, the calibration factor between the survey and trial 2 bids was 0.67, and the calibration factor between the survey and final bids was 0.83.

Fox et al. also estimated four models of the calibration function for each type of pork; two used the trial 2 bid as the dependent variable, and two used the final bid. All models resulted in a positive and significant relationship between the hypothetical (survey) bid and the subsequent (actual) bid. Gender was not significantly related to either the trial 2 or final bid. The models also found no significant relationship between auction characteristics such as group size and the final bids for trials 1 and 7. These models resulted in calibration factors for irradiated pork of 0.69 for the trial 2 bid and 0.86 for the final bid. Non-irradiated pork calibrations factors were 0.55 for trial 2 and 0.28 for the final bid. Estimated calibration factors from the regression models are similar to the more rudimentary calibration factors presented above except in the case of the final bid for non-irradiated pork. Fox et al. suggest that this is because some respondents who opposed the irradiation process may have initially overstated their WTP for the non-irradiated sandwich.

Fox et al. were able to reject two null hypotheses: that the calibration factor is equal to one, suggesting that hypothetical and real WTP are not significantly different, and that the calibration factor is 0.5, suggesting stated hypothetical WTP should be deflated by 50% in order to reflect individual's true WTP. The latter is an especially important finding in light of a 1994 recommendation from the NOAA panel

commissioned to study the CVM; this recommendation proposed a rule that CVM values be deflated by 50% in the absence of an alternative calibration factor. The empirical evidence of Fox et al. supports the argument that individuals' hypothetical values may be exaggerated, but indicates that hypothetical WTP values are not biased upward to the extent suggested by the initial NOAA recommendation. Fox et al. conclude that laboratory experiments are useful in correcting for hypothetical bias, but note that calibration factors may be commodity specific, as evidenced by the difference in the factors for the two types of pork. They also recognize that most public goods are unlike the risk valuation they undertook, which conveniently could be captured by a tangible object--an irradiated versus non-irradiated pork sandwich. Hence, Fox et al. suggest that the CVM-X method should be restricted to the valuation of private goods and studied further before any efforts are made to utilize it in the valuation of public goods.

Other efforts to address hypothetical bias have focused on eliciting true willingness-to-pay values through carefully refined CVM survey design; studies of this type have used both private and public goods. Johannesson et al. (1998) compared dichotomous choice contingent valuation responses with real purchase decisions; the good examined was a box of chocolates. The survey wording emphasized that the subjects should consider whether they wished to purchase the box of chocolates "here and now". Their study indicates that hypothetical 'yes' responses overestimate true 'yes' responses; this effect seem to be most pronounced at the highest bid amount. To examine the discrepancy between hypothetical responses and real responses, the investigators adopted a more stringent approach, counting a hypothetical 'yes' response as a 'yes' only if the respondent indicated he was 'absolutely sure' (as opposed to 'fairly sure') that he

would purchase the box of chocolates. Johanannesson et al. found that these more conservative hypothetical responses underestimate the true 'yes' responses. In addition to examining hypothetical bias and ways to minimize it, the results of this study also raise the issue of respondent certainty, which will be discussed later in this chapter.

Cummings and Taylor (1999) note that efforts to eliminate hypothetical bias through refining survey techniques have been limited to a reminder, prior to the valuation questions, to respondents to consider their budget constraints. They argue that such reminders have failed to eliminate hypothetical bias (see, for example, Loomis et al., 1994 and Neill, 1995). Consequently, Cummings and Taylor directly address the hypothetical bias problem through the development of a survey instrument that discusses the problem with respondents; this technique is known as the "cheap talk" design.

Cummings and Taylor conducted experiments involving four different public goods: contributions to the Georgia chapter of the Nature Conservancy to acquire and preserve land (the NC good), contributions to an Albuquerque organization to fund the publication and distribution of a citizens' guide to low income, Hispanic households that informs them of possible drinking water contamination (the ABQ good), contributions to the Nature Conservancy to purchase rain forest acreage in Costa Rica (the RF good), and contributions to a Georgia organization that builds and maintains walking and bicycling trails in the Atlanta area (the PF good). For each of these four goods, some individuals vote on a hypothetical referendum and some vote on a "real" referendum that actually requires the individual to pay if the referendum is passed. In each case, the null hypothesis to be tested is that there is no significant difference between the 'yes' responses in the hypothetical and real situations. If the null hypothesis of no hypothetical

bias is rejected, then the "cheap talk" design is used, and responses to this design are compared with responses to the real referendum. The "cheap talk" method differs from the hypothetical method only in the additional words that are read to the participating individuals; the method describes the hypothetical bias problem, reasons why hypothetical bias might occur, and asks individuals to vote in the hypothetical referendum as though the referendum was real. Thus, the "cheap talk" approach explicitly addresses the problem of hypothetical bias with individuals that are about to value a particular public good.

The subjects in Cumming and Taylor's study were students in undergraduate courses at Georgia State University. Prior to voting on the provision of one of the four public goods, the students were asked to sign a consent form acknowledging voluntary consent and their receipt of a participation fee, abide by the rules of the referendum, and participate in a series of double auction experiments, which lasted about 50 to 60 minutes. The request that students participate in the auction experiments prior to the valuation task was made in an effort to avoid "endowment" or "found money" effects. Such effects may occur if individuals in the study feel an obligation to spend their participation fee during the course of the experiment; Cumming and Taylor state that requiring students to participate in the auctions beforehand created a sense among the students of having earned their participation fee.

Results from the initial experiments (comparisons of responses to hypothetical and real referendum for each good) lead to a rejection of the null hypothesis that the 'yes' responses do not depend on the type of referendum (hypothetical or real) for all goods except the biking and walking trails; these results supported the existence of hypothetical

bias. The "cheap talk" method was then conducted for the NC, ABQ and RF goods. For each of the goods, the null hypothesis that the 'yes' responses do not depend upon the type of referendum ("cheap talk" or real) could not be rejected, suggesting that the "cheap" talk method is indeed effective at reducing hypothetical bias.

Taylor (1998) cautions that eliciting real payments in a laboratory setting requires the use of a demand revealing mechanism; otherwise observed discrepancies between responses to real and hypothetical valuation questions may stem from free-riding behavior in the real payment situation (biasing real WTP values downward), rather than from hypothetical bias in the hypothetical payment situation (biasing hypothetical WTP values upward). She notes that the referendum payment mechanism used in her work with Cummings (1999), as well as in Cummings et al. (1997) and in Bjornstad et al. (1997) may not have been truly incentive-compatible because the referendum is not "closed". A referendum that is not "closed" leaves open the possibility that the public good in question may be provided by persons outside of the voting group; individuals in the voting group may elect to free-ride off the potential providers outside the voting group. The discrepancies observed in the percentages of 'yes' responses to hypothetical and real referendum questions in Cummings et al. (1997), Bjornstad et al. (1997), and Cummings and Taylor (1999) cannot be attributed to hypothetical bias with certainty because these studies employed non-closed referenda. Taylor notes that prior research has not employed closed referendum because the conditions that must be met are difficult to achieve with actual public goods.

Taylor conducted an experiment in which participants were offered the opportunity to contribute to the publication and distribution of a bilingual "Citizen's

Guide" that discussed drinking water contamination and would be distributed to low-income Hispanic households in Albuquerque, New Mexico. This good was also used in the Cummings and Taylor (1999) experiments. However, the good used by Taylor provided the guide to additional households that had not yet received it and would not receive it, unless the participants in Taylor's experiments provided the funds. Thus, her experiment is closed because the low-income Hispanic households that did not have the guide would not receive it unless the participants in the experiment voted to donate the funds. The experiment participants were drawn from undergraduate classes at Georgia State University. The consent and participation fee procedures of Cummings and Taylor (1999) were followed, and as in this earlier study, participants were asked to engage in 50 to 60 minutes of oral double auctions to avoid "found money" effects.

Taylor (1998) found that the percentage of 'yes' responses to the hypothetical referenda was significantly greater that the percentage of 'yes' responses to the real referenda. Probability models, incorporating participants' socioeconomic characteristics and experimental design elements, were also estimated. Experimental design factors included the number of participants in an individual's voting group, the individual's oral double-auction earnings, and whether the individual responded to a real or hypothetical referendum. The results of the estimation indicate that the probability that an individual votes 'yes' to provide the good is twenty percent lower if he is presented with a real, as opposed to hypothetical, referendum; the estimation results also suggest that an individual's income and race are significant factors in voting behavior.

The extent of hypothetical bias in CVM estimates is likely to be an ongoing area of interest in CVM research. Attributing discrepancies between willingness-to-pay in

hypothetical versus real situations to hypothetical bias is particularly difficult in the case of public goods because of the counter effects of free-riding and "yea-saying" behavior on hypothetical WTP; free-riding behavior results in downward bias of hypothetical WTP, while "yea-saying" results in upward bias of hypothetical WTP. In addition, as noted by Taylor (1998), comparisons between real and hypothetical WTP values for public goods can be hindered by the presence of free-riding behavior in the real valuation scenario. Use of the closed referenda format may help to mitigate free-riding behavior in both the real and hypothetical valuation contexts, but it is unlikely that this format can be plausibly maintained for most public goods.

3.3.1.2.8 Sampling and Aggregation Issues

The ultimate goal of the survey research is to use a sample of individuals to infer the value that the population of interest places on a particular good or program. WTP estimates obtained from a particular sample may have valuable policy implications. Unfortunately, the process of choosing one's sample and obtaining meaningful responses from that sample is fraught with potential problems. Mitchell and Carson (1989) provide an extensive discussion of the problems associated with sampling and aggregation. This section will briefly describe the major difficulties the researcher may confront.

Mitchell and Carson (1989) argue that sampling issues have not received adequate attention in the CVM literature because CVM researchers have been focused on addressing the measurement problems described above and many researchers do not have rigorous statistical training. They identify two classes of biases that can result from the sampling and inference process: sample design and execution biases, and inference biases. Sample design and execution biases include population choice, sampling frame,

sample non-response, and sample selection biases. Population choice bias occurs when the population chosen for study is not the population who will enjoy the benefits or bear the costs of the public good being valued. Mitchell and Carson (1989) state that this type of bias is most likely when there is divergence between those who pay for the amenity's provision and those who will benefit from its provision. Sampling frame bias occurs if the sample chosen for study does not reflect the population of interest; in this case it is inappropriate to use estimates obtained from the sample to infer the population's value for the good in question. Factors that can contribute to this type of bias include non-geographically based populations and frequent difficulty in obtaining up-to-date addresses for mail surveys. Non-response bias occurs if some of the individuals selected in the sample fail to respond to the survey. Non-response can take two forms: unit non-response, where the individual or household fails to respond to the survey, and item non-response, where the respondent answers some, but not all, of the survey questions. Mitchell and Carson (1989) state that non-response rates of 20%-30% are common in CVM research, and that a small amount of non-response is welcome because the researcher wishes to avoid including thoughtless responses in his analysis. However, large amounts of non-response can pose a serious threat to the researcher's ability to make inferences based on the sample. Factors that can contribute to non-response include sampling the general population, complex scenarios, and the elicitation of values for unfamiliar amenities.

There are four general categories of item non-response: don't know, refusal, protest zero bid, and inconsistent responses. The term "protest zero" is used to describe the situation where a respondent indicates that he is not willing to pay anything for the

environmental amenity, not because he places no value on it, but because he objects to the valuation question. Mitchell and Carson (1989) suggest that protest zero bids may be the most problematic form of item non-response, because the researcher is faced with the task of distinguishing between those who are protesting and those who truly do not place any value on the resource in question.

Both unit and item non-response result in a loss of valuable information. Whether or not non-response results in bias depends upon whether or not non-respondents are significantly different from respondents. To determine the degree to which non-response bias exists, the researcher must try to determine the degree to which significant discrepancies exist in the response rates of groups of households (between-group sample non-response bias) and if there are systematic differences between those within one group who did respond and those who chose not to respond (within-group sample selection bias). Either or both of these forms of non-response bias may be a problem in a CVM study; both forms of bias and possible solutions are further discussed in Mitchell and Carson (1989). Mitchell and Carson suggest that mail surveys are particularly likely to have some form of non-response bias. Techniques for reducing unit non-response bias include following up on the initial request for survey information via a telephone call or additional mailing and pretests of the survey design to check for its clarity. Mitchell and Carson (1989) also applaud Bishop and Boyle's (1985) conservative estimation of WTP for an Illinois nature preserve (1985); Bishop and Boyle assumed that the individuals who failed to respond to a deliverable survey package had zero value for the resource. Dillman (1978) presents a set of survey procedures that have been shown to increase response rates; his techniques attempt to create a survey package (the cover letter and

actual questionnaire) that strikes the potential respondent as attractive and conveys that her views are of importance, thereby piquing the potential respondent's interest.

Inference bias can occur if the researcher seeks to infer something about a policy change that was not explicitly valued in her study. Forms of inference bias include temporal selection and sequence aggregation biases. Temporal selection bias occurs if society's valuation of a policy scenario changes over time. For example, society's preferences for environmental amenities changed in the 1960s and 1970s, and the nuclear reactor accident at Three Mile Island helped to contribute to growing opposition to nuclear power (Mitchell and Carson, 1989). Mitchell and Carson suggest that while values for familiar amenities are likely to remain stable over time, values for newly recognized amenities and risks are likely to shift, particularly in the case of widespread media attention.

Sequence aggregation bias refers to a situation where the researcher attempts to add the independently estimated values of small policy changes to obtain the value of a large policy change; this be done in the context of a large geographic area (addition of the values obtained from each small geographical region of the larger area) or a set of public goods (addition of the values obtained for each public good in the set). Mitchell and Carson argue that the values for the large area and for the set of goods are likely to be smaller than the sums of the independently obtained values for the small areas and of the independently obtained values for each good in the set, respectively. These problems are respectively referred to as geographical and multiple public good sequence aggregation bias (Mitchell and Carson, 1989).

3.3.2 Conjoint Analysis

As debate about the validity of contingent valuation continues, alternatives to the traditional contingent valuation method are being explored. Conjoint analysis is a type of stated preference method where individuals are asked to rate or rank various goods, instead of explicitly pricing the good (as in the CVM). Conjoint analysis was used in marketing, transportation, and psychological research before being pursued by environmental economists as an additional methodology that could be used to capture non-use values and that may have advantages over the CVM (see Cattin and Wittink, 1982, for a discussion of conjoint analysis in marketing research; Adamowicz et al., 1998). Conjoint analysis, and similar techniques that ask individuals to choose from alternative programs, are often referred to as "trade-off analysis" or "choice experiment" techniques; these techniques all seek to value marginal changes in the attributes of multi-attribute goods by asking respondents to make trade-offs among the various attributes of each good. For example, the attributes of a groundwater protection program might include the cost of the program to the household, the type of contaminants the program reduces or eliminates, and the geographical extent of protection. Health economists frequently use a trade-off analysis approach called QALY (quality-adjusted life year) analysis; QALY analysis avoids explicit monetary valuation by asking individuals to rate different health states according to the strength of their preference or aversion to the health state described by the questionnaire. Each health state may have a number of attributes, such as mobility, physical and social activity, and symptoms (Fabian, 1994). QALY analysis will be discussed in more detail in Chapter Five.

Recent research suggests that conjoint analysis and other forms of trade-off analysis may have several advantages over the traditional CVM approach (Johnson et al., 1995; Irwin et al., 1993; Magat et al., 1988; Desvouges et al., 1983). Higher response rates and fewer protest zero bids have been associated with conjoint analysis (Ready, Whitehead, and Blomquist, 1995). The conjoint format allows respondents to indicate indifference or ambivalence among attribute bundles, potentially yielding a richer set of data, as indifference or ambivalence frequently manifests itself as non-response in the CVM format. Reducing the incidence of protest zero bids is highly desirable in the valuation of health-related commodities, as many individuals will refuse to place a dollar value on human health in the CVM format. Moreover, conjoint results may be more reliable because individuals are more familiar with making decisions in the conjoint format (Mackenzie, 1990) and the trade-off process used in conjoint encourages respondent introspection (Johnson et al., 1995). Since price is treated as just one of several attributes, conjoint analysis may be particularly appropriate for valuing multi-attribute amenities associated with water quality. To this author's knowledge, only two conjoint studies of groundwater quality (Sparco, 1995 and Stevens et al., 1997) have been published and no comparisons of CVM and conjoint estimates of the economic value of groundwater quality have yet been published. However, several comparisons involving other commodities have been published.

Empirical evidence presented by Ready et al. (1995), Magat, Viscusi and Huber (1988), and Desvouges et al. (1983) suggests that CVM and conjoint results often differ dramatically. Desvouges et al. (1983) compared CVM and contingent rankings for water quality in the Monogahlea River. Mean water use values derived from direct CVM

questions were three to four times less than the values estimated from the contingent ranking approach. Magat, Viscusi and Huber (1988) used a paired comparisons approach, which is a form of conjoint analysis, and an open ended CVM format to derive consumers' WTP for risk reductions associated with a set of market goods (bleach and drain openers) that differ only in terms of purchase price and risk of injury. In contrast to the one-step, open-ended CVM elicitation procedure, the paired comparisons method asked each subject to make a series of comparisons between products in a manner that simulated actual choices in the marketplace. The direct CVM approach produced monetary valuations that were 58% lower than the average conjoint valuation.

Polychotomous choice formats (PC) may have an advantage over the more traditional dichotomous choice formats in that they allow respondents to express their strength of conviction and possible ambivalence toward a particular good; this additional information may yield a richer description of the individual's preferences. Mackenzie (1993) compared four types of conjoint formats to value waterfowl hunting trips: paired comparisons, polychotomous choice with three possibilities, contingent ordinal rankings of multiple alternatives, and contingent pseudo-cardinal rankings of multiple alternatives. Results of his analysis suggest that the contingent rating model had the greatest estimation efficiency, allowing for the narrowest confidence intervals for the WTP estimate. However, the four models yielded similar mean WTP estimates. Mackenzie also noted that the comparison of ratings across individuals is problematic in the contingent ratings model where no one rating level is specified to indicate indifference; some respondents may use the entire scale to indicate their ratings, while others may focus on a small interval of the scale. Ready et al. (1995) compared a dichotomous

choice CV format to a polychotomous choice (PC) format. Their DC CVM format asked respondents to decide whether or not they preferred a given program, while the PC format offered six choices (i.e., definitely prefer, probably prefer, maybe prefer, maybe not prefer, probably not prefer, definitely not prefer). In two empirical studies (preservation of wetlands and horse farms), the PC format resulted in a higher rate of usable responses compared with the CVM format. The polychotomous choice approach also resulted in much higher estimates of WTP for the two amenities. This study is discussed in more detail in the next section of this chapter.

Thus, although very few comparisons have been published, most of the willingness-to-pay estimates from conjoint analysis appear to exceed WTP estimates derived from the traditional CVM format. However, Brocks (1996) compared CVM and conjoint estimates and found that the magnitude of the difference between CVM and conjoint estimates depended upon model specification. In addition, the discrepancy between CVM and conjoint technique estimates may stem from a tendency on the part of respondents to pay less attention to the cost attribute of a particular good in the conjoint format than they do in the CVM format; although the price of the good may be clearly stated in the conjoint format, individuals may tend to discount the price attribute while rating or ranking alternative goods, viewing the rating or ranking task as more of an expression of preferences in the abstract rather than an actual commitment to purchase the good at its specified price.

3.4 The Role of Respondent Uncertainty in Stated Preference Studies

Due to problem of possible ambivalence and lack of commitment on the part of respondents as they consider valuation questions, much of the recent work using stated

preference methodology has attempted to incorporate respondent uncertainty into willingness-to-pay estimates. The need to consider respondent uncertainty in CVM estimates was noted by the 1993 NOAA panel on the CVM, which stated that the "valuation question should be posed as a vote on a referendum" and that "a 'no-answer' option should be explicitly allowed in addition to the 'yes' and 'no' vote options on the main valuation (referendum) question" (Arrow et al., 1993). If respondents are not allowed to express their uncertainty, they may feel forced to choose either the 'yes' or 'no' response, and hence their expressed value is not a meaningful reflection of their underlying preferences. However, some (Carson et al., 1994) argue against providing an "easy way" out for respondents, arguing that many who will opt for the 'no-answer' option, if it is available, are in fact capable of deciding to vote 'yes' or 'no' on the provision of a particular amenity. A respondent may choose to indicate uncertainty for a variety of reasons, including indifference between the 'yes' and 'no' choices, inability to make at decision at that time, preference for some other elicitation method, boredom with the survey and a desire to finish the survey, a lack of interest in the specific problem at hand, feelings that the cost of the amenity is too high, a desire for more information, or concerns about implementation of the program or the payment vehicle (Arrow et al., 1993 and Carson et al., 1994).

When 'no-answer' or 'don't know' responses are permitted, the researcher is left with the question of how to treat them in the data analysis. One may opt to drop the 'don't know' responses, but he or she will not only reduce sample size and lose information, but also runs the risk of introducing sample selection bias if the 'no-answer' respondents are significant different from those that do decide to vote 'yes' or 'no'

(Wang, 1997). A study by McClelland and Whittington (1994) suggests that respondents who choose the 'don't know' or 'no answer' option are more likely to be female and less educated. Another option conservatively treats 'don't know' responses as 'no' responses (see Carson et al., 1994). However, Wang (1997) shuns the use of this approach, arguing "common sense suggests that if a respondent is answering truthfully, a 'don't know' response is not the same as 'no'".

Consequently, recent research has grappled with the issues surrounding the treatment of respondent uncertainty, struggling to find ways to incorporate uncertainty into WTP estimates without throwing away valuable information or being overly conservative. As mentioned in the previous section of this chapter, Ready et al. (1995) used a polychotomous choice (PC) question format where the respondent had a choice of six responses to a single bid amount: definitely yes, probably yes, maybe yes, maybe no, probably no, and definitely no. The PC format was presented to half of the sample, while the other half of the sample received the traditional dichotomous choice (DC) format. Two surveys were conducted: one of willingness-to-pay to preserve wetlands and one of willingness-to-pay to prevent the loss of horse farms to development. The PC format yielded a higher usable response rate than the DC format (67.0% versus 60.0% for wetland preservation, and 58.0% versus 53.5% for horse farm preservation), but the difference was significant for the wetland study only.

Based upon responses to the polychotomous choice question, three logit equations were estimated to determine the probability of a 'definite yes' response (a PC response of 'definitely yes' or 'probably yes'), of a 'yes' response (a PC response of 'definitely yes', 'probably yes', or 'maybe yes'), and of a 'definite no' response (a PC response of

'probably no', or 'no'). Following Opaluch and Segerson's (1989) work on decision-making under ambivalence, an "ambivalence region" was constructed from the three logit equations. The region's lower bound is the highest dollar amount to which 50% of the respondents give a 'definite yes' answer, while the upper bound is the lowest dollar amount to which 50% of the respondents would give a 'definite no'. The ambivalence regions for both amenities were wide. For the loss of horse farms, the upper bounds ranged from $314 to $401 per household per year and the lower bounds ranged from -$23 to $142 per household per year, depending upon the amount of horse farms that would be lost (25%, 50%, 75%, or 100%). For the wetland preservation, the upper bounds ranged from $21.91 to $157.75 per household per year and the lower bounds ranged from $0.27 to $1.71 per household per year, again depending upon wetland scenario presented (the natural log of the wetland bid was taken, therefore precluding negative WTP estimates). The indifference bid is defined as the middle of the ambivalence region; this is the bid value at which half the respondents would give a 'yes' response and half the respondents would give a 'no' response. Indifference bids were calculated for both the PC and DC formats. For all wetland and horse farm preservation scenarios, the DC indifference bid is smaller than the PC indifference bid.

An additional logistic regression showed that the probability of a 'yes' response is significantly higher for respondents to the PC format than for respondents to the DC format. The results of this estimation, coupled with the disparity between DC and PC indifference bids, suggests that when confronted with a dichotomous choice format, respondents may conservatively respond 'no' if they are uncertain or ambivalent.

However, Ready et al. did observe DC format respondents who were willing to vote 'yes' even though the bid fell within the ambivalence region.

The authors conclude that incorporating uncertainty via the polychotomous choice format resulted in higher willingness-to-pay estimates than the dichotomous choice format. Median WTP for wetland preservations is $8.92 to $30.23 per household per year in the PC format, and $1.33 to $21.88 per household per year in the DC format (depending on wetland type), and median WTP for horse farm preservation was $254 to $295 per household per year in the PC format and $6 to $220 per household per year in the DC format. However, these results may be due to the study's lenient criteria for the coding of a PC response as a 'yes' response; the authors state "PC questions generate higher rates of 'yes' responses because the respondent can give a 'yes' response without making a strong commitment" and recognize that "greater ease in giving a 'yes' response may also give the respondent less inducement to consider the question carefully before answering" (Ready et al., 1995, pp.193). Lack of commitment on the part of the respondents to the PC format is also supported by larger confidence intervals for the indifference bids in the PC format than in the DC format. Thus while the PC format may have eased the valuation task for many respondents, thereby increasing sample size, it may inadvertently allow the respondents to be cavalier about supporting the provision of environmental amenities, without serious consideration of the money that they must sacrifice to do so. When confronted with a 'yes' or 'no' situation, respondents may be more likely to think about the forgone opportunities the monetary support of preservation necessitates, and thus WTP in the DC choice format may be more indicative of their true level of commitment to the particular environmental amenity in question.

Groothuis and Whitehead (1997) analyze the effect of a 'don't know' option in the dichotomous choice CV format, using data from both willingness-to-pay and willingness-to-accept studies. The willingness-to-pay (WTP) study is for a sea turtle protection program in North Carolina (see Whitehead 1992, 1993), and the willingness-to-accept (WTA) study estimates the compensation required to locate a hazardous waste landfill in a rural county of Pennsylvania (see Groothuis and Miller, 1994; Groothuis et al., 1998). WTP was elicited with a single bid, and three possible responses: 'yes', 'don't know' or 'no'. The WTA format was a referendum where respondents were offered a specific reduction in their state income tax in exchange for accepting a hazardous waste site in their county. In the WTP (sea turtle preservation) study, 32% of the respondents voted 'yes', 32% voted 'no' and 36% voted 'don't know'. Reasons for 'don't know' responses included a need for more information (37%), and a lack of financial resources (30%); it is likely that those who indicated they could not afford to pay would vote 'no' reality. In the WTA (hazardous waste site) study, 55% of the sample refused the offer, 16% were willing to accept the offer and 28% responded 'don't know'. Of those individuals that responded 'don't know', 5% indicated that their reason for the 'don't know' response was an objection to the question, while 95% suggested that they opted for the 'don't know' response because of uncertainty.

Responses from both the WTP and WTA studies were analyzed using multinomial and ordered logit models. The authors' initial conservative assumption was that in the case of the WTP study, a 'don't know' response was really a 'no' response, while in the WTA study, a 'don't know' response was really a 'yes' response. For the WTA model, estimation results lead to a rejection of the null hypothesis that the 'don't

know' response is similar to the 'yes' response. The authors suggest that forcing the respondents to choose or 'yes' or 'no' would have resulted in many refusals to answer the valuation question. Results from the estimation of the WTP model suggested a 'don't know' response really reflected a 'no' response; respondents in the WTP study who voted 'don't know' would likely vote 'no' if confronted with a traditional DC format. While Groothuis and Whitehead acknowledge that their research does not point to a clear-cut methodology for the treatment of 'don't know' responses, they argue that allowing a 'don't know' option is likely to be especially helpful to respondents that have been asked to consider a amenity with non-use values; respondents may be unfamiliar with the amenity and therefore may not have previously constructed WTP or WTA values. In this case, offering the 'don't know' option may prevent a large number of protest bids and non-response.

Wang (1997) disputes the assumption that each individual has a single WTP value in mind for the amenity in question, arguing that an individual's WTP for a specific good or amenity may be best characterized as "a random variable with an unspecified probability distribution". Reasons for the individual's uncertain preferences include questions about the product's quality or quantity, and unpredictable markets for substitute and complementary goods. Wang suggests that as more information about the amenity is presented, the variance of the individual's valuation function should decrease. A maximum likelihood procedure for the treatment of 'don't know' responses in referendum data is developed and applied to data from Whittington et al.'s (1994) survey on the economic value of environmental improvements in the greater Houston-Galveston Bay area in the state of Texas. This technique is based on a "threshold approach"; the

respondent will vote ‘yes’ if the bid falls below a lower bound to his distribution of WTP values and ‘no’ if the bid exceeds an upper bound to his distribution of WTP values. However, if the bid falls close to the mean of the respondent’s WTP distribution, the respondent is very close to being indifferent between voting ‘yes’ and voting ‘no’; consequently he may respond ‘don’t know’. Four model specifications were estimated: (1) Wang’s (1997) maximum likelihood procedure, (2) an ordered probit model (‘no’, ‘yes’ and ‘don’t know’ responses included), (3) a probit model (‘don’t know’ responses are treated as ‘no’ responses), and (4) a probit model where ‘don’t know’ responses are deleted. The estimated mean WTP values for the four models were (1) $11.86 per month, (2) $11.44 per month, (3) $2.65 per month, and (4) $10.24 per month. Treating the ‘don’t know’ responses as ‘no’ responses resulted in the lowest mean WTP estimate, while Wang’s method and the ordered probit model resulted in much higher, but similar, WTP estimates. While Wang’s approach holds promise for incorporating information about respondent uncertainty, he notes that the threshold technique can only estimate the mean of the respondent’s valuation function and cannot estimate the respondent’s entire valuation distribution.

Some researchers have attempted to incorporate respondent uncertainty into contingent valuation estimates by asking respondents to consider their degree of certainty immediately following their response to the valuation question. Johanannesson et al. (1998), in a study that was discussed earlier, found that hypothetical ‘yes’ responses overestimate real ‘yes’ responses when respondents are asked to state their WTP for a box of chocolates and are then given the opportunity to actually purchase the chocolates. However, when they restricted hypothetical ‘yes’ responses to include only the responses

of those who indicated they were 'absolutely sure' about their stated WTP, they found that this conservative method of dealing with uncertainty underestimated true 'yes' responses. Li and Mattsson (1995) employed a traditional dichotomous choice WTP contingent valuation question, which was followed by a "post-decisional" rating of the certainty of the response to the WTP question. This technique reduced both the mean and variance of willingness to pay. Champ, Bishop, Brown and McCollum (1997) used a split-sample approach to explore the polychotomous and dichotomous choice CV formats in a valuation study of a wilderness area in the Grand Canyon. Following the valuation question, respondents who received the DC format were asked to indicate how certain they were about their decision on a scale numbered 1 through 10. The traditional WTP estimate was $79, while the actual payment was $9. However, after all 'yes' responses with a certainty rating of less than '10' were coded as 'no' responses, WTP fell to $12. The results of this study suggest that the respondents' hypothetical responses reflected actual behavior only for those respondents who felt extremely confident about their responses.

Thus, while researchers seem to be reaching a consensus that respondent uncertainty needs to be addressed in the stated preference methodology, there is no clear-cut way to treat uncertain responses in the estimation process. In addition, the question of whether an uncertain 'yes' truly means 'yes' or 'no' is confounded by the possibility of free-riding behavior in the valuation of public goods; individuals may indicate that they are uncertain about their 'yes' response because they are harboring hope that someone else will come along and provide the amenity in question. Free-riding behavior may account for the conflict between the work of Johannesson et al. (1998), which

elicited WTP for a private good, and Champ et al. (1997), which elicited WTP for a public good. Current research in progress continues to explore the issue of how to treat respondent uncertainty in the valuation of public goods (see Alberini, Boyle and Welsh; Ekstrand and Loomis).

3.5 Willingness-to-Pay Estimates for Groundwater Quality

3.5.1 Averting Cost Studies

A few studies have estimated values for water quality using the averting cost method (see Spofford et al., 1989; Smith and Desvousges, 1986; Harrington et al., 1990; and Abdalla, 1989). Abdalla, Roach, and Epp (1992) studied averting measures in the borough of Perkasie, which is located in southeastern Pennsylvania. Trichloroethylene (TCE), a volatile organic compound, was detected in levels as high as 35 parts per billion in one of the borough's wells in late 1987. Since the contamination could not be immediately rectified, the borough was required to notify residents of the contamination in June 1998. As of December 1989, no solution had yet been implemented. Abdalla et al. conducted a mail survey to elicit increases in household averting expenditures taken in response to the contamination. To avoid the problem of joint products (e.g. expenditures on bottled water for taste reasons), individuals were asked about the specific actions they took in order to avoid TCE contamination. In addition to money spent on averting measures, the survey also asked about the amount of time the household had spent in an effort to avoid contamination, and about the factors influencing the household's averting expenditures, such as health perceptions, attitudes, and demographic factors. Individuals were also asked to indicate their perception of their cancer risk due to the contamination

episode. Surprisingly, only 43.2% of the respondents were aware of the TCE contamination in spite of the mandatory notification. Of the respondents who were aware of the contamination episode, 43.75% spent their money and time in an effort to avoid exposure to TCE. These efforts included: increased bottled water purchases in households that used bottled water prior to the TCE contamination, bottled water purchases by new users, the installation of home water treatment systems, bringing outside sources of water to the household, and boiling water. To avoid including the sunk cost of the home water treatment system in the estimation of averting costs, only a portion of the purchase and installation cost for the system was included in the estimation. The contamination period was 88 weeks long, which was estimated to be 0.17 of the system's useful life; thus, only 0.17 of the system's purchase and installation costs were included in the estimates.

Abdalla et al. (1992) calculated the costs of averting behavior for the sample of residents that undertook averting expenditures in an effort to prevent their exposure to TCE (133 respondents), and extrapolated to the total population of Perkasie residents. They estimated the total increase in expenditures from when TCE was first detected (December 1987) to September 1989 to be approximately $61,313.29 to $131,334.06. This range of estimates reflects whether the estimates valued time spent on averting measures at the minimum wage or at the respondent's estimated wage. Logistic regression estimation of the probability that household undertakes at least one adverting measure suggests that households were more like to engage in averting behavior if they received information about TCE, perceived themselves to have a high risk of cancer due to TCE exposure, and had children of ages 3-17 years living in the household. A

household that had children under the age of three years present was more aggressive in its averting behavior, as measured by the estimated increase in averting expenditures.

Abdalla et al. (1992) recognize that these averting cost estimates represent a lower bound to the value of the economic losses from the TCE contamination episode, but argue that in cases where a lower bound estimate is "high enough" to warrant a policy intervention, such as setting a maximum contaminant level for a particular contaminant, the extra time and money typically associated with stated preference techniques need not be expended.

This study is also important in its suggestion of the significant role that risk perception and the knowledge of groundwater contamination issues can play in the valuation of groundwater protection benefits. The important role of risk information and risk perception in studies using stated preference methodology is further discussed in Chapter Five.

3.5.2 Contingent Valuation Studies

Edwards (1988) estimated WTP for protection of groundwater from nitrates in Cape Cod, Massachusetts, obtaining mean WTP values of $363 to $1437 per household per year. Jordan and Elnagheeb (1993) also focused on nitrate contamination; for the state of Georgia, they obtained mean WTP estimates of $120.84 to $148.56 per household per year. Poe (1993) estimated mean WTP values of $224.72 to $684.95 per household per year for protection from nitrates in Wisconsin. Sun (1990) examined WTP values for protection from nitrates and pesticides, obtaining a mean WTP value of $641 per household per year in Doughtery Co., GA. McClelland et al. (1993) estimated a mean WTP of $146.76 per household per year for protection of groundwater

contaminated by landfills for US households. Schultz and Lindsay (1990) obtained a WTP of $129 per household per year in Dover, New Hampshire for protection of groundwater from general contamination. Powell (1991) elicited WTP values for groundwater resources in the Northeast. In the states of Massachusetts, New York and Pennsylvania, he obtained mean WTP estimates of $55.79 to $81.86 per household per year. Caudill (1992) obtained mean WTP estimates of $45.07 to $64.52 in Michigan in an analysis of the benefits of groundwater protection.

Boyle, Poe and Bergstrom (1994) conducted a meta-analysis of the eight groundwater quality studies discussed above; the econometric techniques used in meta-analysis allow for the control of the factors that differ between the studies. The point estimates from each study are treated as observations, and used as the dependent variable; this resulted in a sample of 52 WTP estimates. The independent variables attempted to capture the environmental commodity, respondent characteristics, and methodological procedures. The authors caution that the estimated coefficients should not be viewed as exact marginal values, but are more generally useful due to the information provided by their signs and significance. The authors note that their analysis is limited by the lack of a consistent definition of groundwater contamination across the eight studies. However, their results show that the following variables positively and significantly affect willingness-to-pay for groundwater quality: the mention of nitrates as a potential contaminant, an increase in the probability of contamination, income level of respondents, and the mention of a possible shortfall in the water supply due to contamination. On the other hand, the following variables negatively and significantly affect willingness-to-pay for groundwater quality: the mention of substitute sources of

potable water, a program that contained contamination, and studies that focused primarily on the use values associated with groundwater quality.

3.5.3 Conjoint Analysis Studies

Stevens, Barrett and Willis (1997) attempted to determine both use and non-use values associated with groundwater protection in Massachusetts using conjoint analysis techniques. Their 1995 mail survey of western Massachusetts residents presented information about five options for groundwater protection: an aquifer protection district, a town-wide water treatment facility, a private pollution control device, the purchasing of bottled water, and doing nothing (maintaining the status quo scenario). These options were characterized by three attributes: the method of protection, cost to the individual, and length of payment. Three types of conjoint models were estimated: traditional conjoint (where the rating for a particular program is the dependent variable), a ratings difference model, and a binary response model. The estimation results indicate that the binary response model produced lower mean WTP values than the ratings difference or traditional models. In addition, respondents were willing to pay more for the aquifer protection program (regardless of model specification), suggesting that the respondents were willing to pay extra if the groundwater source was protected, which may be indicative of significant non-use value for groundwater protection. However, this study did not specify the type of contaminants the various protection options would address, nor did it specify the status quo level of risk, or the specific risk reduction each option could achieve.

3.6 Summary

This collection of studies points to a further research agenda in the area of the valuation of groundwater quality. The relative strengths and weaknesses of the CVM and conjoint analysis to value the benefits of groundwater quality need to be further explored. Most of the prior groundwater quality valuation studies were conducted using the contingent valuation framework. None of the groundwater quality studies have attempted to incorporate possibility respondent uncertainty into willingness-to-pay estimates.

Prior groundwater valuation studies have either emphasized a single type of contamination or failed to specify the type of contaminant or contaminants respondents would be protected from. Many of the prior valuation studies of groundwater quality have focused on nitrate contamination, and there is a need to value health effects resulting from other types of groundwater contamination, such as chemical contamination and excessive sodium levels. In addition, with the exception of Poe (1993), none of the studies provided respondents with information about the actual levels of contaminants in groundwater. Research needs to be focused on obtaining willingness-to-pay estimates for groundwater that do not simply reflect the value of a general change in groundwater quality; society's value for specific groundwater protection policies must be examined. Potential groundwater protection programs must be clearly defined in terms of the particular contaminants they guard against, and the magnitude of risk reduction they can achieve. In addition, individuals need to be informed of the health risks posed by the specific contaminant or contaminants addressed by the proposed program. Only under these conditions can researchers hope to obtain meaningful willingness-to-pay values for groundwater protection.

More generally, although the contingent valuation method has been heavily researched in the last two decades, there remain many unresolved issues in the area of non-market valuation. Further comparison of the contingent valuation and other stated preference methods (such as conjoint analysis and QALY analysis) is necessary to explore the relative strengths and weaknesses of each method in the valuation of different types of non-market goods (e.g. environmental amenities, health risks, mortality).

Research concerning the problems associated with stated preference methodology reminds us that every action involves a trade-off. At this time, it appears that stated preference methods other than the CVM hold the most promise for the valuation of health related benefits, as evidence suggests that individuals may have considerable difficulty making trade-offs between money and health in the traditional CVM format. Stated preference methods such as conjoint analysis may result in more usable responses, as they seem to mitigate protest bids and non-response. However, ranking or rating elicitation methods may have the drawback of allowing the respondent to discount the price of the good in question and her budget constraint as she considers what she most "prefers". In addition, the question of how to treat respondent uncertainty persists: do we simply adhere to a strict dichotomous choice question and risk overestimates of WTP, protest bids, and non-response, or do we allow for respondent uncertainty via polychotmous choice and other methods designed to measure the respondent's degree of certainty, possibly encouraging lack of commitment and forcing us to confront the question of just how certain one has to be for his response to count as a vote of support for the amenity in question? Finally, as described in this chapter, there are myriad biases that can result from survey research; so many things can go wrong that the researcher

may feel paralyzed by the fear of inducing bias, recognizing that particular care to avoid one type of bias may increase the likelihood of another type occurring. The message of this chapter is clear: trade-offs in survey research are inevitable. The best the researcher can do is to be aware of the trade-offs she is making and the questions that remain unanswered, while embracing the spirit of "learning by doing".