

# Current issues in environmental economics

*edited by* Per-Olov Johansson,  
Bengt Kriström *and* Karl-Göran Mäler

Manchester University Press  
Manchester and New York

*Distributed exclusively in the USA and Canada by St Martin's Press*

## Current issues in the design, administration, and analysis of contingent valuation surveys

### Introduction

Contingent valuation has evolved and matured considerably during the past 30 years since Davis' pioneering study (1963) on hunting in the Maine woods.<sup>1</sup> During this period contingent valuation has been shown to be a useful tool for learning about preferences for public goods and has increasingly become accepted for use in policy analysis. The method has evolved from an experimental technique of interest primarily to scholars to an operational technique of interest (or concern) to a much wider set of people. In the United States, those interested in the method now include actors on both sides of battles over the level of environmental regulatory standards and the size of natural resource damage awards. This has politicised the environment in which CV studies are conducted in the United States as parties with an interest in these matters defend or denigrate the method's ability to obtain valid and reliable measures of the public's willingness to pay. While a politicised environment has its drawbacks for conducting a scientific debate, it has the potential for advancing the method by posing questions and stimulating high quality research.

In this chapter we address some issues in the current state of the art in the design, administration, and analysis of contingent valuation surveys. We build on and extend the analysis of the method we presented in our book, *Using Surveys to Value Public Goods: The Contingent Valuation Method* (Mitchell and Carson, 1989). Our thinking has been strongly influenced by three recent events and studies. The first is the state of Alaska's contingent valuation survey which sought to measure the natural resource damages caused by the 1989 Exxon Valdez oil spill in Alaska (Carson et al., 1992). The second is the Exxon symposium (Hausman, 1993) where a series of papers, several based on empirical studies, were presented which were highly critical of contingent valuation and its use in natural resource damage assessment. The third is the report submitted by the Blue Ribbon Panel convened by the US National Oceanic and Atmospheric Administration (NOAA) to assess the reliability of contingent valuation

for use in assessing lost passive use values resulting from oil spills (Arrow et al., 1993).

Although the use of CV in natural resource damage assessment is the focus of an intense debate in the United States,<sup>2</sup> recent CV studies have valued a wide variety of non-market goods and the principal use of the technique is for the assessment of the benefits of government programmes. We begin by describing these broader developments before reviewing, briefly, the three events mentioned above which provide a background for understanding the current debate over the method in the United States. We then discuss two misconceptions concerning the nature and use of contingent valuation which continue to cause confusion. The remainder of the chapter is devoted to variety of methodological issues which are important in the design, administration, and analysis of contingent valuation surveys.

### Recent developments in contingent valuation

In the last few years, contingent valuation has become the mostly widely used approach to value public goods. The number of contingent valuation studies and papers has now grown so large that it is impossible to say anything definitive about contingent valuation as a single well defined 'method'; indeed it is perhaps best to think of contingent valuation as a basic approach to the valuation of non-marketed goods which works by constructing a market in the context of a survey interview. Figure 2.1, based on a comprehensive bibliography (Carson et al., 1994) which has been compiled with the assistance of other scholars, depicts the production of contingent valuation literature from 1972-92. The most recent version of this bibliography contains over 1600 citations.

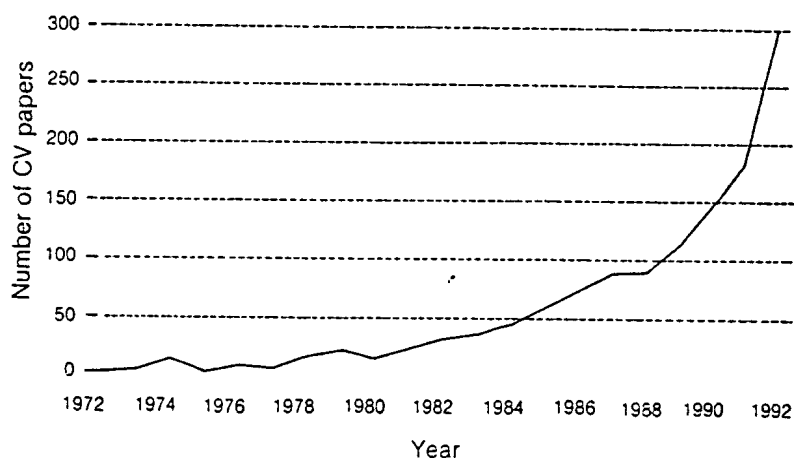


Figure 2.1 Production of CV literature by year, 1972-92

The range of goods which have been valued using contingent valuation is quite large. Our bibliography includes studies which value changes in: the provision of various types of water and air quality, recreation of all types, risk from drinking water, groundwater contamination, wilderness areas, wetlands, tropical rainforests, endangered species, non-environmental human health issues ranging from risks from blood supplies to the size of queues for health care, ecotourism, water supply, sewage treatment, highway safety programmes, senior companion programmes, museums, electricity and water supply reliability.

Contingent valuation studies, which once had a predominantly American flavour, have now been conducted in over forty countries demonstrating the adaptability of the approach to different cultural contexts. Navrud (1992) provides an overview of the recent use of contingent valuation in Europe. CV work in developing countries (e.g., World Bank Water Demand Research Team, 1993) presents many challenges but also has some advantages in terms of substantially lower survey costs and higher response rates. Shyamsundar and Kramer (1993) have even recently conducted a contingent valuation study in Madagascar where the value used was bushels of grain rather than money.

Much of the demand for contingent valuation studies is a derived demand. Policy makers increasingly want and perhaps are becoming increasingly comfortable with monetary values for public goods such as environmental amenities. This demand has resulted in a growing number of carefully done, high quality studies. Inevitably, it has also resulted in a number of studies marked by one or more problems that can befall a contingent valuation study such as: a vague or unclear description of the good, lack of key information about context and substitutes, lack of attention to ways respondents could misperceive the good, implausible or overly hypothetical scenarios (particularly with respect to provision of the good and payment for it), willingness to pay responses which cannot be predicted by the available covariates, inadequate sampling procedures, poor response rates, and sample sizes which are too small for the purpose for which they were intended.

We anticipated the quality problem in a 1988 paper (Mitchell and Carson) entitled, 'How Far Out on the Learning Curve Are We with Contingent Valuation?' In that paper we argued that contingent valuation was past the Stage I (invention) and Stage II (experimental implementation) phases and at that time lay somewhere between Stage III (initial policy implementation) and the final Stage IV (routine implementation). We observed that 'skimping on survey development cost, pretesting, interviewer quality and training, or sample size can easily lead to very sizeable benefit mis-estimates'<sup>3</sup> and expressed a concern that 'poorly conducted studies or studies based on erroneous assumptions may produce embarrassingly implausible estimates which could result in bad policy or serve to

discredit the [CV] method'. This concern arose because contingent valuation studies can appear to be deceptively easy to design and execute to those who lack experience and training in survey research.<sup>4</sup> Moreover, even experienced CV researchers will be unable to obtain valid willingness to pay estimates of sufficient quality for policy purposes if sponsors fail to provide an adequate budget or insist on a deadline which allows insufficient time to carefully develop and test the CV scenario. These 'resource' problems are amplified when the method is extended to new subject areas where the researchers have no body of prior CV experience on which to draw.

What we did not anticipate in our earlier writings was the sharp growth in the literature which argues that contingent valuation is a fundamentally flawed approach no matter what good is valued or how the study is conducted. In retrospect, this objection was an inevitable by-product of the method's acceptance by policy makers and its use in litigation coupled with the ingrained scepticism some economists have about the reliability of survey responses. Helping to fuel this scepticism is the greater attention given to papers which find anomalies and highlight problems relative to papers which 'merely' present benefit estimates from large, well tested contingent valuation surveys.

#### **The Exxon Valdez case**

The need to assess the harm caused by the 1989 Exxon Valdez oil spill in monetary terms focused intense attention on contingent valuation once the government, both state and federal, announced its intentions to undertake a contingent valuation survey for this purpose. Both the government and the Exxon Corporation assembled large groups of scholars to work on contingent valuation issues. Those the government retained to undertake the study were, for the most part, specialists in resource economics and contingent valuation. The Exxon Corporation's team consisted primarily of researchers from outside resource economics. The Exxon researchers conducted a variety of studies designed to support the Exxon legal position that it should not have to pay compensation for lost passive use values and, in any event, that contingent valuation was not a reliable<sup>5</sup> method for valuing natural resource damages. Exxon and the government reached a pre-trial settlement with Exxon paying over two billion dollars in spill clean-up costs and agreeing to pay an additional one billion dollars for natural resource damages.

After the case was settled and while the federal government was drafting a new set of regulations designed to prevent future oil spills, Exxon released some of its studies at a two day symposium it sponsored in Washington in April, 1992. The symposium papers were subsequently published, along with some of the remarks made at the symposium, in a

1993 volume edited by Hausman. Some of these studies were based on empirical research; others were literature reviews or legal analysis. The empirical studies involved telephone and shopping mall intercept surveys and, in one case, face-to-face cognitive interviews. Their authors reach conclusions which reject the use of CV for use in either benefit or measuring damage assessment, on the grounds that their studies show the method to be incapable of measuring nonuse values with sufficient reliability. These studies have been criticised by ourselves and others (Carson and Flores, 1992; Randall, 1993; Smith, 1993; Hanemann, 1994) on a number of grounds.<sup>6</sup>

In November 1992, the state of Alaska released the CV study it had commissioned for possible use in its natural resource litigation with Exxon. This study was conducted over a three year period by a team of scholars which included the present authors and other survey specialists and resource economists. Because it was intended to be used in court where it would necessarily be subject to intense scrutiny by Exxon's experts, the sponsor and those working on the study had a strong incentive to produce a study which would meet high professional standards and a budget commensurate with this goal. The researchers undertook an extensive programme of instrument development involving focus groups, in-depth interviews, pretests and a series of pilot studies during which they explored the sensitivity of the results to a number of design features and considerations. This work (Carson et al., 1992) contains a number of empirical findings of general interest to CV researchers and led to the development of several new design and analysis techniques. In the main survey (hereafter referred to as the Alaska Survey) 1043 respondents, comprising a national area probability sample, were interviewed, face-to-face, by a leading survey firm's professional interviewing staff. A manuscript reporting the findings of this study is currently being prepared for publication.

The third event in this sequence was the convening of a Blue Ribbon panel by the US National Oceanic and Atmospheric Administration (NOAA), the federal agency responsible for issuing damage assessment regulations for oil spills. The NOAA General Counsel, having been presented with diametrically opposing views about the suitability and reliability of contingent valuation for this purpose, invited Kenneth Arrow and Robert Solow to chair a Blue Ribbon Panel whose charge was to advise the NOAA regulation writers on this issue. Edward Leamer, an econometrician, Paul Portney, a resource economist, Roy Radner, a micro-economic theorist, and Howard Schuman, a survey researcher, comprised the rest of the panel, none of whom had previous direct involvement of great extent with contingent valuation. In the course of its work, the NOAA Panel reviewed a large body of material on contingent valuation in general and, in particular, its use to measure existence or passive use values. This material included the relevant journal literature, reports of current appli-

cations including both the Exxon sponsored studies and the newly released state of Alaska study, formal comments submitted to NOAA by parties on both sides of the issue, and oral testimony presented in an open hearing held in Washington. The panel drew the overall conclusion, 'CV studies can produce estimates reliable enough to be the starting point for a judicial or administrative determination of natural resource damages - including lost passive use value' (Arrow et al., 1993).

A careful reading of the NOAA panel report reveals that the panel was concerned by the potential problems that could occur if a contingent valuation survey was poorly designed or executed. The Panel put forth a number of guidelines and recommendations to address issues such as the possibility of an embedding effect and to help ensure the reliability of contingent valuation. In taking this approach the Panel, after a careful review, affirmed that a good quality CV survey can produce reliable information on passive use values for use in policy making.

### **Two misconceptions about contingent valuation**

Two misconceptions tend to fuel the debate over CV in the United States. The first is that contingent valuation willingness to pay questions are no different from other willingness to pay questions. The second is that all contingent valuation studies can be treated as if they are equally reliable. We discuss each of these misconceptions and their implications in turn.

#### *A contingent valuation scenario consists of more than a WTP question*

It is not always understood by those who are not actively involved in conducting contingent valuation research that the CV method involves more than just asking a respondent to express a willingness to pay value for some good. Nowhere is this misunderstanding more evident than in the work of Daniel Kahneman (e.g., Kahneman and Ritov, 1993) and some of the work of the CV critics retained by Exxon (e.g., Kemp and Maxwell, 1993).

Consider two surveys. Survey A asks respondents whether they would pay \$50 for 'significantly cleaner air'. In this survey the good that the respondent is asked to buy is vaguely described, the payment vehicle and schedule is undefined, and there is no description of how and where this good would be provided. Survey B asks respondents if they would vote for a programme that would cost them \$50 in a one time payment to bring about a specific change in air quality (with relevant factors such as location, base quality, and nature of change described in understandable terms). The Survey B question is preceded by interview material which presents detailed information such as the context in which the change would occur such as how the good would be provided and paid for, what substitutes are available.



Survey A type questions have from time to time been asked in public polls. They are best referred to as *attitudinal* willingness to pay questions to distinguish them from contingent valuation WTP questions. An attitude is a subjective judgment of an object which is usually based on some scale such as good-bad, strong-weak (Ajzen and Fishbein, 1977). The correspondence between type A WTP questions and attitudinal questions is twofold. First, the good the type A questions ask about is similar to the type of vaguely defined referent ('clean air', 'trust in government', 'protect wilderness', 'fear of crime') used in polls that measure attitudes. Second, the 'willing to pay' aspect of the question lacks realistic information about the market in which the person would pay for the good. In the absence of this information, asking about 'payment' is simply another way to measure how 'important' the good is to the person.<sup>7</sup>

Instead of a general *attitude* towards a good or class of goods, type B (contingent valuation WTP) questions measure a *behavioural intention* towards paying for a *particular* good. To predict successfully, a behavioural intention question must be as congruent as possible with the purchase decision in terms of what is bought and the context in which it is bought (Ajzen and Fishbein, 1977). In the case of public goods, this means, first, that the specific features of the good must be described and, second, that it must be presented in a clearly defined market scenario. Respondents need to know what they are buying, who is going to provide it, how it would be provided and how the decision on whether to provide it will be made. Only questions that create a realistic market for a precisely defined good can measure the type of income-constrained behavioural intention information suitable for use in economic analysis. The description of the good and the market context together comprise the conditions on which the willingness to pay decision is contingent; hence the method's name.

The distinction between the two types of WTP questions is important because some studies, which purport to test the CV method through split sample experiments such as Kahneman and Knetsch (1992), are based on the assumption that an attitudinal WTP question is adequate for this purpose. The unfortunate result is that these studies use what might be called a 'pseudo' contingent valuation willingness to pay question because they lack the two characteristics necessary to predict a consumer's willingness to pay for a public good.

The first is a clearly defined good. Kahneman and Knetsch, for example, asked Vancouver respondents in a telephone survey to consider independently each of a series of social or environmental 'causes' and to indicate 'the most that you would be willing to pay for each'. A typical description of one of these goods reads in its entirety: 'Improve sport fish stocks in British Columbia fresh water'. Another, presented to a different sample, reads: 'Improve sport fish stocks in Canada fresh water'. Left out of these brief descriptions are the types of specific conditions which one would



expect to see in a good contingent valuation study, such as what is meant by the rather technical 'fish stocks', what their current condition is, the size of the offered improvement, how long the improvement would be maintained, and information about the 'fresh water' (lakes? rivers? both?) where these fish stocks reside. Without such information, the description is too vague to be economically meaningful.

Second, studies which attempt to test the CV method by using attitudinal WTP questions fail to create the necessary context for a meaningful valuation decision which is a realistic market for the good. The Kahneman and Knetsch study does not say who will provide the good, how it will be provided, when it will be provided, and how the respondent will be expected to pay for it. These are the types of conditions a citizen would want to take into account in deciding whether or not to vote for a good's provision in a referendum. Their absence transforms the choice from a realistic one to a symbolic one.

Three things follow from the important role contingent features play in a CV survey. First, if the description of the good or its market context is changed in an important way, it is *expected* that this could change the value the respondents hold for the good. Consider a contingent valuation study of changes in risk from drinking water contamination. A change in the agency the respondents are told would be responsible for carrying out the programme could affect people's willingness to pay; if one agency (e.g., state Government) is less trusted than the other agency (e.g., the Federal Government). This is because respondents who are told the Federal Government would be in charge might be more likely to think they would get the change if they paid for it than those who are told the state would undertake the programme if it is approved. However, if most respondents perceived the two agencies as equally or nearly equally competent to carry out an effective drinking water programme changing this scenario feature probably would not make a difference in respondents WTP amounts. It is obviously important to know which aspects of the description are important and which are not. These often will vary from study to study and it requires careful research during the design phase of a study to identify them.

Another type of change that would not make a difference is changing the size of the commodity without reaching the threshold where a respondent feels he or she is valuing a good that is significantly smaller or larger. For example, a change in a good's description from fifty-five lakes in a given area that would be protected in a specified way from the effects of acid rain to forty-five lakes, might receive a similar per cent of 'yes' votes at a given price from split-samples if respondents did not perceive this to be an important difference. If this view is shared by enough people in the sample the study could conclude that the two outcomes have approximately equal value.

Second, whenever scenarios fail to provide information about contingent elements that are important to respondents the respondents will tend to fill in the information gaps by making default assumptions (Fischhoff and Furby, 1988). Default assumptions are guesses about such things as the likelihood that the good will be provided, what will happen if the good is not provided, what substitutes are available, whether endangered species are involved or not, how long they would have to pay the amount they agreed to, etc. Having respondents make private default assumptions about important scenario features can render a CV study's estimates uninterpretable because there is no way of knowing what they were buying; even if this was known, it would differ from one respondent to another. For example, if a contingent valuation study of wildlife injuries was conducted in the United States and the researcher does not make clear whether any of the species are endangered, it is likely that some respondents would assume some species were endangered and others would not. The former are more likely than the latter to be willing to pay to prevent this type of harm.

Finally, if a good is only described briefly in a CV survey which lacks specificity and market context, it should be no surprise if respondents show gross insensitivity to the scope of the good. Asked about a vague good, the respondent is likely to give a vague answer. This is especially likely if the survey requires the respondent to make a series of willingness to pay judgments in a single survey as was done in several of Kahneman's experiments (Kahnemann, 1986; Kahneman and Knetsch, 1992; Kahneman and Ritov, 1993; Kahneman et al., 1993) because such a format invites quick, relatively unconsidered judgments.

Assuming that attitudinal willingness to pay questions are the same as contingent valuation willingness to pay questions is not a trivial misconception because the two types of questions have very different properties which render the attitudinal questions invalid for measuring economic values. CV proponents have never recommended the use of attitudinal WTP questions for this reason. As attitudinal WTP questions are likely to be vulnerable to the types of effects CV critics test for in their experiments, the use of attitudinal WTP questions to test the contingent valuation method effectively stacks the odds against contingent valuation.

*All CV surveys are not equally reliable*

The second common misperception about contingent valuation studies is the assumption that the quality of one study is as good as another or, as critics would have it, one study is as bad as another. This assumption naturally leads the CV critics to the position that if any CV study shows significant weakness, the CV approach itself is thereby discredited. As Randall (1993: 14) has pointed out, this view amounts to asserting that 'CV results are invariant to CV practice', an obviously absurd premise.

Unfortunately, this misperception is by no means confined to CV critics. Sometimes those who favour the use of the method appear to give uncritical acceptance to the results of any contingent valuation study. This happens when policy analysts use a contingent valuation estimate without due consideration to the quality of the study, or when the values from different contingent valuation studies are averaged to arrive at a consensus estimate, without regard to large differences in such things as the quality of the survey instrument, sampling, and sample size.

Another manifestation of this misconception is when it is assumed that the presence of a set of reference operating conditions (ROCs) (e.g., Cummings et al., 1986) regarding the nature of the economic entity being valued *guarantees* reliable estimates without regard to how the study was designed and implemented. Examples of ROCs are that respondents 'must understand, be familiar with, the commodity . . .', 'must have had (or be allowed to obtain) prior valuation and choice experience with respect to consumption levels of the commodity' and must elicit WTP rather than WTA values (Cummings et al., 1986: 104). As we will show in the next section, conditions such as these do not guarantee success. At best a set of reference operating conditions may serve as a useful warning about potential problems to which the survey designer needs to be sensitive.<sup>8</sup>

The Blue Ribbon Panel did not fall into this misconception. Rejecting suggestions that it recommend limiting the use of contingent valuation only to situations where respondents were familiar with and had prior experience with the environmental good in question, the Panel instead highlighted the importance of good methodology for contingent valuation studies. Several pages of its report are devoted to a careful review of methodological issues and its specific recommendations take the form of a series of methodological guidelines which the Panel believes will enhance the reliability of a CV study. They include the use of face-to-face interviews, careful pretesting, conservative design, the referendum format, accurate description of the programme or policy, and a reminder of substitutes.

The NOAA Blue Ribbon Panel guidelines set a very high (and very costly) standard that contingent valuation studies should meet if they are to be used in natural resource damage assessment under the US Oil Pollution Act of 1990. While many of their guidelines are appropriate *for this use*, it is reasonable to ask whether it is sensible or even rational for all contingent valuation studies to adhere closely to them, no matter what policy purpose the study seeks to serve. We believe a policy maker looking at the different sources of uncertainty that affect a decision of interest should ask questions about the cost and benefits of reducing those different sources of uncertainty. Clearly, the more important the stake in the decision and the more contentious it is (which will almost always be the case in litigation), the more resources should be invested in the CV study to help ensure its reliability.

Providing an adequately specific description of the good to be valued and its market context as described above is a necessary but not sufficient condition for a reliable CV study. The CV survey designer must also successfully *communicate* this information to the respondent. This raises questions of questionnaire design and implementation, a topic to which we now turn.

### Key methodological issues

Survey researchers are well aware that just because respondents answer a question does not necessarily mean that they have grasped its meaning. Researchers know that differences in such things as the CV designer's attentiveness to possible sources of bias, the choice of the information provided, and the wording of the survey instrument can have large implications for the reliability of WTP estimates as can the choice of sampling frame, the mode of administration, the response rate, and the analytical techniques.<sup>5</sup> While there is typically no single 'correct' decision about these survey design issues, survey practitioners draw on a body of past experience as well as on research specially conducted during the design phase of a study to determine which approaches and techniques are likely to lead to reliable contingent valuation outcomes. In this section we discuss some key methodological issues, primarily in the design and administration of contingent valuation surveys.

#### *Scenario design*

In designing a contingent valuation scenario, there are two primary goals. The first is to ensure that respondents clearly understand the characteristics of the good and the context in which it is being offered. The second is to ensure that respondents find the CV scenario plausible, especially as it relates to payment for the good and its method of provision.

Understandability, the first goal, requires the good and its context to be described with sufficient specificity and clarity that respondents understand what they will get for their money if the good is provided. We have already discussed the need for specificity and how attitudinal WTP questions fail the specificity test. The clarity dimension comes into play once specificity is achieved because it can be a difficult task to communicate the scenario (which includes the description of the good and its context) to respondents who vary in educational levels and personal background. One obvious, but common, mistake is to use technical terms that an economist or a scientist may be comfortable with, but which are poorly understood by some of the general public. Confusing wording, information overload, and inattention to respondent beliefs which interfere with communication can also affect respondent understanding. It is the goal of preliminary research to develop and test a description of the good that is understandable.

It is important to note that prior experience with a good is not required as some have argued for respondents to understand its characteristics. While it is often easier to describe a good to a respondent who has had experience with it, this is not always the case. For instance, it may be easier to convey to a respondent who has not visited a recreational site that there is a toxic materials problem there that may need to be cleaned up than it is to convey the same information to a respondent who has visited the site frequently without being aware of the problem. Even in cases where those with experience may easily grasp the nature of a good, there are ways to convey it clearly to the inexperienced.

Words alone are often insufficient to describe a CV scenario unless the good is already familiar to respondents. We have found visual aids such as maps, pictures and diagrams enhance respondent understanding and attention and help ensure that the good being sold is not confused with other public goods. For example, land that is proposed to be purchased for a new wildlife refuge can be described by showing maps and pictures of the area and describing the wildlife and the services it would provide for them. Visual aids are especially useful in conveying complicated concepts such as changes in the provision of the good over time.

Understandability can be enhanced by describing the context in which a good provides services. On this issue, the NOAA Blue Ribbon Panel recommended a 'high standard of richness in context to achieve a realistic background' (Arrow et al., 1993: 4608). Since context can involve many dimensions, the designer should focus on those context features that preliminary research shows are likely to influence the value respondents place on the good. For an air quality good this would certainly involve information about whether any human health improvements would occur if the good is provided. Information about the availability of close substitutes is an important context feature for goods of all types. In the Alaska Survey we used the show card shown in Figure 2.2 to help communicate the relative scarcity of each of the bird species which had a large number of individuals killed by the spill. Left to their own default assumptions, respondents might erroneously have assumed that the bird species with the largest number of deaths, the murre, was seriously threatened by the spill. The possibility of spending money on other types of public goods is another important type of substitute.<sup>10</sup> One way of doing this is to ask a series of attitudinal questions about these goods at the beginning of the survey. These questions can serve as a good 'warm up' and, in addition, sometimes prove to be good predictors of willingness to pay.

A unique feature of many CV surveys compared with most other types of surveys is the amount of information that must be conveyed to respondents in order to adequately describe the good, its context, and the market in which it is offered. Because there are cognitive limits to the information burden that can be imposed in a survey, the survey designer needs to

## Card 4

## Bird species affected by the 1989 Alaska oil spill

<i>Species</i>	--In the entire spill area--	
	Number of dead birds recovered (rounded)	Estimated population before the spill
Murres	16,600	350,000
Sea ducks	1,150	100,000
Murrelets	1,150	50,000
Cormorants	1,050	30,000
Pigeon guillemots	500	20,000
Kittiwakes	400	10,000
Grebes	350	8,000
Loons	300	3,000
Storm-petrels	300	300,000
Fulmars	250	150,000
Gulls	200	100,000
Bald eagles	100	5,000
Other sea birds	250	300,000
<i>Totals</i>	22,600	1,516,000

Figure 2.2 Bird species affected by the 1989 Alaska oil spill

Source: Carson et al. (1992), 'A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill', Report to the Attorney General of Alaska.

minimise the information and convey it in a way that maintains respondent interest and attention. In order to avoid information overload in the Alaska survey we conducted extensive research to discover which information was essential for the respondent to have in order to accurately understand the



survey scenario and which was not. We still found it necessary to devote twelve pages of the instrument<sup>11</sup> to describe the area, the effects of the original spill and a programme to prevent a second spill that would have an equivalent impact on Prince William Sound's natural resources. The following are some of the techniques we found helpful in maintaining respondent interest and attention during the presentation of this material:

- \* Using simple language, eliminating unnecessary words, and presenting the material in a sequence that was logical to the respondents.
- \* Interspersing the narrative with occasional questions to involve the respondent.
- \* Using in-person interviews.
- \* Carefully training the interviewers to read the material at an appropriate pace, using conversational inflection, pauses and eye contact throughout the interview.
- \* Using various visual displays to illustrate the text. These included four maps, ten photos, and four cards showing information or a diagram to illustrate the text.

In addition to designing a scenario that respondents will understand, the researcher must also make it believable. A plausible or realistic scenario will be taken seriously by a respondent as presenting a credible choice situation which could or will happen.

The contingent valuation method has sometimes been criticised for eliciting hypothetical values, the implication being that the necessarily hypothetical nature of a CV survey inevitably produces unreliable WTP amounts. This belief confuses formal hypotheticality with substantive hypotheticality. In the formal sense, CV respondents are well aware that they are not actually voting when they tell the interviewer how they would vote (in a study that uses a referendum scenario). What matters is the substantive impression created by the scenario. This may be explicitly hypothetical or it may be realistic.

An example of the former are the types of thought experiments beloved by economists which preface a choice situation with 'Imagine', 'What if?' 'Suppose?' or 'Presume a . . .'. However plausible it may be to professional economists, this type of framing suggests to many respondents that the choice is of little or no personal consequence, a judgment that mitigates against an economically meaningful answer to the willingness to pay question.

Realism, a scenario characteristic that the Blue Ribbon Panel regarded as important (Arrow et al., 1993: 4607), is another matter. Realism involves *simulating* a plausible real life choice situation with sufficient verisimilitude that respondents take it seriously. But what kind of real life situation? Buying something in a shop? Environmental amenities are not sold in shops but voters do make decisions about taxing themselves to provide these types of amenities. The fact that referenda are not uncommon in real life in



the United States (nor are surveys which attempt to predict their outcome) led the Blue Ribbon Panel to recommend the CV designers use the referendum format on the grounds of its realism. That Americans today resent paying taxes and are reluctant to increase their tax burden helps ensure that respondents do not take the decision to tax themselves lightly.

A scenario can use the referendum format and still fail to simulate a real situation if other important scenario features, such as the provision mechanism, lack realism. While it may be attractive to an economist to value the good in the abstract, the need to create a plausible provision mechanism means that one will almost always value the policy to provide the good. The trade-off involved in describing a policy is that whatever provision policy is chosen, some of its features will almost always be unattractive to some respondents and hence the willingness to pay responses for the good will be less than they would be for the good if its delivery could just simply be guaranteed. There are two main reasons for this.

The first is that a respondent may find the payment mechanism less than optimal from the perspective of the respondent's financial situation and sense of fairness. As would be expected in scenarios that respondents take seriously, they frequently protest when asked to pay for things they consider to be the responsibility of other parties. These parties are most frequently a specific business, business in general or, occasionally, another level of government. While some protest zeros are inevitable, it is obviously desirable to keep them to a minimum. In order to do this, the scenario must make plausible the obligation of the public to pay if the good is provided. This was a difficult task in the Alaska Survey, as many respondents felt it was the oil companies' responsibility to prevent future spills. After experimenting with several types of payment vehicles, we found 'a special one time charge in addition to your regular federal taxes' worked best. Protest zeros were minimised by telling respondents that the oil companies that take oil out of Alaska would also pay a one time tax out of their profits to help pay for the programme.

The second reason why WTP amounts tend to be reduced because of the need to specify a credible payment mechanism is that some respondents may lack faith that the supplier of the good specified in the policy can or will actually deliver the good with 100 per cent probability and in an efficient manner.<sup>12</sup> In general this problem is much more pronounced for large goods than small ones for which provision difficulties are often not seen as real obstacles. Disbelief that a large good can be delivered, such as a major reduction in national air pollution, can cause respondents to be willing to pay a lower amount for it than for a smaller good such as a major reduction in a local city's air pollution. The job of the survey designer is to find a provision mechanism which minimises these undesired effects.

Specific features of a provision mechanism also may present plausibility difficulties. In the Alaska Study we explored the possibility of protecting

Prince William Sound from future spills by constructing a new pipeline through Canada. In our early focus groups we found some participants did not consider such a pipeline to be a plausible solution because it would also be vulnerable to spills and involved international complications. This led us to test another provision mechanism by which a new large spill would be prevented by having all tankers in Prince William Sound accompanied throughout their journey by two specially equipped oil spill prevention escort ships. This way of protecting the Sound from another large spill seemed much more plausible to focus group participants.

The scenario must invoke the respondent's budget constraint by clearly conveying that provision of the good will have actual financial consequences for the household. A plausible scenario should also offer a way of paying for the good that will seem reasonable to the respondents. Often the ideal payment vehicle is one that is actually used for this purpose, such as the respondent's municipal water bill for a study of the value of drinking water risk reductions in a particular town or a park entrance fee for a study to value future park improvements. Sometimes preliminary research may indicate that respondents accept a vehicle, such as higher prices and taxes for national water quality benefits, that is less directly related to the good at hand or a vehicle, such as utility bill, that may not in fact be used for this purpose.

Another aspect of the payment vehicle which can cause problems because it is particularly difficult to convey is a multi-year payment. For various reasons, respondents do not seem to feel a strong personal obligation to multi-year financial commitments made in a CV survey beyond two or three years. It is our view that multi-year payment vehicles are only reliable when the service provided is one that is frequently observed and the service is one which respondents believe would be discontinued if regular payments for it stopped (e.g., sewage treatment in a city).

The elicitation mechanism is another scenario feature of considerable importance in creating a plausible choice situation. We have found respondents feel comfortable with a referendum format that uses a take-it-or-leave-it price where the respondent is asked to say whether he or she would vote for or against the provision of a good if it cost his or her household a specified dollar amount.<sup>13</sup> This format works well because it answers the question respondents frequently ask at earlier points in the interview: 'How much is this going to cost me?' It also has the desirable theoretical characteristic of being incentive compatible.<sup>14</sup> The Blue Ribbon Panel decided that the arguments in favour of the referendum format were persuasive and prescribed its use in one of its guidelines (Arrow et al., 1993: 4608).

#### *Process of designing a CV survey*

It is hard to emphasise enough the need to conduct design research during the process of designing a contingent valuation survey. In this section we

describe how we do this in the course of designing in-person surveys such as the one we used in the Alaska study. We usually begin by sitting down with the policy and technical experts to learn about why an valuation effort is necessary and the good to be valued. This is often an iterative process which typically serves a useful purpose of clarifying the economic issues and the nature of the good. Initial information about public perception of the type of good to be valued can often be gathered from media files and by examining reports of contingent valuation studies which valued related goods.

We next hold one or more small group discussions called focus groups. The initial groups use wide ranging discussions to explore how respondents think about the target situation and whether they hold any prior beliefs which might interfere with a contingent valuation scenario on this topics. These discussions are avowedly exploratory and the design team expects to find surprises.<sup>15</sup> Subsequent focus groups are usually useful for getting reactions to initial versions of the scenario and to follow up on issues that emerged in the earlier groups.

An initial draft of the survey instrument is put together during the focus group stage of the research which we try out by conducting a series of in-depth interviews. These allow us to ask the respondent detailed questions about particular aspects of the survey instrument and work on developing the flow of the CV scenario. A contingent valuation scenario is likely to undergo substantial and continuous revisions at this stage as it evolves.

We next conduct small pretests (ten-fifty interviews) in which the revised survey instrument is administered by professional interviewers in the field. At this point, the survey designers look for the reactions of experienced interviewers with respect to any difficulties in administering the survey. Because the pretests involve a larger number of respondents than the in-depth interviews and typically a wider range of respondent types, they can reveal new types of problems. After additional revisions, it may be judged by the survey design team that the survey instrument is ready for a pilot test.

The pilot test is the first opportunity to gather data on a random sample of a reasonable size and to begin making quantitative judgements about the quality of the information gathered by the survey instrument. It is also possible at this stage to start identifying and refining covariates which predict the willingness to pay responses. The initial pilot willingness to pay responses can also be used to help determine the monetary thresholds used in the discrete choice elicitation question. A detailed analysis of the pilot often reveals places where respondents appear to misunderstand the text and can often pin-point administrative problems such as missed skipped patterns that can be corrected in future interviewer training sessions.

If the survey design team is fortunate, the first pilot study may show that the survey instrument is working well and only needs minor changes to work even better. However, the initial pilot study may identify issues which

require further attention. If this is the case, more research is necessary using some or all of the types of research described above. At some point, the survey design team will either feel confident that a final survey instrument can be fielded which will yield valid and reliable data from a large random sample of the population of interest or they will come to the conclusion that they cannot reliably measure the value of the good within the available budget.

#### *CV survey administration*

Survey administration costs almost always comprise a large portion of the total cost of any high quality contingent valuation survey. Because of this they are a tempting target for those with tight budgets to find ways to cut back on by using smaller samples or less expensive survey modes such as mail surveys. It is important to bear in mind that the data from a contingent valuation survey can only be as good as the effort that went into its collection.

The ideal set of circumstances for encouraging a respondent to take CV questions seriously is an in-person interview with a government sponsor administered by a professional interviewer with the survey instrument going to a substantial effort to inform the respondent that the survey results will play a role in the government's decision and that if the good is provided the public will bear its cost. This is certainly not the only set of circumstances that can motivate a respondent to take the scenario presented as a choice with potential economic consequences. There are other sponsors such as universities which can often be a good substitute for government sponsorship. There are some situations where telephone or mail surveys may work reasonably well because respondents see them as the natural means to collect information for making a policy decision.

With respect to the cost of a contingent valuation study, the mode of survey administration is almost always the key determinant because in-person surveys cost substantially more than telephone surveys and telephone surveys cost more than mail surveys. The advantage of using in-person surveys to conduct contingent valuation surveys is their ability to present a considerable amount of information while maintaining respondent interest and attention, to control of the sequence of that information, and to command the maximum degree of respondent seriousness in making the decision about whether or not to pay for the good. The superiority of in-person interviews is so large that a decision to use another mode of administration, such as telephone surveys, will usually need strong justification. Mail surveys are a poor third for a number of reasons of which the most important, at least in the United States, is the inability of many people to comprehend the type of written information which CV surveys must convey (Mitchell and Carson, 1989; III). As we mentioned earlier, one reason for using cheaper survey modes might be that the purpose of

the survey does not require as high quality information as the natural resource damage assessment guidelines recommended by the Blue Ribbon Panel.

No matter which survey mode is chosen, before conducting the final survey decisions must be made about the sample, the response rate, and the extent of the interviewer training; all three of which are sometimes neglected in contingent valuation studies. If one is seeking a sample of the general population in some well defined geographic area, there are well known approaches using complete physical enumeration in in-person surveys or random digit dialling (RDD) in telephone surveys.<sup>16</sup> With in-person surveys, stratification and clustering are used to improve efficiency and reduce costs, respectively. With RDD telephone interviews, budgeting for making a large number of call backs to randomly selected numbers is usually necessary to obtain a high quality sample. High response rates are expensive in every survey mode and beyond a fifty or sixty per cent response rate, costs tend to increase rapidly. Obtaining response rates above seventy per cent is very expensive and it is almost unheard of to obtain a response rate greater than eighty per cent in a regular survey of the general population.<sup>17</sup> No matter what the response rate, weighting is necessary to adjust the sample to the characteristics of the target population. The higher the response rate, the more accurate the weighted sample is likely to be.

Since contingent valuation surveys are typically more complex and hence more difficult to administer than a standard public opinion or marketing survey they require careful attention to interviewer training. We advocate the use of professional interviewers because they are trained to adhere strictly to the text of the instrument whereas graduate students or other types of interviewers who might be recruited on a one time basis are easily tempted to adapt the wording to the respondent or to explain the meaning of text which puzzles the respondent. This type of intervention destroys the purpose of a survey which is to obtain information from a sample of people in response to material which is presented consistently. Problems of meaning and presentation should be solved prior to administering the final survey, not compensated for by those who administer it.

#### *Contingent valuation survey analysis*

Two main principles should guide the analysis of contingent valuation data. The first of these is the use of an analysis strategy which provides protection against a small number of observations or unverifiable assumptions being the major force driving the estimates. The second is that the analysis of contingent valuation survey data should go well beyond obtaining estimates of summary statistics such as the mean or median and explain variation in the willingness to pay responses obtained. A further principle goes without saying and that is complete reporting.

It is always possible to choose a set of statistical assumptions which will make the results of a contingent valuation study look ridiculous. For example, if one assumes that the willingness to pay response follow a log-logistic distribution, for most likely parameter values the estimate of the mean will be infinity.<sup>18</sup> However, we know that willingness to pay is bounded by income. We also know that a small number of the willingness to pay responses will be invalid. This is true of any survey question. As such procedures or statistics which are relatively insensitive to the possibility of a small number of aberrant observations should always be used in a contingent valuation survey.

Good contingent valuation studies have willingness to pay responses which vary systematically with the covariates one might expect to influence that measure. Poor quality contingent valuation studies often produce what appears to be random data where the covariates in the survey explain little or no variance in the data. There are three caveats which apply to this statement. First, there may be gross errors in the values of some of the covariates which may mask the relationship between the covariate of interest and WTP. These gross errors must be addressed in some manner. The income variable is particularly notorious for almost always containing some gross errors as well as a large number of missing values. Second, it is necessary to devote some effort to collecting the covariates. This is frequently neglected in contingent valuation studies but should be a standard objective of the initial survey design and pilot study work. It is a necessary ingredient in assessing the construct validity of a contingent valuation survey and in understanding the implications of how the of WTP responses are distributed. Third, there can be complex patterns of correlation between the covariates of interest and willingness to pay. Some of these patterns have been observed in a number of contingent valuation studies and the specification of valuation functions, in general, is becoming easier over time.

## Conclusion

Contingent valuation as a method of valuing non-market resources has received considerable attention in the four years since we published our book (Mitchell and Carson, 1989). In part this is a product of the method's greater acceptance. The number of studies in the literature has more than doubled and its geographic spread is almost as dramatic. In part this is a product of critical attention. In the United States its use measure passive use values for policy and legal purposes has been contested by affected parties.

In this chapter we discussed the implications of the current debate over contingent valuation in the United States for understanding the nature of the method and how it should be implemented. Two misconceptions about



the method underlie the attempt by critics to demonstrate that the method is fundamentally unreliable. These are that contingent valuation questions are no different from other willingness to pay questions and all CV studies can be treated as if they are equally reliable. We argue that willingness to pay questions in the absence of a contingent market incorrectly measure attitudes rather than behavioural intentions and that it is not meaningful to consider the results of a CV survey without taking into account how they were obtained. Clarification of these misunderstandings shifts the debate to the question: are there conditions under which contingent valuation studies can obtain data that is sufficiently reliable to be useful for policymakers? This is the question which the NOAA Blue Ribbon Panel addressed and answered in the affirmative.

In the last section we discussed some methodological issues involved in obtaining reliable contingent valuation results, including those emphasised by the Blue Ribbon Panel. Our specific recommendations in this section are informed by our experience in conducting the Alaska Survey and by other findings in the growing CV literature. The design of scenario elements, such as the description of the good, the plan for providing it, and the payment vehicle, should be such that respondents clearly understand the nature of the good and the context in which it would be provided and find the choice situation plausible. Determining which design elements will result in reliable, unbiased, results can be a difficult task which may require an extensive programme of preliminary research. Unreliable CV studies typically fail to devote the resources and effort needed for this task.

#### Notes

<sup>1</sup> See Hanemann (1992) for a discussion of the early intellectual history of contingent valuation which puts it in the context of the development of the other methods developed by resource economists to value non-market goods.

<sup>2</sup> For a discussion of issues surrounding natural resource damage assessment in the United States see Ward and Duffield (1992) and Kopp and Smith (1993).

<sup>3</sup> We were particularly sceptical that low cost survey techniques such as self administered questionnaires (e.g., mail surveys) and telephone interviews were suitable for most CV purposes.

<sup>4</sup> In our book (Mitchell and Carson, 1989) we attempted to make the case that the design and execution of a valid and reliable CV survey is a challenging task, both methodologically and theoretically. Unfortunately, it is not possible to provide foolproof instructions to novice CV researchers.

<sup>5</sup> 'Reliable' has two meanings in the CV literature. In our work (Mitchell and Carson, 1989) we have used it as it is used in measurement theory to refer to a measure which is replicable or stable over repeated administrations. In this usage it is normally paired with 'validity' which refers to a measure which provides an accurate representation of what one is trying to measure. The Blue Ribbon Panel used reliability in its ordinary sense as a measure that is 'capable of being relied upon, dependable'. This encompasses both the validity and reliability dimensions.



<sup>6</sup> These grounds include inappropriate question wording and instrument design, use of less reliable survey techniques, small samples, and apparent misinterpretation of the data.

<sup>7</sup> By invoking the trade-off between money and the provision of the good the question is a more stringent way of measuring how important the good is to the respondent than simply using an importance scale. Survey researchers often regard questions that pose trade-offs, such as jobs vs. the environment, as a more credible expression of environmental concern than the use of an 'importance scale', because trade-off questions remind the respondent that environmental progress typically involves a cost of some kind. For an extended discussion of these issues see Mitchell and Carson (1989: chapter 5).

<sup>8</sup> Even admonitions to use willingness to pay measure rather than willingness to accept compensation measures should not be taken as an absolute, as some clever contingent valuation scenario designers (e.g., Garbacz and Thayer, 1983; Howe and Smith, 1991) have been able to construct situations where it is plausible to accept compensation and the strategic incentives are likely to be correct because of competition with other programmes or localities.

<sup>9</sup> For some of the major works in what is a very large survey methods literature on this subject see: Biemer et al. (1991); Schuman and Presser (1981); Tanur (1992); and Turner and Martin (1984).

<sup>10</sup> The types of the goods that might be used in a survey on an environmental good include crime protection, education, helping the homeless.

<sup>11</sup> It should be noted that the typeface was large and there was a generous amount of white space on each page.

<sup>12</sup> Provision plans usually feature a government operated programme because public goods are normally provided by the government. This can make it difficult to convey provision certainty because in the United States today there is a deep scepticism about the ability of government to fulfill its promises and act effectively, a scepticism which can lead respondents who desire the good to vote against a proposed programme because the expected value of what they believe they would get for their payment is too low.

<sup>13</sup> Typically a range of amounts is used, one of which is randomly assigned to each respondent.

<sup>14</sup> It should be noted that a number of additional conditions need to hold for a discrete choice question to be incentive compatible. These conditions are typically violated by asking multiple valuation questions or by the use of a voluntary contribution payment vehicle.

<sup>15</sup> Two examples may help convey how important these surprises can be. In a study of groundwater contamination we learned that people believe ground water travels underground ten to twenty times faster than in actually does. In another study we found that focus group participants believed human caused haze in the Grand Canyon was harming the rock formations. Knowledge of these beliefs made it possible to design the CV instruments so the good was not over-valued because people thought it would create more services than it would actually create.

<sup>16</sup> It is assumed that a probability sample, where every member of the desired population has a known and positive probability of being selected for the sample, is used. Many surveys use less desirable quota or convenience samples. The latter, in

particular, are impossible to use for drawing population inferences although they may be useful for pretesting wording.

<sup>17</sup> The reported response rates are frequently inflated by dropping out of the denominator households who were never contacted. A complete accounting of the final status of each randomly chosen dwelling unit or number is usually needed to accurately determine the true response rate for a survey.

<sup>18</sup> This is true whether the willingness to pay responses are from an open-ended or discrete choice question. For continuous data, it is often not recognised that one is making an implicit distributional assumption when the sample mean is calculated in the standard way of summing the observations and dividing by the sample size. However, that procedure is the maximum likelihood estimator for the mean of a normal distribution. Distributional assumptions are usually much more apparent in dealing with discrete choice data.

### References

- Ajzen, I. and M. Fishbein (1977), Attitude-behaviors relation: a theoretical analysis and review of empirical research, *Psychological Bulletin*, 84: 88-918.
- Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman (1993), Advance notice of proposed rulemaking, extension of comment period and release of contingent valuation methodology report, *Federal Register* 58(January): 4601.
- Biemer, Paul P., R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman, eds. (1991), *Measurement Errors in Surveys*, New York, Wiley.
- Carson, Richard T., and Nicholas E. Flores (1992), A second look at 'does contingent valuation measure preferences? experimental evidence' how evident is the evidence? Unpublished paper. University of California, San Diego, November.
- Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud (1992), A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill. Report to the Attorney General of the state of Alaska, November 1992.
- Carson, Richard T., Jennifer Wright, Anna Alberini, and Nicholas. Flores (1994), *A Bibliography of Contingent Valuation Studies and Papers*, La Jolla, CA: Natural Resource Damage Assessment, March.
- Cummings, Ronald G., David S. Brookshire, and William D. Schulze (1986), *Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Method*, Totowa, NJ: Rowman and Allanheld.
- Davis, Robert K (1963), *The Value of Outdoor Recreation: An Economic Study of the Maine Woods*, Dissertation, Harvard: Harvard University.
- Fischhoff, Baruch and Lita Furby (1988), Measuring values: a conceptual framework for interpreting transactions with special reference to contingent valuation of visibility. *Journal of Risk and Uncertainty* 1: 147-184.
- Garbacz, Christopher and Mark A. Thayer (1983), An experiment in valuing senior companion programme services, *Journal of Human Resources* 18: 147-152.
- Hanemann, W. Michael (1992), Preface, *Pricing the European Environment*, Ståle Navrud, ed. New York: Oxford University Press.
- Hanemann, W. Michael (1994), Valuing the environment through contingent valuation, *Journal of Economic Perspectives* 8: 19-43.

- Hausman, Jerry A., (1993) ed., *Contingent valuation: a critical assessment*, New York: North-Holland.
- Howe, Charles W. and Mark G. Smith (1991), Optimising the reliability of urban water supplies, Paper presented at the annual meeting of the European Association of Environmental Resource Economists. Stockholm School of Economics.
- Kahneman, Daniel (1986), Comments, *Valuing Environmental Goods*, ed. Ronald G. Cummings, David S. Brookshire, and William D. Schulze, Totowa, NJ: Rowman and Allanheld.
- Kahneman, Daniel and Jack Knetsch (1992), Valuing public goods: the purchase of moral satisfaction, *Journal of Environmental Economics and Management* 22: 57-70.
- Kahneman, Daniel and Ilana Ritov (1993), Determinants of stated willingness to pay for public goods. Unpublished paper. Department of Psychology, University of California, Berkeley, March, 1993.
- Kahneman, Daniel, Ilana Ritov, Karen E. Jacowitz, and Paul Grant (1993), Stated willingness to pay for public goods: a psychological perspective, *Psychological Science* 4: 310-315.
- Kemp, Michael A. and Christopher Maxwell (1993), Exploring a budget context for contingent valuation estimates, *Contingent Valuation: A Critical Assessment*, J. A. Hausman, ed. New York: North-Holland.
- Kopp, Raymond and V. Kerry Smith (1993), Valuing natural assets: the economics of natural resource damage assessment, Washington DC: Resources for the Future.
- Mitchell, Robert Cameron, and Richard T. Carson (1988), How far out on the learning curve are we with contingent valuation? *Proceedings of the Conference on the Role of Social/Behavior Science in Water Resource Management*, Anne D. Baumann and Yacov H. Haimes, eds. New York: American Society of Civil Engineers.
- Mitchell, Robert Cameron, and Richard T. Carson (1989), Using surveys to value public goods: the contingent valuation method, Washington, DC: Resources for the Future.
- Navrud, Ståle (1992), *Pricing the European Environment*, New York: Oxford University Press.
- Randall, Alan (1993), Discussions of papers presented at the symposium: an assessment and outlook for contingent valuation. *Contingent Valuation: A Critical Assessment*, ed. J. A. Hausman, New York: North-Holland.
- Schuman, H. and S. Presser (1981), *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, New York: Academic Press.
- Shyamsundar, Priya and Randall Kramer (1993), Does contingent valuation work in non-market economies? Paper presented at the annual USDA W-133 Meeting. Santa Fe, New Mexico, March.
- Smith, V. Kerry (1993), Non-market valuation of environmental resources: an interpretive appraisal, *Land Economics* 69: 1-26.
- Tanur, Judith M. (1992), ed., *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, New York: Russell Sage.
- Turner, Charles F. and Elizabeth Martin (1984), eds., *Surveying Subjective Phenomena* Vols. 1 and 2, New York: Russell Sage.

- Ward, Kevin M. and John W. Duffield (1992), *Natural Resource Damages: Law and Economics*. New York: Wiley Law Publications.
- World Bank Water Demand Research Team (1993), How reliable are contingent valuation studies in developing countries? evidence from a multi-country study. Unpublished paper. Infrastructure and Urban Development Department, World Bank.