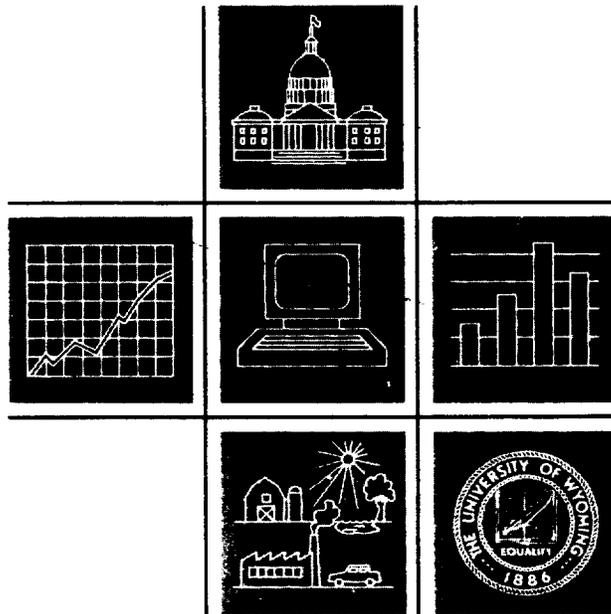


The Institute for Policy Research



...addressing public policy issues for the
'80's.

VOLUME I.A

VALUING ENVIRONMENTAL GOODS: A
STATE OF THE ARTS ASSESSMENT
OF THE CONTINGENT VALUATION METHOD

EXPERIMENTAL METHODS FOR ASSESSING ENVIRONMENTAL BENEFITS

Volume I.A

Valuing Environmental Goods: A State of the Arts
Assessment of the Contingent Valuation Method

Executive Summary

by

Ronald G. Cummings
University of New Mexico

David S. Brookshire
Donald L. Coursey
University of Wyoming

William D. Schulze
University of Colorado

With the Assistance of: Mark S. Walbert

Contributions By:

Kenneth Arrow
Richard Bishop
A. Myrick Freeman
Thomas Heberlein
Daniel Kahneman
Alan Randall
Sherwin Rosen
V. Kerry Smith
Vernon Smith

USEPA Contract #CR-811077-01-0

Project Officer

Dr. Alan Carlin
Office of Policy Analysis
Office of Policy, Planning and Evaluation
U.S. Environmental Protection Agency
Washington, D.C. 20460

OFFICE OF POLICY ANALYSIS
OFFICE OF POLICY, PLANNING AND EVALUATION
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

DISCLAIMER

Although prepared with EPA funding, this report has neither been reviewed nor approved by the U.S. Environmental Protection Agency for publication as an EPA report. The contents do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

EXECUTIVE SUMMARY
TABLE OF CONTENTS

<u>CHAPTER NO.</u>	<u>TITLE</u>	<u>PAGE NO.</u>
A.	INTRODUCTION	1
A.	INTRODUCTION	1
B.	CVM BACKGROUND AND SUMMARY OF RESEARCH	3
B.	CVM BACKGROUND AND SUMMARY OF RESEARCH	3
C.	HYPOTHETICAL BIAS: IMPLICATIONS OF RESEARCH IN PSYCHOLOGY AND OTHER DISCIPLINES	7
C.	HYPOTHETICAL BIAS: IMPLICATIONS OF RESEARCH IN PSYCHOLOGY AND OTHER DISCIPLINES	7
D.	INDUCING PREFERENCE REVELATIONS: METHODS USED IN EXPERIMENTAL ECONOMICS	9
D.	INDUCING PREFERENCE REVELATIONS: METHODS USED IN EXPERIMENTAL ECONOMICS	9
E.	COMPARISON STUDIES: WHAT IS ACCURACY?	11
E.	COMPARISON STUDIES: WHAT IS ACCURACY?	11
	1. Concepts Related to Accuracy	11
	1. Concepts Related to Accuracy	11
	2. Comparison Studies and the Accuracy of CVM Measures	13
	2. Comparison Studies and the Accuracy of CVM Measures	13
F.	ASSESSING THE CVM	16
F.	ASSESSING THE CVM	16
	1. A Suggested Framework	16
	1. A Suggested Framework	16
	2. The Market, Reference Operating Conditions and the CVM	16
	2. The Market, Reference Operating Conditions and the CVM	16
	3. Final Remarks	19
	3. Final Remarks	19
G.	THE ASSESSMENT CONFERENCE: FORMULATING FINAL CONCLUSIONS CONCERNING THE STATE OF THE ARTS OF THE CVM	20
G.	THE ASSESSMENT CONFERENCE: FORMULATING FINAL CONCLUSIONS CONCERNING THE STATE OF THE ARTS OF THE CVM	20
H.	STRUCTURAL BIASES IN THE CVM	21
H.	STRUCTURAL BIASES IN THE CVM	21
	1. Strategic Bias	21
	1. Strategic Bias	21
	2. Starting Point Bias	21
	2. Starting Point Bias	21

	3. Information Bias	22
	4. Vehicle Bias	24
	4. Vehicle Bias	24
	5. Conclusions	24
	5. Conclusions	24
I.	HYPOTHETICAL BIASES IN THE CVM	25
I.	HYPOTHETICAL BIASES IN THE CVM	25
	1. Preference Research Issues	26
	1. Preference Research Issues	26
	2. The Comparability of WTP and WTA	32
Measures		
	2. The Comparability of WTP and WTA	32
Measures		
	3. Attitude vs. Intended Behavior	36
	3. Attitude vs. Intended Behavior	36
	4. Hypothetical Biases in the CVM:	38
Conclusions		
	4. Hypothetical Biases in the CVM:	38
Conclusions		
J.	THE ACCURACY OF CVM MEASURES OF VALUE	42
J.	THE ACCURACY OF CVM MEASURES OF VALUE	42
	1. Overview of the "accuracy" issue	42
	1. Overview of the "accuracy" issue	42
	2. What is Accuracy?	43
	2. What is Accuracy?	43
	3. Reference Accuracy and Public Good	45
Values		
	3. Reference Accuracy and Public Good	45
Values		
	4. The Need for Accuracy or	45
Calibration Research		
	4. The Need for Accuracy or	45
Calibration Research		
	5. The Reference Operating Conditions	45
	5. The Reference Operating Conditions	45
K.	THE STATE OF THE ARTS OF THE CONTINGENT VALUATION METHOD	50
	1. The CVM Without Apology	50
	1. The CVM Without Apology	50
	2. Conclusions Concerning Accuracy	51
	2. Conclusions Concerning Accuracy	51
	3. Final Remarks	52
	3. Final Remarks	52
L.	CRITICAL ISSUES FOR FUTURE CVM RESEARCH	54
L.	CRITICAL ISSUES FOR FUTURE CVM RESEARCH	54
TABLE 3	ALTERNATIVE REFERENCE OPERATING CONDITIONS	47
TABLE 3	ALTERNATIVE REFERENCE OPERATING CONDITIONS	47

A. INTRODUCTION

This document is intended to serve as an Executive Summary of Volume 1.B which has as its focus an assessment of the state of the art of the Contingent Valuation Method (CVM). As is detailed below, the CVM involves the use of survey methods as a means for deriving estimates for individual valuations of non-market environmental "commodities"; typically, such "commodities" take the form of contingent changes in environmental quality, such as improved air or water quality.

Given the hypothetical nature of the CVM -- where a hypothetical commodity is "exchanged" in a hypothetical market for payments (valuations) which are hypothetical -- considerable controversy exists as to the extent to which applications of the CVM can yield, in any meaningful sense, accurate measures of individual values for a posited change in environmental quality. It is important that the reader understand the context for which this controversy is relevant. President Reagan's Executive Order 12291 requires that federal agencies such as the EPA consider the benefits and costs of federal regulations or actions prior to their implementation. For EPA regulations concerning such things as air or water quality, costs may be amenable to estimation but benefits attributable to a large part of these regulations are non-market, 'public goods' in nature. Agencies such as the EPA then have strong incentives and interests in identifying and developing means by which benefits attributable to public goods -- such as environmental improvements -- may be assessed.

Methods other than the CVM exist for valuing public goods, primarily the Travel Cost Method (TCM) and the Hedonic Price Method (HPM). The environmental (and other public good) 'commodities' for which the TCM or HPM might be used for valuation purposes are very limited, however, since appropriate secondary data to apply the TCM or the HPM are often lacking. For the broad range of air quality and environmental safety issues of potential regulatory concern to the EPA, the CVM may be, metaphorically, the only game in town for estimating relevant benefits. Obviously, the fact that the CVM may be no worse than other methods is not a sufficient reason for the use of the CVM values as 'acceptable' economic measures of social benefits in policy assessments. However, the fact that the CVM is the only game in town for providing information of relevance to critical policy issues of the day is a powerful incentive for scholars to meet the intellectual challenge to devise

means by which the CVM (or other methods) can be made effective in responding to society's needs.

Within this milieu, it seems fair to say that all scholars, whatever their predilection towards the CVM, who are directly or indirectly involved with the method appreciate the immediate need for a reflective pause for the CVM. Such a pause is required for thinking through the many propositions that have been posed as indicative of sources for bias in CV measures, as well as related counter-arguments. Most importantly, a reflective pause is required for a re-examination of means by which we can effectively apply the scientific method in our efforts to assess the CVM. Developments with the CVM have reached an important watershed at which a state of the arts assessment of the method is timely.

In an effort to prepare a state of the arts assessment of the CVM, our inquiry in Volume 1.B consists of three major parts. In Part I, a literature review is provided; it is summarized here in sections B through F which follow. The literature review encompasses the historical setting of the CVM, its development and uses over the last decade, and the potential relevance of research in experimental economics for the CVM. The Part I report also considers the implications for the structure of survey instruments used in the CVM, as well as for results of research findings in other disciplines, including psychology. Finally, the Part I report focuses on the question of the "accuracy" of measures derived with the CVM.

Part II of Volume 1.B presents results from a conference on "An Assessment of the State of the Arts of the CVM" which was held on July 2, 1984 in Palo Alto, California. Information provided from the Part I report was critically reviewed in presentations by Professors Alan Randall, A. Myrick Freeman, Richard Bishop, Thomas Heberlein and V. Kerry Smith. Also, a review panel consisting of Professors Kenneth Arrow, Daniel Kahneman, Sherwin Rosen and Vernon Smith critically reviewed both the Part I report and the Conference presentations to the end of presenting the "profession's" view of the state of the arts of the CVM. The results from the Assessment Conference, as they relate to our final conclusions concerning the state of the art of the CVM, are summarized below in sections G-L which is the substance of Part II of Volume 1.B.

B. CVM BACKGROUND AND SUMMARY OF RESEARCH

The goal of the contingent valuation method (CVM) is the establishment of a hypothetical market in order to derive values analogous to market prices. Such values -- referred to as bids -- may then be incorporated into a benefit-cost framework. The basic structure of a contingent valuation market was originally defined by Randall et al. (1974) and further expanded in other studies. Given that the purpose of the hypothetical market is to emulate real markets, a goal in the design was to ensure that the structure and payment mechanisms of the market "were realistic and credible to respondents" (Randall et al., 1974). In particular, substantial detail pertaining to the commodity being valued is provided to the participant prior to the introduction of willingness-to-pay questions. For the non-market environmental good for which values are to be established through bids, alternative environmental changes are described verbally to the respondent with a specific focus on the quality, location and time dimensions of the posited changes. Where possible, devices such as photograph sets and maps and additional technical information are employed to define better the characteristics of the "commodity" (the environmental change). The goal in this process is to ensure that all participants have a uniform and consistent perception of the non-market commodity and the operational nature of the hypothetical market. Substantial operational and structural detail is required in the design of the hypothetical market. The method of hypothetical payment, termed the payment vehicle, is chosen so as to be a familiar mechanism for payment (e.g., access fees at National Parklands or higher utility bills).

The hypothetical market is used in the following way. Beginning with a price initially suggested by the enumerator (i.e., a starting point bid or price), the respondent answers yes or no as to whether the posited price would be paid (accepted) for an increment (decrement) in the public good (level of environmental quality) described prior to the bidding question. An iterative bidding process is continued where, for example, one dollar is successively added to (subtracted from) the previous bid until a bid is reached where the respondent is unwilling to pay more in the willingness-to-pay (WTP) case, or unwilling to accept less in the willingness-to-accept (WTA) case, for the environmental change under discussion. After the iterative

procedure is completed, additional socio-economic information is collected.

Variants on the basic hypothetical market structure have included the establishment and use of budget constraints so that respondents are required to consider the reductions in budget category expenditures implied by their stated willingness-to-pay for an environmental good. Multiple public goods have been introduced into the framework in an attempt to force the respondents to consider substitution possibilities. Respondents also have been provided information on other individuals' bids in order to examine the possibility of strategic bias. Finally, the iterative bidding process has at times been replaced by a payment card where individuals offer a bid by simply checking off a value which falls closest to their maximum (or minimum) willingness to pay.

Early research using the CVM as a means for estimating values for public goods as described above was not guided by a single research agenda. The development period was essentially a heuristic research process. Researchers were drawn to different methodological problems and therefore followed different research directions. The wide range of existing CVM studies can be divided into two groups: a) research efforts that employed the CVM for specific valuation problems and b) research efforts with a methodological as well as a valuation focus. Given our concern with methodological issues, the following brief discussion focuses on results from research in category (b).

To set the stage for our discussion of the results of the review found in Chapter 3 of Volume 1.B, we group these studies into: 1) those which use the axioms of utility theory as a basis for deriving testable hypotheses and 2) those involving experiments concerning design and structural issues related to operational applications of the CVM.

Studies in the axiomatic category focused primarily on questions related to strategic behavior and the notion of preference research. A series of studies has examined the generalized issues of strategic behavior as raised by Wicksell (1896) and Samuelson (1954 and 1958). Two approaches for investigating the prevalence of strategic behavior have been undertaken: the use of experimental markets in a laboratory setting; and an examination of the distribution of bids obtained from applying the CVM. Strategic or free-riding behavior has not been found to be prevalent in experimental (laboratory) markets where a wide variety of auction and pricing schemes have been employed. Research by Bohm (1972), Scherr and Babb (1975), and Smith (1977 and 1979) supports this conclusion.

Evidence from CVM applications is much less conclusive because the tests for strategic bias are not anchored in market structures where actual values can be calculated for comparison purposes. Specifically, the tests are based upon an assumed "true" bid distribution (e.g., the distribution is assumed to be normal, or alternatively, to be related to the distribution of income in the U.S.) and upon the characteristics of individuals (such as environmental preferences) as predictors of strategic bid behavior. If the assumptions and the structure of such tests are accepted, then results from CVM experiments suggest the absence of strategic behavior in the bidding process (Brookshire et al., 1976 and Mitchell and Carson, 1981).

The role within the contingent valuation framework of certain axiomatic structures derived from economic theory has been explored by various researchers. The extent to which CVM values are appropriately constrained by individual budgets has been examined via the use of explicit budget information; generally, there appears to be no statistical difference between bids obtained with and without budget information. The extent to which bids offered by subjects are made within a context wherein substitution possibilities are considered has also been examined. In this regard, groups of subjects are given differing sets of information regarding substitution possibilities and resulting bids are compared. It is generally the case that more explicit information regarding substitution possibilities results in significantly lower bids -- a result that is consonant with the well-known "bounded rationality" hypothesis.

CVM results are also reviewed in Volume 1.B for a variety of design and operational biases potentially stemming from CVM hypothetical markets. Among these, issues concerning starting-point bias and vehicle bias warrant brief mention.

In early CVM studies, willingness-to-pay questions were posed in the form "would you be willing to pay \$X (for the proposed environmental change)", after which an iterative bidding process was used in efforts to determine a maximum willingness-to-pay. Obviously, the final iterated bid, Y, may be biased by one's choice of an initial, "starting" value X, giving rise to what is referred to as a "starting-point" bias. Several studies have examined the dependence of Y on X as a means of testing for starting-point bias with mixed results. Starting-point bias was found to be predominant in the Rowe et al. (1980) study. However, in Randall et al. (1978), Brookshire et al. (1980), Brookshire et al. (1981) and Thayer (1981), starting-point bias was not found. In any case, since about 1982, starting points are seldom used in applications of the CVM. Rather, subjects are given a "payment card" -- an array of values

ranging, for example, from \$.50 to \$50.00 in 50-cent increments -- and are asked to simply check off the amount that they are willing to pay for the environmental good. Thus, the subject chooses the starting point.

An issue related to starting-point bias concerns the question as to whether or not an iterative bidding process is required in order to obtain a subject's maximum willingness-to-pay. If Z is the amount checked off on the payment card and Y is the final bid obtained after an iterative bidding process (which takes the form: "if the good is not obtainable at \$Z would you pay \$1.00 more?"), at issue is the question: is Y significantly different from Z? In these terms, analyses by Schulze et al. (1983), and Sorg and Brookshire et al. (1984) indicate that Y is indeed significantly different from Z -- generally, Y will be some 40% higher than Z. Desvousages et al. (1983) find mixed evidence as to starting-point bias and Y - Z differences. Thus, the relationship between initial bids, Z, and maximum willingness-to-pay remains an open question.

"Vehicle bias" refers to potential biases that may result from the choice of a payment mechanism, or payment vehicle, for the subject's offered bid. Typical payment vehicles used in CVM studies include higher taxes, higher utility bills and access fees. Vehicle bias was found to be predominant in studies by Rowe et al. (1980), Greenley et al., Daubert and Young (1982), was not found in Randall et al. (1978) and mixed results were found in Brookshire et al. (1980 and 1981).

The Rowe et al. (1980) study, which examined a wide variety of biases in the CVM, suggests an overall maximum percentage distortion on bid values attributable to these biases. They concluded "the level of distortion was found to be up to 40% ..." (p. 18) This study did not, however, include estimates of potential distortions of maximum willingness-to-pay from the use of iterated bids (Y-values) vs. non-iterated bids (Z-values).

A final issue relates to theoretical and empirical differences between willingness-to-pay and willingness-to-accept measures. Willig (1976), for the case of price alterations, as well as Randall and Stoll (1980) and Takayama (1982) for quantity alterations, have demonstrated theoretically that the differences should be small. However, studies by Hammack and Brown (1974), Sinclair (1976), Banford et al. (1977), Coursey et al. (1983) and Knetsch and Sinden (1983) indicate a priori expectations are not fulfilled. Willingness-to-accept measures are typically an order of magnitude larger than willingness-to-pay measures. This issue will be taken up later in this summary.

The research efforts summarized above were not designed to demonstrate the efficacy of the hypothetical market in terms of accurate revelations of preferences, but to explore the size of possible distortions (e.g., biases) created by the use of various design options for a hypothetical market or to examine the extent of market-induced strategic behavior. However, the issue of "accuracy" remains central to our inquiry -- see Freeman's lament that "... the problem of accuracy has been almost totally ignored in the economics literature ..." (Freeman, 1979, p. 98).

Oskar Morgenstern (1973) succinctly identified several aspects of the interplay between accuracy and economic data that are relevant for assessments of the CVM. The level of accuracy should be established with a clear understanding of the "particular purpose for which the measurement is made" (p. 4) -- in our case, benefit-cost analysis. Morgenstern further suggests that it is inappropriate to "treat material in an 'accurate' manner at a level exceeding that of the basic errors," which for our purposes are the errors in benefit measurements. To illustrate this point, Morgenstern provides two examples:

"The classical case is, of course, that of the story in which a man, asked about the age of a river, states that it is 3,000,021 years old. Asked how he could give accurate information, the answer was that 21 years ago its age was given as 3,000,000 years." (p. 64)

or

"... in order to determine the precise height of the Emperor of China whom none of his subjects had ever seen, it suffices to ask each of the 300 million Chinese, what he thinks the height is and average their opinions. This will necessarily give a very precise figure." (p. 64)

In section E of this summary the authors develop a suggested framework which may be useful in addressing the accuracy issue as it is relevant for the CVM. This framework, based on the scientific notion of reference accuracy, first requires that we review the substance of research results from the fields of psychology (and related works in the subdisciplines of economics) and experimental economics.

C. HYPOTHETICAL BIAS: IMPLICATIONS OF RESEARCH IN PSYCHOLOGY AND OTHER DISCIPLINES.

There is a wide range of criticisms of the CVM related to the fact that the "market" and payment for the CV commodity are all hypothetical. Our analyses of these issues in Volume 1.B result in the following propositions.

1. "Hypothetic bias in CV measures can result from the fact that payment in the CVM is hypothetical". In this regard, CVM studies by Bohm and by Bishop and Heberlein produce results which are interpreted by the authors as demonstrating the proposition that valuation procedures involving hypothetical payment will yield biased results. Recent work by Carson and Mitchell suggest that Bishop and Heberlein's findings of hypothetical bias depend upon estimation methods which, when altered somewhat, result in contrary findings: i.e. the absence of bias related to hypothetical payment. In large part, the credibility of Bohm's findings of payment-related bias is dependent upon one's criteria for "accuracy", an issue discussed below in Section E. Also related to this proposition are research results by scholars in the field of psychology. Slovic and others find substantial differences in decision strategies with actual and hypothetical payment. As shown below in Section D, however, some part of these results may be weakened by the fact that the experimental setting used by psychologists did not include the incentives for preference revelation used in experimental economics.

2. "Hypothetical bias in CV measures may result from the fact that the CVM commodity is hypothetical." This proposition may take one of several forms. For example, first it is argued that, given hypothetical commodities with which individuals are unfamiliar, the preference research process requires considerable more time than the short period of the CVM interview. Secondly, it is argued that biases may result from the lack of consonance between the commodity "offered" in the CVM and the individual perceptions of that commodity; different individuals will perceive and, therefore, value different commodities. There exists considerable evidence that supports the proposition that for commodities with which individuals are unfamiliar, biased valuation measures may result from differences in individual Perceptions of the CVM commodity. Under the best of circumstances, individuals' "information processing" capabilities may be quite limited (Slovic, Kunreuther and White, 1974); with unfamiliar commodities presented within a hypothetical setting the short time of the CVM interview may simply be insufficient for accurate perceptions of the

commodity and the preference research process. Therefore, it is unlikely that CVM will yield meaningful valuations for the commodity.

3. "Hypothetical bias in CV measures may result from 'framing' effects (Tversky and Kahneman, 1981 and Kahneman and Tversky, 1982) and from the fact that the CVM may elicit responses reflecting attitudes rather than intended behavior." Thus, the framing of WTP questions and the manner in which hypothetical commodities are described, may affect valuations offered in the CVM. Framing biases may be particularly important when individuals are unfamiliar with the commodity. Some framing contexts may elicit attitudes rather than indications of intended behavior, and psychologists find attitudes to be a poor indicator of intended behavior. Further, to the extent that individuals isolate various decision contexts -- think in terms of "mental accounts" -- the context within which the CVM commodity is considered by the individual may be affected by the manner in which the CVM questionnaire is framed.

One finds some mentions in the literature of means by which some aspects of hypothetical bias might be mitigated, if not eliminated. For example, Azjen and Fishbein suggest that attitudinal biases may be mitigated by posing questions within specific contexts, with specific targets, actions and time frames.

D. INDUCING PREFERENCE REVEALATIONS: METHODS USED IN EXPERIMENTAL ECONOMICS

Experience gained from laboratory experiments conducted by experimental economists has a number of important implications for the contingent valuation method of valuing public goods. These implications are described in detail in Chapter IV of Volume 1.B. Presently, contingent valuation surveys are designed to collect field data relevant for social policy analyses. In the terminology of experimental economics, each survey instrument has its own set of rules and therefore furnishes a specific set of individual messages about the public good. The survey method permits control over changes in the institutional rules for allocating a public good but little or no control over individuals' valuation of the good. A researcher may propose a new questionnaire design, and test that design in the field. However, lacking control of information concerning preferences, the results of that survey cannot be unambiguously interpreted. Evaluation of each survey's results is complicated by the classic problem of underidentification. Field experiments must be interpreted in terms of assumptions about both individual preferences and assumptions about behavior implied by the rules of the survey. However, the fundamental objective behind a laboratory experiment in economics is to create a manageable "microeconomic environment in the laboratory where adequate control can be mandated and accurate measurement of relevant variables guaranteed" (Wilde, 1980, p. 138). As pointed out by Smith, control and measurement can only be measured in relative terms, but undoubtedly are much more precise in the laboratory than in the field.

The technique of laboratory experimental methods may be well suited for testing the relative performance of different contingent valuation surveys and for designing and evaluating new survey instruments of interest to policymakers. Any desired configuration of preferences over an abstract collection of public or private goods can be induced for a group of individuals (Smith, 1976, 1982). Each individual is assigned a payoff rule indicating the amount of money he or she will receive for various outcomes of the social decision process. As long as the individual prefers more money to less, a preference ordering is induced over the outcomes of the social decision process.

Within this context, laboratory methods might be used to study the comparative performance of survey instruments. The research objective in comparative studies is to understand how and why different field instruments solicit

messages from individuals by conducting similar surveys in the laboratory. Fortunately, the results of these types of studies usually provide insights for modifying existing CVM institutions and for directing future research. Subsequently, testing of new questionnaire formats with novel allocation rules may be quickly and inexpensively accomplished in the laboratory.

Of primary interest for applications of the CVM are three sets of methods/techniques used in experimental economics to the end of inducing subjects to accurately reveal their preferences: the Vickery Second Price Auction, some form of a tatonnement process, and repetitive trials. The Vickery Second Price Auction involves an auction process for n units of a good among I subjects ($I > n$) wherein each subject is informed that successful bidders will pay the bid of individual $n+1$; i.e., the bid of the highest bidding unsuccessful subject. It can be demonstrated that the Vickery Auction results in values which are Pareto optimal and which can accurately reflect the subject's preferences.

For auctions involving public goods, a tatonnement process is commonly used in laboratory experiments. Such auctions make use of a process based upon a Groves-Ledyard (1977) mechanism for providing a collective good. In a public-good auction, individuals submit desired quantities of the commodity and the cost share or contribution for the commodity that they would voluntarily accept. To each individual is reported the average group quantity and his or her share of total cost given the contributions of others in the group. Each individual then has the right to either veto or agree to the tentative results. Group agreement prevails if and only if each individual agrees to the outcome and the group covers the cost of the proposed amount of the public good. If agreement is reached, then each individual receives the public good and must pay his or her cost share. The veto provision provides the tatonnement process in the sense that no contracts can occur until all individuals in the group are in equilibrium or agreement. This provides at least a partial solution to the problem of free-riding or the incentive to contribute less than true maximum willingness-to-pay. One individual can veto the results of the auction even if every other individual in the group agrees about a given quantity and distribution of cost shares.

Finally, researchers in experimental economics have found that subjects require a number of "dry-runs" or trials of the auction procedure before they become cognizant of the fact that truthful revelation of preferences is in their best interest. In other words, subjects require experience with the auction mechanisms in order to learn (although some

subjects never do learn) that true preference revelation is a dominant strategy.

The relevance of these methods used in experimental economics for the CVM is relatively straightforward. First, the Vickrey Auction and the tatonnement processes provide insights as to means by which CVM subjects might be given incentives for accurately revealing their preferences. The requirement of repetitive trials in laboratory experiments so that subjects learn what is in their interest, may imply a corresponding need to provide some sort of "learning" mechanism for subjects in applications of the CVM. In this regard, the iterative bidding process described above may serve this purpose.

E. COMPARISON STUDIES: WHAT IS ACCURACY?

How accurate are values obtained from CVM studies? Are these values as accurate as values obtained from other traditional approaches such as the travel cost method (TCM) or the hedonic price method (HPM)? Obviously, if both the CVM and, for example, the HPM give the same value for the same commodity under the same circumstances and if this can be shown to be true when repeated for many environmental commodities, and if the HPM is viewed as generating accurate measures of value, then this would provide strong evidence for the relative accuracy of measures derived with the CVM. In efforts to address these issues, the following line of argument is developed below. First, we consider how "accuracy" might be defined and consider the implications of one such definition -- Reference Accuracy -- for assessments of the CVM. Secondly, criteria for accuracy are applied to the CVM in a context where CVM values are compared with corresponding values derived from market-based studies using either the TCM or the HPM.

1. Concepts Related to Accuracy. There are three concepts related to criteria concerning "accuracy". First, the traditional definition of scientific accuracy as seen in statements such as the "measurement is accurate to within ± 50 percent of the measure's value." Such a definition of accuracy is essential, because estimates of accuracy which economists have implicitly employed, such as the standard error of a regression coefficient in a hedonic equation, do not reflect the many possible sources of inaccuracy such as improper choice of functional form, simultaneous-equation bias, or inappropriate assumptions on the distribution of the disturbance term, etc. The only way to incorporate a broader estimate of the total possible range of error is to catalogue the documented range of deviation in measured values for a particular technique. For example, Leamer, in an article aptly entitled "Let's Take the Con Out of Econometrics" (Leamer, 1983) argues that the only way to assess the true accuracy of econometric estimates is to perform sensitivity analysis on such factors as choice of functional form. Summing up demonstrated possible sources of error as a percent of estimated values then allows determination of an economic equivalent of "reference accuracy".

Reference accuracy is defined as the limit that errors will not exceed when the device is used under reference operating conditions. In scientific applications the "device" is a measuring instrument such as a scale used for obtaining weight, whereas in economics the "device" would be

the technique used such as the CVM, TCM or HPM. "Reference operating conditions" in scientific applications refers to limits on the relevant circumstances under which the measurement is taken, such as temperature, atmospheric pressure, etc. In economic applications such as the CVM, limits also exist. For example, to maintain the hypothetical nature of the CVM and avoid strategic bias, the technique possibly should not be employed for current political issues where individuals perceive their answers will influence immediate outcomes (Rowe and Chestnut, 1983).

We will further specify reference operating conditions for the CVM in section F below, but note that on the basis of the discussions above in section D, the technique should use willingness-to-pay as opposed to willingness-to-accept measures of value and should not be applied to commodities with which people have little or no experience in making prior choices or which involve a high degree of uncertainty.

A second aspect of scientific accuracy, significant digits, should be noted since it is often a point of irritation when non-economists, especially natural scientists, examine benefit estimates produced by economists. An example will make the point clear. An economist might report that the average bid in an application of the CVM was \$11.41. The natural scientist will respond that reporting the result in this way is inappropriate since four significant digits are used, which does not reflect the accuracy of the measurement method. (Note that the standard deviation reported with the average bid is not relevant for assessing accuracy since a large standard deviation may result solely from different individuals having different values (tastes) for the same public good and since a highly biased average bid may have a small standard deviation.) Four alternative ways of reporting the average bid used above as an example and the implied accuracy of each are as follows:

Number of Significant Digits	Average Bid	Implied Accuracy
4	\$11.41	±\$. 005
3	\$11.4	±\$. 05
3	\$11.4	±\$. 05
2	\$11	±\$. 50
2	\$11	±\$. 50
1	\$ 1 x 10 ¹	±\$5.00

Note that the implied accuracy is one half of the value of the last reported digit. Economic value estimates are almost always reported as though they have at least three significant digits. We will argue below that in fact, they have a level of accuracy which implies no more than one significant digit, i.e., an accuracy no better than about $\pm 50\%$ of the measured value.

A third view of the accuracy of scientific measurements relates to the "order of magnitude" of the estimate. For example, a scientist may argue that the amount of CO gas dissolved in the earth's oceans (an important quantity in estimating the likelihood that burning fossil fuels will alter the earth's climate through the greenhouse effect) is only known to within one order of magnitude. What this would imply for estimating the accuracy of economic measurements is shown on the vertical scale in Figure 1, which is logarithmic in that each unit of distance on the scale, moving from bottom to top, represents a tenfold increase in magnitude. Thus, a hypothetical willingness-to-pay bid of \$10 obtained using the CVM payment card (or check-off) approach might be raised to \$14 by applying iterative bidding. If a willingness-to-accept question were to be used along with iterative bidding, this last bid would likely be raised at least by a factor of five, to \$70. The arrows position these example bids along the logarithmic scale. Note how the \$10 and \$14 are close together near the \$10 mark on the dollar scale - "of the same order of magnitude" - while the \$70 bid is close to the \$100 level on this scale, an order of magnitude larger than the previous two bids. Thus, one might argue that the iterative and non-iterative willingness-to-pay bids are "close" - of the same order of magnitude - while hypothetical willingness-to-pay and hypothetical willingness-to-accept measures are not "close" and may differ by about one order of magnitude. Physical scientists and health scientists often argue that "order of magnitude" estimates are the best that can be made for complex environmental processes at issue in many benefit-cost studies. As a result, economists may be in a relatively comfortable position if they can avoid errors as large as one order of magnitude such as those implied by the difference between hypothetical willingness-to-pay and willingness-to-accept measures of value.

2. Comparison Studies and The Accuracy of CVM Measures.

We now consider the implications of the above-described notions of scientific accuracy for values estimated by the CVM and by market-based methods. First, consider the CVM. The range of possible error for the CVM derived solely from possible biases may be roughly established as follows. Rowe et al. (1980) report that, in examining the effect of

starting-point, vehicle, information and strategic bias, only strategic bias did not have a significant effect on bids. They conclude that the sum of starting-point, vehicle and information bias can be as large as 40% of the estimated value. One additional source of bias is relevant. Schulze et al. (1983) show that use of a payment card to record bids, results in bids as much as 40% lower than use of iterative bidding. Even though, based on the experimental evidence of the last section, we reject hypothetical willingness-to-accept measures of value outright, the sum of the demonstrated possible biases is about 64%. In other words, an upper bound bid of \$10 could be reduced to \$6.00 by the sum of the effects of starting-point, vehicle and information bias and further reduced to \$3.60 by using a payment card to collect bids. Averaging \$10.00 and \$3.60 gives an example midpoint bid of \$6.80. If we report this bid, \$6.80, as having an accuracy of $\pm 50\%$, the implied range would be \$10.20 to \$3.40, very close to the range implied by known potential biases in the CVM. Thus, we conclude that given the current state of the arts, the CVM is not likely to be more accurate than $\pm 50\%$ of the measured value.

How accurate are the HPM and the TCM? Unfortunately, detailed estimates of the possible errors associated with these techniques, which might be jointly termed indirect market measures of environmental values, are not readily obtainable. Even though indirect market techniques are regarded by some as yielding accurate, market-analogous values, a large number of theoretical and econometric issues surrounds the estimation of willingness-to-pay using either the HPM or the TCM as originally proposed by Rosen (1974) and Clawson (1959), respectively. For example, a possible identification problem which may apply broadly to indirect market methods has been analyzed by Brown and Rosen (1982). Additionally, simultaneous equation problems such as those resulting from the supply of community public goods from tax revenues in property value studies may create difficulty for the HPM. A special problem exists with respect to assumptions made concerning the value of time spent in travel when willingness-to-pay estimates are derived using the TCM (see for example, Cesario, 1976). All of these problems indicate that obtaining willingness-to-pay for environmental commodities using indirect market methods is more difficult than estimating an ordinary demand equation to obtain the value of a private good. However, we can show that even estimation of ordinary demand equations is subject to surprisingly large errors. Since no systematic study has been done of the possible errors in indirect market methods, we will assume that the errors in these methods are at least as large as those which can be shown to exist for ordinary demand estimation.

Coursey and Nyquist (1983) examine one of the possible major sources of bias by applying a number of estimation techniques which allow for alternative assumptions about residential distributions (including least squares, least absolute errors, Huber, Cauchy, exponential power and Student's t) in estimating demand equations for six commodities in three different countries. Thus, 18 separate demand equations were estimated using six different procedures for each one. Strong evidence was found that the assumption of normality on the disturbance terms was quite generally violated and use of robust alternatives to normality was appropriate. Further, estimates of the intercept, income elasticity and own-price elasticities were highly sensitive to the choice of estimation technique. Changes in estimated intercepts from use of different techniques varied from 5% to 747% and exceeded 50% in 8 of 18 demand equations. Changes in estimated income elasticity across techniques varied from 3% to 851% and exceeded 50% in 5 of the 18 demand equations. Finally, changes in estimated price elasticities ranged from 14% to 183% across techniques, with a change greater than 50% in 12 of the 18 demand equations.

A careful study of the sensitivity of indirect market methods to functional form, included variables, simultaneous equation bias, assumptions on the disturbance term, etc. would be highly desirable. However, the large errors possible in estimating the parameters of ordinary demand equation may serve to suggest that indirect market methods also are likely to be accurate only to one significant digit. More precisely, such measures may have an accuracy of no better than $\pm 50\%$ of measured values.

If demonstrated errors in the CVM and the two indirect market methods, HPM and TCM, are likely to limit accuracy to no better than $\pm 50\%$ of measured values, what are the implications for the comparison studies? If, for example, the measured value for a particular commodity using the CVM is \$10.00 and the same commodity, under the same circumstances is valued at \$28.00 using the TCM, are the two measures different? Many of the authors of the comparison studies would argue that these measures are not only different but that because the TCM is based on actual as opposed to hypothetical behavior, it must be the correct value. In contrast, we argue that these two example values are not distinguishably different since the CVM value has a range of at least \$5 - \$15 and the TCM value has a range of at least \$14 - \$42 and these two ranges overlap.

Table 1 presents a summary of the available comparison studies which have used the HPM or TCM as the basis of comparison for the CVM. Some of the studies listed made

some efforts to establish a range of values for the valuation methods employed. However, none of the studies took into account the full potential range of error for any of the techniques. In fact, many of the studies made no attempt to assess the accuracy of even one of the methods employed. Where a study did establish a partial range, Table 1 reports alternative values. Chapter VI in Volume 1.B reports on these studies in detail.

The most striking aspect of the data in Table 1 is that only in case (a) of the Desvousges, Smith, and McGivney study is there a significant difference between values drawn from alternative techniques, using our criterion for accuracy. In other words, an assumed accuracy of $\pm 50\%$ implies that the CVM has only been shown to be different from indirect market methods in one out of 15 possible comparisons presented in Table 1.

The reader may easily draw an incorrect conclusion at this point. This result does not establish the accuracy of CVM measures for any particular commodity. Rather, it simply appears that values derived from the CVM fall within the range of "reference accuracy" (within the rather large error bounds developed above) for those commodities where indirect market measures can be obtained. In other words, these results suggest that CVM values may be "accurate" in cases where individuals have had some opportunity to make actual previous choices over that commodity in a market framework. These studies do not demonstrate that people are capable of providing marketlike values using the CVM for commodities which are not already being traded, at least to a limited degree, in existing markets; in this latter regard, examples include such "commodities" as existence and option values for preserving an environmental asset over which people have no experience in making prior choices. We will examine this argument in greater detail below.

F. ASSESSING THE CVM

1. A Suggested Framework. In considering the question as to appropriate criteria against which to assess the accuracy of measures derived by the CVM, it is useful to recall the rationale for our interest in the Method. As discussed above, benefit-cost analysis is used in determining the optimal levels for a public investment. At a conceptual level, applications of benefit-cost analysis may be viewed as efforts to deduce market outcomes that would obtain if such investments were made under market conditions. Given benefits and costs determined by market institutions, public goods would be provided at levels at which marginal benefits equal marginal costs.

Of course, for most pure public goods, particularly environmental goods, market institutions do not exist. The CVM is then used as a substitute for the "missing" market; it is used to simulate the market in the sense of eliciting revelations of preferences (a willingness-to-pay) analogous to those which would have resulted under market conditions. Like the market institution, the CVM must then be viewed as an "institution". Thus, the general criterion against which to assess the CVM becomes clear: the extent to which the CVM institution, and preference revelations drawn therein, is comparable to the market institution and preference revelations drawn therein.

To make this general criterion operative for our purposes, we need to bring together the issues reviewed above. Thus, in what follows we develop the following line of argument. First, drawing from what we know of market institutions and from lessons learned from our literature review, we define Reference Operating Conditions (ROC) that are relevant for the notion of "reference accuracy" which may apply to values derived from the CV institution vis-a-vis market institutions. The ROC's are shown to imply limits on the range of environmental goods to which application of the CVM might yield "accurate" (within the reference accuracy context) measures of value. In these terms, the relevance of the ROC's will be demonstrated inferentially by an appeal to the results of Comparison Studies (discussed above in section E) and the psychological/economic issues raised in earlier discussions of "hypothetical bias" (discussed above in section C).

2. The Market, Reference Operating Conditions and The CVM. In our society "the market" consists of many amorphous "markets" which differ in such things as degrees of organization and the necessity for negotiation. Thus, as observed by Knight,

"In economics (a market) means the whole area, often indefinitely defined, within which buyers and sellers of a commodity come together and fix a common price The wheat market is practically the world the market for ... brick from a small factory may not extend beyond a few miles."
(Knight, 1951, p. 68).

To consider some further examples, the market for groceries is relatively well organized and exchange involves little if any negotiation. Towards the other end of the spectrum, the market for used furniture is less well organized and exchange can, in some settings (e.g., the flea market), involve considerable negotiation.

Also of importance for our consideration is the fact that economic deductions drawn from "the market" are complicated by the fact that commodities traded in a market are often heterogeneous. Thus, Knight asks: "... is wheat in Paris the same commodity as wheat in Chicago? ... is a physically equivalent ... can of peas with a label which is a guarantee of quality effectively the same commodity as if it had an unknown name?" (p. 69) In terms of the efficacy of the market vis-a-vis fixing "a common price," these complexities are substantively increased when dissimilar commodities are jointly offered. An example might be a house; to paraphrase Knight, are two physically equivalent (floor space, rooms, paint, appliances, etc) houses, one located in neighborhood A and one in neighborhood B, the same commodities? Most often, the answer is no, inasmuch as other "commodities" are offered in joint supply with the house: crime rates, quality of schools, proximity to beaches, theaters, etc. - and, possibly, environmental (air) quality. Each of these commodities, valued and desirable in their own right, are obtained only in the housing "package." Since one cannot, in choosing a house, pick the crime rate from one neighborhood, the school system of another and air quality from still another, the implicit market valuation of these commodities, "attributes" of the house in a given neighborhood, will be imperfect measures of "true" values associated with these attributes.

Whatever the characteristic of any given market, one of the most important characteristics of the set of interrelations involving the "... process of competing bids and offers" which we call "the market" is its capacity to "... generate high quality information at low cost." (Heyne, 1983, p. 125) Thus, "... the most important single cause of exceptions to (market laws) ... is found in the condition: people do not know the facts." (Knight, 1951, p. 69) The better organized the market, the better people will "know the facts." In these regards, prices provide valuable

information and "... the more such prices there are, the more clearly and precisely they are stated and the more widely they are known, the greater will be the range of opportunities available to people." (Heyne, 1983, P. 125).

Thus, key "reference operating conditions" (ROC's) relevant for the market institution are: (I) the process of competing bids and offers which generates information in the form of experience and familiarity with commodities as well as with the valuation process, and (ii) incentives for an individual to acquire and "process" information imposed by his or her limited income juxtaposed with a more or less strong desire to maximize satisfaction.

The importance of the ROC's described above is exemplified in experimental economics wherein efforts are necessarily made to simulate these conditions in the laboratory setting. Thus, in experimental economics, subjects offer bids within a well-defined information context which allows them to calculate their net (monetary) gains; moreover, repetitive trials are used to provide subjects with the opportunity to learn maximizing strategies. Results from experimental economics in general point to the importance of market-like incentive structures and the trial-feedback-learning process in any effort to form incentive-compatible institutions and, more importantly, to elicit true, market-like preference revelations. From the above and from our earlier discussions, we tentatively suggest the following ROC's as being relevant for state of the art applications of the CVM.

(1) subjects must understand, be familiar with, the commodity to be valued.

(2) subjects must have had (or be allowed to obtain) prior valuation and choice experience with respect to consumption levels of the commodity.

(3) there must be little uncertainty.

(4) WTP, not WTA, measures must be elicited.

ROC's 1 and 2 derive directly from the market institution (which provides high quality information at low cost). Moreover, in terms of ROC 1, results from psychological research (section C above) point to distortions in decision processes (framing biases, etc.) that arise when individuals are unfamiliar with decision contexts. Regarding ROC 2, results from experimental economics emphasize the importance of iterative trials which serve to provide subjects with valuation and choice experience -subjects must "learn" maximizing strategies. ROC 3 derives directly from research in psychology and experimental economics: under conditions of uncertainty, valuation decisions may be subject to distortions resulting from the use of a wide range of

heuristic devices. Finally, as discussed above in section D, WTA measures are generally found to be highly distorted vis-a-vis "true" valuations possibly as a result, psychologists might argue, of cognitive dissonance.

A major state-of-the-arts problem is that we know little about the errors associated with violations of the Reference Operating Conditions. Received research results suggest that if WTA measures are used rather than WTP measures, the WTA measure may be five or more times larger than WTP. In terms of ROC's 1-3, however, we lack the data that would allow us to quantify Reference Accuracy. As noted above, results from psychological and experimental economics research tell us only in qualitative terms that distortions - errors - will result when these ROC's are unsatisfied.

In Table 2, data are given concerning the extent to which ROC's were satisfied in selected applications of the CVM; these applications are described in considerable detail in Volume 1.B's Chapters III and VI. Thus, in Brookshire et al.'s study of air quality in Los Angeles, subjects may well have been familiar with the commodity "smog". With average turn-over of housing in the L.A. area of 3-5 years (in the late 1970's) subjects generally can be assumed to be knowledgeable of the air quality attribute as it relates to housing and housing costs (advertisements for housing in the L.A. newspaper will many times include a description of air quality), in which case subjects may be thought to have had general experience in value choices with respect to "consumption levels" of the commodity (improved air quality). Uncertainty played a negligible role in the CVM application and WTP measures were elicited. Analogous arguments apply to the study of municipal infrastructure by Cummings et al.

What of the CVM studies presented in Table 2 which do not satisfy one or more of the ROC's, particularly ROC's 1-3 about which we know little in terms of Reference Accuracy (e.g., the study designed to derive existence and option value for visibility in the Grand Canyon by Schulze et al. and Burness et al.'s toxic waste study)? In such cases we can say no more than that there exists no positive evidence that would support the accuracy of such measures vis-a-vis market or market-related values. It must be said that negative evidence in this regard does exist. Order of magnitude differences between initial valuations and valuations derived after prior experience (from iterative trials) with choice mechanisms, are suggested by research in experimental economics. Research in psychology has firmly established the distortions in choices which attend decision environments characterized by uncertainty and unfamiliar learning/decision contexts. In short, we can neither

confirm nor deny the accuracy of CVM values derived in applications which do not satisfy the ROC's. Given the present state of the art, however, available evidence suggests that such measures may be seriously distorted.

3. Final Remarks. In closing, the authors recognize that, while an assessment framework based on Reference Accuracy and the resulting Reference Operating Conditions may in form parallel objective frameworks for assessing accuracy in other sciences, it may fall well short of "objectivity" vis-a-vis assessments of the CVM. This follows from the obvious fact that while the ROC's per se may be objectively deduced from market institutions, their application to assessments of a CVM study may generally be subjective. For example, one may ask: what degree of "familiarity" with a commodity is required to satisfy ROC 1; how much value and choice experience (or how many repetitive trials) are required to satisfy ROC 2; and how much uncertainty is "little uncertainty" (ROC 3)? In response to these questions our knowledge of markets and lessons drawn from experimental economics and psychological research tell us little more than that in moving from pure public goods to common market goods, we can expect something of a continuum in meeting ROC's, as exemplified in Figure 2. Thus, in moving from existence values for the Grand Canyon to a hamburger, we expect individuals to be increasingly familiar with the "commodity" and to have had more numerous market-related experiences; as we move along this continuum, uncertainty as to outcomes of transactions and the potential for problems related to cognition is reduced.

In efforts to deal with these issues, the state of the arts is one wherein we can simply say that evidence exists which supports the proposition that indirect market experience with a commodity may serve to satisfy the ROC's: when the environmental good is a distinct attribute of a market-related good (water quality in a time/travel cost recreation trip or air quality as an attribute of housing locations/costs), experience and familiarity with the market good seemingly spills over to the individuals ability to value the attribute. Thus, while not totally answering the "what degree" and "how much" questions regarding the satisfaction of the ROC's; comparison studies may suggest classes of environmental/public goods which may be taken a priori as those which would satisfy the ROC's for the Contingent Valuation Method.

G. THE ASSESSMENT CONFERENCE: FORMULATING FINAL CONCLUSIONS CONCERNING THE STATE OF THE ARTS OF THE CVM

The preceding sections reviewed in Part I of Volume I.B which includes results from the authors' review of the literature and their efforts to suggest tentative, pre-Conference conclusions as to the state of the arts of the CVM. This work was critically reviewed in papers presented at the Conference by Professors Bishop, Freeman, (V. Kerry) Smith and Randall as well as by the Review Panel consisting of Professors Arrow, Kahneman, Rosen and (Vernon) Smith. Written comments concerning our Part I (of Volume I.B) work were also provided by R.C. Mitchell and R.T. Carson (Resources for the Future). Given the pointed relevance of Mitchell and Carsons' discussions of Part I, the comments are included in the Appendix to Part I.B's Chapter XIII.

In the sections that follow, attention is focused on our "final" conclusions concerning the state of the arts of the CVM. Such conclusions draw from discussions and interchanges which took place at the Assessment Conference. Our final conclusions thus are a result from the intellectual assessment process wherein constructive critical interchange between scholars is used to mold conclusions which hopefully reflect some degree of consensus. State of the arts conclusions regarding the CVM are developed below in the following manner. In Section H, attention is focused on the weight of structural bias in the CVM: biases which have been argued to result from such things as starting points, choice of payment vehicles, strategic behavior and information. Section I considers an issue that has long been considered as being of central importance for assessments of the CVM: the potential for hypothetical bias in CVM measures of value. In Section J we re-consider the question of primary importance for our state of the arts assessment of the CVM: how does one evaluate the accuracy of CVM measures? These discussions are brought together in Section K wherein we consider the bottom line: What is the state of the arts for the CVM? Our review of Volume I.B's Part II concludes with Section L wherein we define critical issues for future research with the CVM. In each of the following sections, references to Chapters - Chapters I through XII - refer to Chapters in Volume I.B.

H. STRUCTURAL BIASES IN THE CVM

Four structural types of bias in value measures derived with the CVM were given particular attention in Part I of Volume I.B and were of particular concern at the Assessment Conference. These potential biases, discussed in turn below, are: strategic bias, starting point bias, information bias and vehicle bias.

1. Strategic Bias. In general, the views of Conference participants concerning strategic bias in CVM measures parallel those developed in Chapter V. Freeman notes the absence of strong empirical evidence for free-riding behavior, which in his view suggests that individuals will not behave strategically in purely hypothetical or contingent market settings - a point of view seconded by Rosen. Professor Arrow finds neither theoretical arguments nor empirical evidence compelling in terms of strategic behavior by CVM subjects.

Both Freeman and Rosen emphasize, however, the potential dependence of the "no strategic bias" conclusion on the fact that, within hypothetical settings, subjects in the CVM study are not offered obvious opportunities to manipulate outcomes; i.e., as noted in Chapter V, the potential for strategic bias is less, the more hypothetical the valuation process in the CVM. Such dependence, if it exists, raises two related problems, however. First, and most obviously, a trade-off is suggested between strategic bias and hypothetical biases - this issue will be discussed in detail below in Section I. Secondly, a number of researchers are currently advocating alternative structures for the CVM wherein emphasis is placed on the subject's perception that his/her response will influence policy. Thus, Randall's theoretical model (Chapter VIII) is based on the assumption that subjects believe that the results of the valuation exercise will influence policy; within this framework, the "penalty" for a non-preference-researched response is argued to be that the subjects' opportunity to influence policy is wasted or misused. Such focus on influencing policy, as noted by Randall, is suggestive of referendum formats; indeed, Kahneman views the CVM as it stands as effectively simulating a referendum. Carson and Mitchell (Appendix) look to referendum formats - political markets - as an alternative framework for the CVM and as a means for identifying "reference operating conditions" relevant for assessing the accuracy of CVM measures (Appendix, part 4).

Ceteris paribus, the use of referendum-type formats as a means to investigate hypothetical bias may be questioned on the grounds that the more real is one's perception of the

relevance of his/her responses in terms of influencing policy, the greater is the potential for strategic bias (see, in Chapter XII, Rosen's "personal computer" analogy). It is not clear that such is the case, however. As implied by Carson and Mitchell, couching the CVM within the context of a referendum may in fact amount to the adaptation of the CVM to an institution which differs markedly from the market institution which common applications of the CVM attempt to simulate. The possibility of tying the CVM to alternative institutions (vis-a-vis the market institution) is an interesting and potentially important point and is considered in some detail below in Section F.

2. Starting Point Bias. In Chapter III we noted that when the CVM valuation process is initiated by the interviewers' question: "Would you be willing to pay \$X," post-bidding valuations tended to cluster around \$X.

The dependence of CVM values on the initial or "starting point" value of \$X was described as a "starting point bias." We noted empirical evidence supporting the existence of such biases - Carson and Mitchell (Appendix, section 2.a) suggest still stronger evidence for such biases and argue that studies suggesting the absence of such biases may be flawed by the low power of tests used to examine hypotheses concerning starting point bias. At least two methods have been suggested for eliminating/mitigating starting point bias: the use of a payment card (c.f. Chapter III), and Freeman's naval gunfire analogy of "bracket and halving" (Chapter X).

Professor Kahneman (Chapter XII) proposes quite a different context for treating and interpreting starting point bias. Kahneman suggests that the finding of starting point bias is indicative of a CVM "commodity" for which subjects are unable to answer valuation questions. For some types of commodities, lack of experience or familiarity with the commodity results in subjects' having great difficulty in putting dollar values on the commodity - subjects are not "hiding" anything from the interviewer nor are they attempting to be clever, they simply do not know how to answer the valuation question in a meaningful way. Thus, rather than adopting means to eliminate starting point biases, Kahneman seemingly views means to identify the existence of such biases as an important part of the study design: the presence of such biases indicates that subjects are too ignorant of the commodity to be able to value it meaningfully, in which case the CVM should not be applied to the commodity in question. Kahneman offers further "sad news" (XII. C): use of a payment card does not eliminate the problem inasmuch as value ranges on the bidding card

provide the potential for "entering biases" (indications of "reasonable" responses).

When starting points are used in CVM studies, we concur with Carson and Mitchell that the evidence suggesting starting point biases is indeed compelling. While, as is discussed in Section I, Kahnemans' concern that a subjects' lack of experience/familiarity with a particular environmental good may result in his/her having difficulty in placing monetary values on the good - indeed, "familiarity, and/or experience is an ROC in Chapter VI - received empirical evidence does not seem to support the notion that such difficulties are made manifest by starting point biases. Following Mitchell and Carson's suggestions (Appendix), higher powered tests for such biases may well result in starting point biases showing up in CVM studies involving commodities with which subjects are reasonably familiar - see the seven studies wherein derived CVM values are shown to compare favorably with values derived from indirect market methods (Table 6.12 in Volume I.B). Thus, we would argue that starting point bias may well reflect other phenomena, e.g., the subjects' interpretation of starting points as indicative of actual costs for a proposed environmental improvement. Moreover, it would appear that payment cards can be structured so as to eliminate the potential for the "entering biases" of concern to Professor Kahneman. Thus, while an issue of concern, the authors conclude that starting point problems should be amenable to control through care in the design of the CVM payment card.

3. Information Bias. In Chapters III and V, the authors pointed to the confusion that one finds in the literature as to the substance of what is referred to as "information bias;" at the heart of this confusion is the failure on the part of many writers to distinguish between effects on CVM valuations arising from the subject's exposure to more information ("more" in quantitative and/or qualitative terms) regarding the commodity or valuation process as opposed to the subjects exposure to different information - "different" in the sense that two sets of information imply two different market (valuation) structures or two different commodities.

Randall (Chapter VIII) suggests that such confusion is eliminated as follows. Rational subjects base their contingent market decision on (i) the value of the commodity offered; (ii) the rule by which the agency decides to provide or not to provide the commodity; and (iii) the rule that determines the payment to be exacted from the subject. Since, according to Randall, only (i) is relevant for valuing nonrival goods, the pertinent question is: do (ii) and (iii) encourage accurate reporting of (i)? In this

vein, Randall argues that different information which affects (ii) or (iii) should affect reported measures of willingness to pay. Such changes in information then result in effects on WTP measures that are expected a priori. Such effects, therefore, are not biases. In this manner, Randall rejects the notion of "information bias."

Related to Randall's point (iii) - as well as to (ii) - is the design question as to whether or not a subject in the CVM should be given information concerning bids by other subjects. Arrow argues that such information should not be given due to the potential effect of this information in eliciting strategic behavior. Moreover, Arrow views such "second hand" information as possibly leading to biases resulting from subjects' dependence on more informed judgments of others, as implied by their bids. Freeman argues that such information could lead, in effect, to a form of starting point bias. Along a slightly different line, Kahneman sees information concerning (iii) as an integral part of the valuation process - any one individuals' "true" willingness to pay is inextricably related to what all other individuals are paying for the commodity in question, i.e., Kahneman implicitly rejects the economists' commonly-used assumption of independent utility functions.

However, Randall's arguments concerning (I) - (iii) address only one part of the sources of information of concern in Chapter V: changes in information affecting value structures and/or commodities; his arguments do not seem to speak directly to the relationship between reported valuations and the quantity/quality of descriptive information concerning the commodity. In these regards, it would seem that in cases where systematic differences in valuations are associated with changes in the quantity or quality of information describing the CVM commodity, the implied "bias" may well be attributable to difficulties in "information processing" described in Chapter V. Arrow points to the difficulties in balancing the potential benefits of providing subjects with descriptive information with the subject's difficulties in processing that information. Freeman sees such biases as positive vis-a-vis assessments of the CVM inasmuch as they may be interpreted as indicative of subjects' approaching the valuation process in a meaningful way; i.e., subjects use information provided to form perceptions of the CVM commodity and base their valuation responses on that information.

Thus, in terms of information which has the effect of altering the nature of the CVM commodity, rules for providing the commodity and/or rules which determine actual payment, we would concur with Randall's judgment that one

would expect such changes to alter bids, in which case a bias per se is not implied. On the related subject concerning a subject's exposure to bids offered by other subjects, we find the argument that such information may result in undesirable biases compelling; in this regard, we note that, while a substantive issue which perhaps warrants future inquiry, Kahneman's rejection of the assumption of independent utility functions weakens results from virtually all benefit assessment methods. Finally, in terms of biases which may result from different levels of purely descriptive information given to CVM subjects, two concluding observations appear salient. First, an integral part of pre-tests of questionnaires must be the effort to balance the subject's need for information with his/her general capacity to absorb - process - the information. Secondly, as suggested by Freeman, one must avoid interpretative generalizations of CVM results to environmental changes other than those specifically described in the CVM instrument.

4. Vehicle Bias. Conference participants, particularly Professors Arrow, Kahneman and Randall, took sharp issue with Chapter V's discussion of vehicle bias. The essence of our discussions of vehicle bias in Chapter V is reflected in Freeman's (Chapter X) statement of the vehicle bias problem: our inability to determine which payment vehicle, if any, provides "true" (unbiased) values and which payment vehicles lead to biased values. Arrow, Kahneman and Randall argue that the search for an unbiased payment vehicle is misguided - "biases" are not implied by systematic variations in offered values and payment vehicles.

The essence of Arrow and Kahneman's argument (see Kahneman's ROC Number Seven in Chapter XII.C) is that the social arrangements by which payments are to be made - the payment vehicle - is an integral part of the CVM commodity per se, i.e., one cannot separate the value of the commodity from the procedures by which the commodity is provided and payment is made. Of course, this is Randall's argument (iii) concerning information bias which was discussed above. In this regard, Kahneman rejects the notion that values based on one set of "social arrangements" may be transferred to a different set; Arrow sees differing preferences - and therefore values - related to purchases via use permits, general taxation and/or general price effects, as rational. Thus, Arrow suggests that WTP depends on the structure of "P".

These arguments are surely compelling and have important implications for the design of and interpretation of results from the CVM. First, following Kahneman (Chapter XII.C), reflecting the fact that our commodity is not a market

commodity, but a commodity which can only result from social action (government intervention), the CVM's mode of payment is selected on the basis of realism - what payment vehicle would most likely be employed, in fact, if the commodity were to be provided? Secondly, paralleling Freeman's interpretative limitations related to information bias, we explicitly acknowledge, without apology, the potential dependence of obtained valuations on the adopted payment vehicle.

5. Conclusions. In terms of the potential structural biases in CVM values which this Section addressed, the current state of the arts in the CVM may be described as follows. First, all else equal, strategic bias does not appear to be a major problem in applications of the method. Two caveats are relevant for this conclusion, however. Interactive information concerning other subjects' values, as might attend efforts to bring standard CVM practices together with experimental techniques, may introduce incentives for strategic behavior. Further, efforts to reduce the potential for hypothetical bias (discussed below) in the CVM, a la Randall's proposed dependence on a subject's belief that his/her response will actually affect public policy, may invite strategic behavior in applications of the CVM which rely on market institutions - the implications of structuring the CVM in alternative institutions are discussed below in Section L.

Secondly, the authors submit that the use of carefully structured payment cards can effectively mitigate starting point bias in applications of the CVM involving commodities with which subjects have had some degree of market-related experience - where subjects are reasonably "familiar" with the commodity. For other commodities, Kahneman's concern with starting point bias - with or without a payment card - may be well-founded, but it is unclear to the authors how one would distinguish between anchoring-sorts of biases in these cases and biases attributable to the myriad hypothetical-related issues concerning decision-making under uncertainty, attitude/behavior and others which arise when individuals begin at the bottom of a learning curve relevant to an environmental commodity.

Thirdly, the "information bias" rubric seems to serve no useful purpose for assessments of the CVM; indeed, it may be counterproductive. In terms of the quantity/quality of descriptive information concerning the CVM commodity, it seems reasonable to expect that pre-tests of questionnaires can be used to balance information needs with information processing capacities for "appropriate" commodities. Once again, the familiarity issue arises as does the relevance of the authors' suggested ROC's. In the case of unfamiliar

goods, in the authors' minds, it appears sanguine to expect that processing capacities can be balanced with the bulk of information that might be required to elicit reasonably informed valuations from subjects.

Finally, in terms of information concerning rules pertaining to the provision of the commodity and/or to payment, we see little to distinguish these information "biases" from those considered under the rubric of "vehicle bias." In these regards, we consider the state of the arts as one wherein the notion of vehicle bias, broadly defined, is without substance. One acknowledges that such rules are an integral part of the valuation process. Values derived via the CVM are then interpreted as simply applying to the specific commodity described in the questionnaire, provided under the "social arrangement" (rules for provision and payment vehicle) described in the questionnaire. In this context, one views with equanimity the rational fact that different payment/provision institutions - social arrangements - may result in different valuations.

I. HYPOTHETICAL BIASES IN THE CVM

The issues associated with hypothetical bias, and the implications of such biases, served as a source of interesting exchanges at the Assessment Conference. Reflecting some degree of consensus among conference participants, the major issues related to hypothetical bias, as they are relevant to our state of the arts assessment of the CVM, are: the preference research issue(s); the comparability of WTA and WTP measures; and the attitude v. intended behavior issue. Those issues are considered in the discussions that follow.

1. Preference Research Issues. Under the rubric of "preference research" developed in Chapter V, three distinct lines of argument can be discerned from the Conference papers and discussions: the role of incentives for accurate valuations; the importance of a subject's familiarity/experience with the CVM commodity; and the (related) learning issue.

(a) Incentives and accurate valuations. In Chapter V.B, arguments by Freeman (1979) and by Feenburg and Mills (1981) concerning the lack of incentives for "accurate" valuation responses in the CVM were distilled into a hypothesis of the form: valuations with actual payment equal valuations without actual payment (i.e. with hypothetical payment). Underlying this hypothesis was Freeman's notion that, since individuals suffer no utility loss from inaccurate responses to CVM valuation questions, they lack incentives to engage in the mental effort (and consumption of time) required to research preferences and formulate meaningful evaluations. Our review and interpretation of the literature related to the above hypothesis -primarily the works by Bohm (1972), Bishop and Heberlein (1979), Coursey et al. (1983) and Slovic (1969) - resulted in our conclusion that results from research to date belie the above stated hypothesis, i.e., substantive differences in values result when real and hypothetical payments are involved. Obviously, the implications of this conclusion would not bode well for the CVM. If hypothetical payment does not provide incentives for accurate responses in the CVM, and absent means for quantifying such biases, the viability of the method may be seriously questioned.

Mitchell and Carson (Appendix) take sharp issue with our conclusion. Based on their reworking of data used by Bohm and by Bishop and Heberlein, they find that results from these works concerning actual/hypothetical payment are much weaker than those reported in the authors' original papers.

In turn, however, we should note Bishop and Heberlein's critiques of Mitchell and Carson's reworking of their data, given in Chapter IX. Moreover, Mitchell and Carson challenge the relevance of results from the Coursey et al. study inasmuch as the study's focus is on WTP-WTA differences, and results related to actual/hypothetical payment differences are simply inferential. Finally, referring to the literature in cognitive psychology, their discussions with Slovic suggest that, first the general literature on this topic shows equivocal findings; and second, that results from Slovic's 1969 study do not strongly support the sweeping conclusion offered by us in Chapter V.

Of course, Mitchell and Carson do not argue that hypothetical payment does not result in bias; rather they argue that the question remains open. Arrow seemingly agrees that the question is open. He argues (Chapter XII-B) that in the pseudo-reality of the CVM, well-structured questionnaires which create real-like markets may well be capable of generating real-like results. Randall (Chapter VIII) offers a stronger argument: notwithstanding hypothetical payment, incentives for a subject to research preferences and formulate accurate valuation responses are provided by the subjects' concern with foregoing an opportunity to influence policy - we have noted above the potential conflict between this position of Randall's and the strategic bias issue noted by Arrow, Freeman and Rosen. Perhaps still stronger in these regards are results from laboratory experiments conducted at the University of Arizona reported by Vernon Smith (Chapter XII-E). Based on these experiments, Smith concludes that interrogated WTP/WTA values (corresponding to hypothetical payment/compensation) were found to be better predictors of post-trading equilibrium values for prices than a priori predictions from expected utility theory. Moreover, while pre-trade predictions of trading volumes were typically inaccurate, Smith notes that predicted (hypothetical) valuations were generally close (around 95%) to actual market-clearing prices.

There remain, however, the results of Bishop and Heberlein's recently completed study of Sandhill deer hunting permits (Chapter IX). As in their early goose-hunting permit study, Bishop and Heberlein find significant differences between bids involving cash and hypothetical payments in all of their WTA experiments (Table 9.2 in Volume I.B) and in three of the four auction formats used in their WTP experiments (Table 9.3 in Volume I.B). Based on these findings, Bishop and Heberlein conclude that the evidence for bias related to hypothetical payment is rather convincing. Moreover, they argue, no matter how closely the

Reference Operating Conditions are met, hypothetical bias (attributable to hypothetical payment) will remain.

Bishop and Heberlein's conclusions, as well as the results from their impressive Sandhill study, are not readily dismissed. No matter how weakened by Mitchell and Carson's analysis, there exist research results from several studies (reviewed in Chapter V) supportive of those offered by Bishop and Heberlein. But there exists a great deal of evidence which challenges the weight of Bishop and Heberlein's conclusions. In this regard, we note the above-cited observations by Mitchell and Carson and by Arrow, as well as, particularly, the experimental results reported by Vernon Smith. Moreover, results from Chapter VI's analyses of seventeen comparison studies demonstrates remarkable (in our view) consonance between values derived with the CVM and values derived from indirect market methods - a degree of consonance which is, at worst, inconsistent with the full weight of Bishop and Heberlein's conclusions, particularly as their conclusions refer to commodities which to some extent satisfy our ROC's. Similarly, these demonstrations argue against the strong conclusion suggested by us in Chapter V.

In offering, then, a state of the arts conclusion concerning the incentives issue generally, and biases attributable to hypothetical payment particularly, the authors feel compelled to soften their conclusions in Chapter V and to concur in principle with Mitchell and Carson: at worst, evidence from research to date provides equivocal results concerning the hypothetical payment issue; at best, for public goods which satisfy the ROC's, evidence from comparative and experimental studies suggests that minimal biases in CVM measures may result from hypothetical payment.

(b). Familiarity/experience as a prerequisite for CVM commodities. A second preference research issue developed in Chapter V concerns the extent to which subjects in the CVM interview can place meaningful values on commodities with which they are unfamiliar - they have no experience in trading/valuing the commodity in question. Hypotheses related to this issue developed by the authors in Chapter V focused on time and information requirements by subjects if they were to research preferences in a meaningful way to the end of formulating accurate valuation responses. In our search for research results relating to these hypotheses, myriad problems associated with such things as cognitive dissonance, mental accounts, information processing - more generally, bounded rationality - we were compelled to conclude that results from the received literature offered little that would support the notion that subjects, during

the relatively brief period of the CVM interview, could define their preferences for a new, unfamiliar commodity in any meaningful way - thus, our use of ROC's 1 and 2 developed in Chapter VI.

The familiarity issue, and our requirement for experience/familiarity with CVM commodities as a Reference Operating Condition, was the subject of considerable controversy at the Assessment Conference. Freeman (Chapter X) essentially accepted the familiarity/experience issue as being on equal footing with the hypothetical payment/incentive issue as a potential source of bias in CVM measures, and expanded the familiarity argument in the following way. In contrast to conventional theory, Freeman argues that individuals have more accurate knowledge of their preference orderings in the neighborhood of those consumption bundles that they have actually experienced. In instances where individuals are moved into unfamiliar regions of their preference orderings, accurate preference orderings - and therefore accurate valuations - will result only after the individual can learn (via trial and error experiences) about this "new" region of consumption bundles. Thus, if the CVM involves small changes around neighborhoods of experienced consumption bundles (the individual is, therefore, somewhat familiar with the commodity), valuation responses will be more accurate than for CVM studies involving changes (or new commodities) which move individuals to regions of preference orderings with which the subject has no experience.

V. Kerry Smith acknowledges the potential importance of the familiarity issue, but takes the argument along two somewhat different lines. First he argues that the relevant state of the arts is one wherein we can say little, qualitatively or quantitatively, about the implications of the familiarity problem inasmuch as we have no model of how individuals behave/respond in the CVM milieu; he notes Hoehn and Randall's (1984) interesting beginning in this regard, to which we would add the logic suggested by Freeman (Chapter X). Secondly, and somewhat curiously, Smith argues that, in accepting the ROC's which require that subjects be familiar with the CVM commodity and its (at least) indirect market exchange, we require that the subject's choice experience is the equivalent of his/her knowledge of the features (outcomes) of the implicit market; i.e., such CVM studies elicit the subjects' perception/estimation of implied market outcomes for hypothetical changes rather than the subject's personal valuation of the commodity.

V. Kerry Smith's latter point warrants a closer look. If the CVM commodity was a loaf of bread, the subject's knowledge of market outcomes (the price that bread commands

in the supermarket) would surely be reflected in the subject's bid. But the familiarity requirement for public goods is not this strong, nor is the requirement for indirect market experience. In Chapter VI's example of air quality in Los Angeles, satisfaction of the familiarity ROC was argued on the grounds that subjects were (i) aware of (familiar with) air quality differences in various areas in the basin, and (ii) that equivalent houses in areas with better air qualities would cost "more." Individuals may have rough ideas of how much more beach-side homes cost than the housing counterpart in Pasadena, but it would be heroic to assume their access to hedonic measures which attribute values to the myriad attributes of the beach-side house (proximity to beach, crime rates, etc., and air quality). Faced with the question: "Living in Pasadena, what would you pay for (beach-side) levels of air quality?", a basis for the subject's calculation of a market solution a la Smith is not readily apparent. Thus, while Smith's call for modeling efforts concerning individual behavior within the setting of the CVM is (and was, at the Conference) well-received, his assertion that CVM applications for commodities satisfying the familiarity ROC's imply the generation of implicit market outcomes, rather than an individual's revelation of preferences, is not (to the authors' minds) convincing.

Kahneman argues that the requirement of familiarity does not go far enough in terms of imposing limits on applications of the CVM which may lead to a priori expectations of reasonably accurate responses. In Chapter VI, the authors, in describing the implications of the ROC's, noted that the ROC's precluded the derivation of value estimates for unfamiliar, and uncertain, commodities, such as those related to option, preservation and bequeathment values. Kahneman suggests the use of a distinct ROC which precludes the application of the CVM for deriving any value with ideological content - i.e., only user values should be the subject of CVM applications. In support of his argument, Kahneman draws on the notion of "symbolic (or incoherent) demand." Symbolic demand reflects an individual's hierarchy of values which, Kahneman argues, must inject itself into any economic or political context. Manifestations of symbolic demand - manifestations of ideological "loading" - are seen in subjects' inability to differentiate between values attributable to related, but nonsubstitute goods; e.g., a subjects' inability to differentiate, in value terms, between improved air quality in area A, areas A and B, and air quality throughout the U.S. (this particular example of symbolic demand is found in Schulze et al. 1984, Chapter I). Thus, to the extent that familiarity and uncertainty ROC's do not eliminate all

possible applications of the CVM to commodities with ideological content, we are asked to expand the ROC's to preclude such applications.

© The learning issue. While inextricably related to the familiarity question discussed above, questions concerning "learning" are sufficiently distinct to warrant their separate treatment. At issue in these regards is the efficacy of various methods and techniques in assisting subjects in the CVM to first, more effectively research their preferences; and/or secondly, to more completely understand the nature of the contingent market and incentive-compatible behavior appropriate for that market. Methods/techniques of concern in these regards are: the iterative bidding process; the use of repetitive valuation trials; and more generally, the transferability of techniques used in laboratory experiments to applications of the CVM.

A recurring theme through Chapters III - VI is the authors' view that the iterative bidding process must be used in CVM applications if meaningful measures of subjects' maximum willingness to pay are to be derived. This admittedly strong view was based primarily on three arguments developed in those chapters. First, the heuristic argument (Chapters III and IV) that, at the outset, subjects may not fully appreciate the "all or nothing" character of the contingent market and that the bidding process "prods" the individual to more completely research his/her preferences vis-a-vis the contingent commodity; as in any auction, demands on the subject's judgment as to the extent to which he/she really wants the commodity, increase as the stated price increases. Secondly, results from experimental economics demonstrate that subjects require time and repetitive valuation trials before they begin to fully appreciate the nature and implications of the valuation process. Third, and finally, the considerable empirical evidence which demonstrates significant differences between initial, one-shot values and final values derived with the bidding process.

While acknowledging that initial, one-shot, bids may underestimate a subject's maximum willingness to pay, Mitchell and Carson (Appendix) reject the notion that the iterative bidding process solves the problem; in so doing, they challenge each of the three arguments used by us in developing our contrary conclusion. The heuristic "prodding" argument is turned 180 degrees to suggest that the bidding procedure may in fact "bully" subjects into bidding more, given their awkward social position of having to say "no" to the interviewer's inferred request for a higher bid. While agreeing that CVM scenarios should

include iterative elements which permit learning, Mitchell and Carson argue that the iterative trials of experimental economics are unnecessary to accomplish this end, and moreover, do not make the case for using the iterative bidding process. The necessary use of iterative trials in experimental economics, they argue, may well be related to the nonintuitive, second-price auction institution. In terms of one's understanding of the WTP format, they point to the data presented in Table 4.1 of Chapter IV (Volume I.B) which shows (for WTP trials) minor differences in bids across the repetitive trials. Finally, the interpretative weight of our empirical evidence demonstrating differences between initial and post-bidding values is implicitly challenged by Mitchell and Carson by the question: "To what does one attribute the observed differences: downward bias (as we argue) or a "bullying" effect?

Bishop and Heberlein (Chapter IX) also criticize the "categorical conclusion" regarding the need for iterative bidding suggested by us in earlier chapters. Like Mitchell and Carson, they point to the weak statistical tests in demonstrations of bid differences with and without iterative bidding processes and report results of their analysis of three bidding game studies wherein starting and iterated bids are positively correlated with hypothetical payment, but not correlated with actual cash payments. Referring to results from their Sandhill study, Bishop and Heberlein suggest that iterative bidding encourages subjects to exaggerate their willingness to pay; one should note, however, that only one iteration was used in their study. Finally, noting that iterative bidding precludes the use of mail surveys in application of the CVM, they suggest as an "ultimate conclusion" that the iterative bidding process may simply not be worth the trouble and expense.

In Chapters IV and VI, the authors devoted considerable attention to developments in experimental economics and the potential promise of laboratory methods/techniques used by experimental economists for structuring and testing questionnaires to be used in CVM field interviews; particular stress is given to the use of "Vickery Auctions" and tatonnement processes - basic methods used in experimental economics - as means by which more accurate responses might be obtained with the CVM.

Our enthusiasm for lessons learned from experimental economics, vis-a-vis their meaningful transferability to the CVM, was not totally shared by Conference participants. Bishop and Heberlein criticized our stress on the need to conduct laboratory experiments while ignoring the contributions of field experiments - a position supported by Arrow. In chiding the authors' "one-sided" emphasis on the

virtues of laboratory experiments they point to the highly simplified and artificial settings of all laboratory experiments, and question the transferability of such results to real-world situations - a criticism echoed by Mitchell and Carson as well as by V. Kerry Smith.

The emphasis given to Vickery auctions and the tatonnement process in Chapter IV was found particularly disconcerting by a number of Conference participants. In terms of the Vickery auction - a "discovery" viewed by Bishop and Heberlein as a red herring - Mitchell and Carson (Appendix) as well as Bishop and Heberlein (Chapter IX) acknowledge the effectiveness of the method in assessing institutional structures for private goods involving actual exchanges (see also, V.K. Smith, Chapter XI, Section 4.C), but fail to see how the method is to be used for hypothetical markets for public goods wherein exchange is impossible; in this regard, these authors argue that our reliance on the Coursey et al. (1983) experiment, involving the private good SOA, does not support our general conclusions. Given the nonintuitive format of the Vickery auction, and (as we report in Chapter IV) the repetitive trials required for subjects to learn incentive-compatible behavior implied by the format, both Bishop-Heberlein and Mitchell-Carson question how such repetitive trials are to be implemented within the CVM framework (see, also, Freemans' remarks in Chapter X). Iterative bidding, these authors maintain, does not substitute for the repetitive exchange trials of the Vickery auction format. Similarly, in terms of our suggested use of tatonnement processes as a part of the CVM, Bishop-Heberlein assert that, for hypothetical public goods of interest for the CVM, Groves-Ledyard procedures for implementing such processes may not cause respondents to reveal true preferences and may result simply in increased costs, increased confusion and lower response rates. In this regard, reliance on tatonnement processes for the large groups of subjects generally included in CVM studies "boggles" the minds of Mitchell and Carson.

While we accept the "Red Herring" comment of Bishop and Heberlein in the spirit of intellectual mischief in which it was intended, we do feel that the role of experimental economics in contingent valuation research has been misunderstood, most likely due to a failure in our exposition in Chapter IV. Rather than serving as guidance for the structure of hypothetical survey questions for the CVM, the demand revealing mechanisms developed by public choice theorists and experimental economists show how to obtain value estimates which are close to "true values" in laboratory situations. It turns out that even in the laboratory, it is fairly difficult to obtain "true" demand revealing values. First, one must use an incentive

structure such as a Vickery auction for private goods. However, this not sufficient. In addition, individuals must be given a number of repetitive learning trials to understand the auction mechanism and learn that demand revelation is their best strategy. Only by using both, a demand revealing mechanism and by allowing sufficient learning experience to accrue via repetitive trials, do about 70% of the subjects actually reveal demand in laboratory settings. Thus, based on their observations, the Bishop and Heberlein study (described in Chapter IX) which actually attempted to repurchase hunting permits likely did not reveal demand for hunting permits since no opportunity for repetitive learning trials was given to participants and subjects most certainly had no prior experience selling their hunting permits. It then follows that experimental economics sheds little light on Bishop and Heberlein's hypothetical values, but suggests their "true value" obtained from actual behavior may have been biased for reasons other than those acknowledged by them. The primary lesson from experimental economics is, therefore, concern methods by which values may be obtained which are demand revealing as a basis of comparison for alternative, hypothetical measures of value.

These discussions conclude our capsulization of the controversies surrounding the preference research issues: issues concerning the need for incentives for accurate valuations, the subjects' need for familiarity/ experience with CVM commodities, and the efficacy of iterative bidding and methods/techniques drawn from experimental economics for assisting subjects in their preference research processes. As to the implications of these discussions for the state of the arts of the CVM, conclusions in this regard are but deferred until we have considered other issues related to hypothetical bias. Thus, the authors' conclusion concerning issues related to preference research are given below in sub-section I.4.

2. The Comparability of WTP and WTA Measures. In Chapter VI, the authors submit as a Reference Operating Condition for assessing the accuracy of CVM values, the requirement: "WTP, not WTA, measures are elicited." (VI.E). The rationale for the authors' imposition of this ROC was based on two related lines of argument. In Chapter III (Section 4) we note that in spite of theoretical arguments (which relate to private goods) that WTA should equal WTP, empirical studies (Table 3.2 in Volume I.B) consistently demonstrate wide divergences between WTA and WTP measures; generally, estimated WTA measures are orders of magnitude greater than estimated WTP measures (Table 3.2 in Volume I.B). In Chapter IV (Section C), we argue that such observed disparities between WTA and WTP may be attributed

to cognitive dissonance, which, in the context of IV.C's discussions, is reflected (via the Davis et al. experiment) by subjects' failure to recognize dominant strategies in a Vickery auction, i.e., in some cases, iterative trials, whereby subjects learn that full demand revelation is their dominant strategy, results in the convergence of WTA to WTP measures. Such convergence was found to generally obtain (in the Coursey et al. experiment) under nonhypothetical circumstances, but not under hypothetical circumstances, an anomaly attributable to the lack of a market-like environment in the hypothetical experiments. In retrospect, we note the implications of this finding for earlier-discussed criticisms of our enthusiasm for the use of Vickery auctions in the hypothetical setting of the CVM (Section I.1.c). We also note the consistency of laboratory results with Randall et al.'s (1983) argument (also, see Randall's arguments in Chapter VIII) that WTP underestimates "true" values while WTA overestimates such values.

A considerable amount of interesting and constructive criticism of our WTA/WTP arguments and conclusions was offered by Conference participants. First, various participants questioned our attribution of WTA-WTP differences in hypothetical settings to "cognitive dissonance" and our implied reliance on results from iterative trials in one experiment (the Coursey et al. (1981) experiment) as a means for eliminating cognitive dissonance. Thus, Bishop and Heberlein question the lack of symmetry of learning effects from iterative trials on WTP and WTA measures in the Coursey et al. experiment: iterative trials affect WTA measures but, seemingly, not the WTP measures. Moreover, Freeman (Chapter X) questions our attribution of WTA-WTP differences to "cognitive dissonance" and the link between cognitive dissonance and our learning-via-iterative-trials arguments. In this regard, cognitive dissonance refers to the beliefs of a subject (on which preferences are based) which are persistent over time and in the face of contrary "facts," and which are changed by subjects via their selection of information sources which are consistent with "desired" beliefs (Ackerlof and Dickens, 1982, p. 307). Thus, all else equal, the cognitive dissonance argument would lead us to expect little if any changes in bids with additional information (learning; Arrow, 1932). In these terms, a subject's lack of understanding of a Vickery auction (or any other valuation institution) may be viewed as distinct from an individual's value-related beliefs which are subject to cognitive dissonance. Our "evidence" from experimental economics, with reference to iterated trials, then suggests the subject's need to learn a "new" institution, but does not

necessarily establish cognitive dissonance as an explanation for WTP-WTA differences in nonlaboratory experiments (Table 3.1 in Volume I.B) as we infer in III.4 and IV.C.

As to our observations of large WTP-WTA differences, this issue is addressed by Randall in Chapter VIII wherein he argues that, for a fairly wide range of contingent market designs, one can confidently expect that reported WTP and WTA measures will, respectively, understate and overstate an individual's true valuation. The generality of this conclusion (which we implicitly accepted in Chapters IV and VI) is challenged by Freeman as inconsistent with the "familiarity" issue discussed above in I.1.b: in instances where individuals lack accurate information regarding their preferences - the CVM commodity takes the individual to preference orderings beyond the neighborhood of experienced consumption bundles - individuals may make errors in any direction, i.e., WTP or WTA may be greater or less than values that would result from experience with the new commodity bundles. Along these lines, it is interesting to note that in Bishop and Heberlein's Sandhill study in (Volume I.B, Chapter IX, Tables 9.2 and 9.3) hypothetical WTA values are less than cash offers ("true" valuations?) and WTP measures exceed cash offers; they also note large WTP-WTA differences in cash offers as well as offers involving hypothetical payment/compensation.

Kahneman strongly supports our "use WTP, not WTA" ROC, but first suggests that it be generalized and second, rationalizes the generalized ROC along different lines. His generalized ROC is: use the CVM only for commodities that have a "transactions structure"; do not use the CVM for commodities that have a "compensation structure." A "transactions structure" refers to a commodity-exchange context easily associated with voluntary exchange - one pays for a commodity or action which makes him/her better off. A "compensation structure" refers to a commodity-exchange context wherein overtones of involuntary exchange are present - how much you must be paid to accept more polluted air. The rationale for Kahneman's suggested ROC is his appeal to "prospect theory" which, in essence, assumes that individuals evaluate gains and losses differently; more specifically, it assumes that individuals value losses disproportionately higher than (identical) gains. Thus, one would expect a subject's valuation of a gain (WTP) to be substantively different from his/her valuation of a loss of identical magnitude (WTA).

We must confess that the link between Kahneman's rationale and his recommended ROC is not perfectly clear. One might appeal to prospect theory as a means for explaining why WTP and WTA measures should be expected to differ, but this

would not argue for or against the preferability of one measure over another. It might argue, however, that one must use value functions based on WTP for valuing environmental improvements, but that a different value function, based on WTA measures, must be used in valuing (costing) environmental degradations; i.e., one cannot move toward the origin along a "benefit" curve. But this observation could apply with equal force to our conclusion that WTP, not WTA, measures be obtained via the CVM. Our rejection of WTA measures derived with the CVM is, upon close inspection, based on the argument that they are less "stable" than WTP measures; i.e., they are more affected by iterative trials, questionnaire design, etc. We do not make the case that cognitive dissonance, or other psychological/economic factors, are more or less relevant for WTP or WTA measures. Large differences observed between the two measures obtain in CVM studies, and that WTA measures are "high" may be inferred as a motivation for our recommended ROC.

Vernon Smith (Chapter XII.E) casts the WTP/WTA argument in a different light. He asks if we are not confusing WTA/WTP differences for the same individual with such differences among individuals. He notes that such differences among individuals, even if large, should not be disturbing since such differences provide the basis for exchanges - large differences may simply imply a low volume in market trading. In terms of WTA-WTP differences for the same individual, Smith seemingly rejects the assumption of small income effects which underlies the Willig (1976) argument leading to approximate equality between WTP and WTA. His experiment demonstrates, first, that several subjects persistently reported WTA and WTP that were substantively different; secondly, his experiment demonstrates that, despite differences in WTA and WTP values reported by individuals in the experiment, when such values are used in a market demand/supply context, the resulting prediction of post-trade market-clearing prices is more accurate than predictions drawn from expected utility theory. Thus, Smith argues that empirical evidence belies the theoretical expectations of "equal" WTP and WTA for individuals - note here the consonance of this observation with those of Kahneman - but that in a market context such differences across individuals can result in accurate pre-trade predictions of actual (post-trade) prices (valuations) at which commodities are traded.

There are some particularly interesting implications of Vernon Smith's argument which warrants further examination. Consider the following data from Smith's experiment given in Figure 12.4 (Volume I.B).

Measure	Trial:		
	1	2	3
(a) Predicted price from the expected utility model	\$ 1.25	\$ 1.25	\$ 1.25
(b) Predicted price from WTA and WTP	1.25	1.43	1.148
(b) Predicted price from WTA and WTP	1.25	1.43	1.148
(C) Actual, post-trading equilibrium price	1.30	1.51	1.52
(C) Actual, post-trading equilibrium price	1.30	1.51	1.52
(d) Sum of WTA	16.147	10.62	13.86
(d) Sum of WTA	16.147	10.62	13.86
(e) Sum of WTP	12.142	10.80	12.214
(e) Sum of WTP	12.142	10.80	12.214

Smith's experiment suggests a method for addressing accuracy/calibration questions related to CVM measures. For example, for a commodity which is exchanged in the market, a CVM study might be conducted which collects WTP and WTA measures from each subject. Demand (supply) curves are estimated from WTP (WTA) measures. Comparison of the resulting predicted price with actual market price has obvious implications for the accuracy of CVM estimates of value. Most importantly, Smith's experiment provides empirical weight for Kahneman's argument that benefits (the area under a WTP-demand curve) attributable to an environmental improvement may be expected to differ from costs (the area under a WTA-supply curve) for an environmental degradation. In this regard, the reader should note the different "areas" (sums) for WTP-benefits and WTA-costs implied from Smith's results given above, particularly values (d) and (e) for the first trial in Smith's experiment.

Related to Vernon Smith's argument is the point raised by Rosen (XIII.D). Rosen argues that WTP/WTA differences may in fact reflect "selectivity" i.e., populations from which WTP and WTA measures are taken are not homogeneous populations. In this regard, Rosen points to Brookshire et al.'s earthquake study: those living on a fault may well be expected to value earthquake risks differently from those who do not live on a fault.

Based on these interesting exchanges, it would appear to us that the following conclusions are relevant for the WTP/WTA issue. First, we agree with Freeman and Bishop-Heberlein that a compelling case has yet to be made as to the general relationship between WTA and/or WTP measures and "true" valuations; certainly our attribution of such differences to cognitive dissonance is little more than an assertion. As is argued below, this implies the need for considerably more attention being given to the collection and analysis of psychological and attitudinal data in future CVM studies. Secondly, we agree with Freeman that the above-discussed "familiarity" issue is relevant for assessments of WTP/WTA differences; however, the little available empirical evidence does not support the notion that such differences are systematically related to the subject's familiarity with commodities. Referring to Table 3.2 (Volume I.B), WTA/WTP differences ranged from 2:1 to 5:1 in experiments involving private goods (goose permits in Bishop and Heberlein (1979) and a better-tasting substance in Coursey et al. (1983)). Thirdly, we find Kahneman's "prospect theory" arguments to be, at a minimum, intuitively appealing, and certainly consistent with (if not supported by) considerable empirical findings. The notion that individuals value gains (from transactions structures) differently from losses (from compensation structures) may not, however, lead one to reject CVM applications to the estimation of WTA values; rather, it may suggest particular uses of WTP and WTA values: WTP for gains and WTA for losses. Finally, we concur with Bishop-Heberlein (Chapter IX) that the "burial" of WTA may be premature and that additional research is required which focuses on explanations of WTP-WTA differences. Meanwhile, it appears to us, our ROC "use WTP, not WTA" may serve as an operationally useful guideline for ongoing research with the CVM.

3. Attitudes vs. Intended Behavior. In Chapter V (Section E) the authors reviewed the "attitude versus intended behavior" issue raised by Bishop and Heberlein (1979 and 1983) which focused on the question: do CVM value measures reflect attitudes rather than intended behavior, and to what extent do attitudes correspond with intended behavior? Essentially, we adopted Randall et al.'s (1983) position that

since CVM questions asked for intended behavior rather than attitudes, problems of correspondence between attitudes and behavior were likely minimized. We acknowledged, however, the relevance of Ajzen and Fishbein's (1977) design criteria for improving attitude-behavior correspondence (specific targets, actions, context and timing). As an aside, Bishop and Heberlein (Chapter IX) may have found our treatment of this subject to be uninformed or shallow, but in light of the major emphasis given results from psychological studies throughout Chapter V, we find ourselves nonplussed by their assertion of our "Indifference and hostility" (Chapter IX, Section E.7) to the relevance of psychological research for economic inquiry. We confess, however, to understating the importance of attitude-behavior issues in psychology research.

Bishop and Heberlein's elaboration of the attitude-behavior issue in Chapter IX.E.7, is insightful, illuminating, and we believe, rich in its implications for the state of the arts of the CVM. Their major focus is on attitudes (as they relate to reported WTP and behavior (actual payment of WTP) and the factors which result in close correspondence between the two. Attitudes are determined by the interaction of three components: cognition (dispassionate facts/beliefs), affectation (evaluative/emotional reactions to cognitive information) and intended behavior (intentional "conclusions" derived from affective responses to cognitive information). Interaction between these three components is of primary importance; e.g., an affective change may motivate the individual to acquire more information (a cognitive change) which may then lead to a change in intended behavior. They argue, that a cash offer for a goose/deer license may elicit an affective response, and therefore a behavioral response, that is distinct from the affective response to a hypothetical offer -- witness their observed differences between valuations involving real and hypothetical payment. This analogy is consistent with Kahneman's arguments concerning WTP-WTA differences: WTA questions involving compensation structures elicit affective responses that differ from those elicited by WTP questions involving transactions structures.

Of primary interest are the factors which lead to close correspondence between attitudes and behavior. As an example in this regard, define AC (awareness of consequences) as a measurable manifestation of the cognitive component of attitudes vis-a-vis a CVM "commodity," and AR (acceptance of personal responsibility) as a measurable manifestation of the relevant affective component of attitudes. One can then define design and analytical criteria for assessing the probable

correspondence between reported willingness to pay and what a subject might actually pay for a CVM commodity. Design criteria are those proposed by Ajzen and Fishbein (1977) to which we add questions related to AC and AR (see Bishop and Heberlein's examples in IX.E.7). In analytical terms, one's assessment of the probable correspondence between attitudes and behavior -- which relates to the probable accuracy of estimated values -- is based on the values of AR and AC variables. For the commodity in question, the greater is a subject's awareness of consequences (familiarity with the commodity?) and acceptance of personal responsibility, the greater is our expectation of close correspondence between attitudes and behavior (and, therefore, the more accurate the resulting measure of value).

As noted above, Bishop and Heberlein's elaboration of the attitudes-behavior issue allows for sharp focus on the need for attitudinal information for assessments of CVM results as well as for the types of information that would be useful in these regards. While not affecting the weight of their contribution, however, their discussions raise several questions of interest for our broad state of the arts assessment of the CVM. First, in operational terms, we simply note in passing the indexing task implied by their proposed criteria for correspondence between attitudes and behavior; e.g., what constitutes "high" values for AC or AR variables? Secondly, absent from their discussions is the relationship between attitude-behavior criteria and the other psychology-related issues discussed in Chapter V and reviewed by them. As an example, Bishop-Heberlein's discussion of the three interactive components of attitudes would seem to bear directly on the familiarity issue discussed above. If the cognitive component is empty -- subjects are unfamiliar with the commodity, or have little in the way of relevant facts/beliefs -- what might we expect in terms of affective responses and formulated behavioral intentions? A response to this question is implied in Kahneman's discussion of starting points (Chapter XII.C): subjects are simply incapable of assigning values to the commodity. Bishop-Heberlein's counterpart to this conclusion would seem to be: low AC values imply divergence between attitudes and behavior and thus (one supposes) inaccurate values.

A third question raised by Bishop and Heberlein's attitude-behavior discussions concerns the conflict between their position on the viability of estimating such things as option and existence values with the positions taken by us in Section VI.E and by Kahneman in XII.C. Appealing to familiarity/experience factors underlying our ROC's 1 and 2, we argue that one can expect a priori that such values must involve (using Freeman's model, Chapter IX) consumption

bundles well beyond the neighborhood of bundles with which the subject has experience; thus, our rejection of uses of the CVM for estimating such values. Kahneman rejects the use of survey methods for valuing all but user values -- explicitly excluding option/preservation values -- in his discussion of "symbolic demand".

Responses to questions related to ideological values, he argues, must reflect the subjects' hierarchy of values which tend to be injected into responses involving political or economic content. While acknowledging, first, that assessments of the validity of existence values via the CVM will not be easy and, secondly, that results from field experiments hold little promise for the use of the CVM in deriving such values, Bishop-Heberlein seemingly take the position that the CVM might indeed be used for estimating option or, particularly, existence values. The relative accuracy or meaningfulness of such measures would be assessed via analyses of the correlation between reported existence values and AC/AR variables. In their acid rain example, high existence values would imply (i) "high" awareness that acid rain damages will affect future generations (an AC variable) and (ii) a "high" indication that the subject feels personally responsible for reducing these effects (an AR variable; see IX.E.7).

In terms of the different positions concerning the use of the CVM for nonuser values described above, we should acknowledge possible exceptions to our conclusion that the familiarity/experience ROC's preclude the estimation of nonuser values; but we do not find Bishop-Heberlein's arguments (and the acid rain example) compelling in this regard. "High" AC values, which indicate familiarity with the acid rain problem, and "high" AR values simply do make their case: other values in the affectation "account" -- perceptions of how the subject is affected in a "user value" sense -are relevant. At issue then is the subject's ability to differentiate between that part of his/her affective reaction to acid rain that is attributable to personal effects (a use value) and, generally, more altruistic affective reactions vis-a-vis future generations. Echoing Kahneman's notion of symbolic demand, it is this latter process, a process with which we expect the subject to have little experience, that we question. We would expect, a la Kahneman, that the sum of the user and nonuser parts will greatly exceed the subject's valuation of the whole.

4. Hypothetical Biases in the CVM: Conclusions In the authors' view, discussions at the Assessment Conference were particularly productive in giving perspective and context to the myriad issues concerning hypothetical bias discussed in

Chapters III - VI. As noted in those Chapters, the potential for hypothetical bias in the CVM enters through the hypothetical nature of payment as well as the hypothetical commodity and the institution within which the commodity is exchanged -- the contingent market. We now ask, in light of the Assessment Conference, what is the state of the arts of the CVM in terms of the potential magnitude of hypothetical biases?

In terms of hypothetical payment, we view the potential for related biases with a great deal more equanimity than that suggested in the conclusions to Chapter V. In this regard, Mitchell and Carson's arguments as to the weakness of empirical results used by us in arriving at our more pessimistic conclusions are well made. The weight of the "incentives for accuracy" argument must, at worst, be questioned in light of Vernon Smith's experiments, wherein WTP/WTA interrogations were "good" predictors of market outcomes, and the results from comparison studies wherein the CVM generated value estimates that were remarkably close to estimates derived from indirect market methods (holding the question of the accuracy of any method aside, for the moment). We concur with Arrow's observation that hypothetical/real payment differences may not be as serious as one might fear: well designed survey instruments wherein the exchange setting is "pseudo-real" may indeed elicit real-like results. This is not to argue that incentives/hypothetical payment issues are not relevant; it is to argue that, first, the jury is still out -- it remains an open issue -- and, second, that some promise exists for structuring CVM instruments in ways that mitigate, if not eliminate, the magnitude of payment bias.

Within the rubric of "hypothetical bias," we find the most prominent source of bias to arise in instances wherein the CVM commodity, within a contingent exchange setting, is largely unfamiliar to the subject -- the subject has no experience in viewing the commodity within the context of trade-offs. In Freeman's terms, the effect of the CVM is to move the individual to areas of his preference orderings that are far removed from neighborhoods of consumption bundles with which the subject is familiar. Our lack of models concerning subjects' behavior in the CVM setting notwithstanding, we see in Freeman's rudimentary modeling efforts, as well as in Kahneman's notion of symbolic demand and Bishop-Heberlein's discussions of the roles of attitudes, the bases for reasserting our contention that, for state of the arts applications of the CVM, (i) participants in the CVM must understand (be familiar with) the commodity to be valued (our ROC Number 1) and (ii) subjects must have had (or be allowed to obtain) prior

valuation and choice experience with respect to consumption levels of the commodity (our ROC Number 2).

In terms of learning issues, final state of the arts conclusions concerning the efficacy of iterative bidding processes and laboratory methods/techniques for applications of the CVM must be softened considerably from the tone of earlier conclusions offered in Chapters III - VI. We find impressive the substantive effect on bids that result from the iterative bidding process in studies involving, not just the small samples of concern to Mitchell and Carson, but large sample sizes. In our view, iterative bidding does result in substantively higher bids. Iterative effects notwithstanding, Mitchell and Carson, as well as Bishop and Heberlein, are obviously correct in pointing to the lack of evidence that would support (or reject) the attribution of such effects to the preference research processes as asserted by us in Chapters III - V; moreover, we must acknowledge the substance of Bishop and Heberlein's observation that the parallel between the iterative bidding process and the iterative valuation trials used in laboratory experiments, implied by our discussions in IV, is without obvious substance. Nor, it seems fair to say, has the attribution of iterative bidding effects to Mitchell and Carson's "bullying" or "social awkwardness" motives been established. Thus, all that can be said at this point in time is that iterative bidding rather consistently results in higher CVM valuations, but we are unable to explain such differences.

Bishop and Heberlein's lament that economists involved in CVM research are woefully ignorant of research results in the related, and certainly relevant, field of psychology extends with equal force to economists' general ignorance (until only very recently) of developments in experimental economics; the authors concede their general ignorance in this area prior to the development of this book. As the novice enters the literature of experimental economics, he/she must be struck with the impressive developments made in that field which relate directly to the most perplexing questions facing the CVM practitioner: how does one establish incentive structures; how do subjects learn; how does one elicit preference revelation? The real "lessons" from experimental economics of unquestionable importance for the development of the CVM are found in two principal areas. First, laboratory methods can provide us with a relatively inexpensive and efficient method for conducting experiments concerning design and conceptual questions of relevance for the CVM; examples in these regards are questions concerning strategic bias, WTP-WTA differences, effects of psychological variables on subject valuations, etc. Secondly, and of particular importance, developments in

experimental economics may be provocative -- challenging -- to CVM researchers in terms of stimulating new and imaginative lines of inquiry concerning persistent problems encountered with the method. In these regards, the issue is not, for example, whether or not the Vickery Second Price Auction per se will "work" in applications of the CVM; rather, the issue is: can the CVM be structured so as to better provide incentives for true revelations of preferences (as an interesting initial effort in this regard, see Bishop and Heberlein's experiments with a Fifth Price Auction in Chapter IX). As another example, can we (should we) be experimenting with repeated visits (repeated "trials") with CVM subjects, with questions designed to help them learn incentive-compatible behavior vis-a-vis a contingent market?

Thus, lessons from experimental economics are clearly relevant for our state of the arts assessment of the CVM: they indicate the lack of substantial progress made in the method's development in important areas concerning subjects' learning/understanding of incentive structures. Such lessons are not, however, a panacea for resolving the problems of the CVM. Earlier-noted comments by Conference participants concerning our over-emphasis on the ready transferability of methods/techniques used in experimental economics to applications of the CVM for valuing public goods are well made, as are the reminders by Arrow and by Bishop-Heberlein of the important role of field experiments for improving the state of the arts of the CVM.

Turning now to the WTP-WTA issue, relevant state of the arts conclusions were suggested in the closing paragraphs of Section 1.2 above. V. Kerry Smith's call for theoretical inquiry as to subjects' behavior in the contingent market setting is particularly appropriate for efforts to explain WTP-WTA differences. In this regard, see the contrast between Randall's theoretical model, which relies on subjects' perception that their responses influence policy, wherein WTP (WTA) understates (overstates) "true" valuations, and Bishop-Heberlein's contrary evidence as well as Freeman's model which suggests that, for "unfamiliar" commodities, WTP or WTA relationships to true valuations cannot be determined a priori. While we find compelling, on deductive as well as intuitive grounds, Kahneman's argument that subjects value losses differently than gains, we are concerned with the fact that WTA measures appear to vary much more than WTP measures in response to such things as iterative trials. Thus, in operational terms, i.e., as we await results from further theoretical and empirical research concerning this question, we maintain our conclusion suggested in Chapter VI which states that WTP, not WTA, measures should be estimated with the CVM.

Finally, the state of the arts of the CVM in terms of our appreciation of the attitude-behavior issue is, in our view, greatly enhanced by Bishop-Heberlein's discussions in Chapter IX. Means by which the accuracy of CVM measures, in terms of the correspondence between attitudes and actual behavior underlying reported willingness to pay, are directly implied by the interactive relationships between attitudinal components and behavior. While implementation problems remain for resolution, one can see in Bishop-Heberlein's exposition the essential framework for deriving empirical measures for cognitive and affective components of attitudes and, at least conceptually, their use in deriving indices of attitude-behavior correspondence.

J. THE ACCURACY OF CVM MEASURES OF VALUE

1. Overview of the "accuracy" issue. Recurring throughout Part I of Volume I.B, as well as throughout Conference papers and discussions, is reference to a subject's "true" valuation of a public good such as an environmental change. Thus, our standard for accuracy in values derived from the CVM is a subject's reported valuation that reflects a "true" revelation of preferences vis-a-vis the CVM commodity. In this regard, our appeal to market institutions as a framework whose structure we hope to simulate in the process of applying the CVM is motivated by our desire to capture, in applications of the CVM, the incentives for preference revelation that our theories lead us to expect from a market context. In the market context, individuals must introspectively balance the utilities foregone as a result of paying for a good with the utilities gained from acquiring the good; to this end, he/she must, however "completely," search his/her preferences for the good in question vis-a-vis all other possible goods and their prices (relative to his/her income). Thus as has been extensively argued above, the importance for assessments of the CVM of such themes as the subject's familiarity with a commodity (for the preference "search", or research process) and the credibility of payment and payment modes to the subject (for meaningful subjective assessments of implied trade-offs).

In these regards, we must reiterate our earlier-noted concern with V. Kerry Smith's interpretation of our ROC's related to these themes as requiring that the value derived in the CVM be the subjects' estimation of market outcomes as opposed to the subjects' preference revelations; ROC's per se are discussed below. ROC-1 requires that the subject have some familiarity with the CVM commodity and ROC-2 requires some choice experience, direct or indirect, with respect to consumption levels of the CVM commodity. These conditions then loosely require that, as in Freeman's arguments, the consumption bundles (including the CVM commodity) that the subject is hypothetically evaluating are within neighborhoods of consumption bundles with which he/she has had experience. Thus, our concern with accurate revelations of preferences leads us to require that choice setting which is analogous to a market setting, and which is consistent with the expectation that the subject is capable of meaningful searches of preferences. To require an "informed" choice setting does not, in our view, imply that the CVM application must then elicit the subjects' introspective estimate of solutions of a hedonic market.

Given that our standard for CVM values is the true revelation of a subject's preferences, the primary question

becomes: how do we measure that standard? Obviously, if we had a "true" value, assessments of the accuracy of CVM measures vis-a-vis this standard would be straight-forward. The state of the arts relevant to such measures is such that, aside from limited results from laboratory and field experiments involving private goods, these measures are not available. Therefore, in Chapter VI the question of the accuracy of CVM values is addressed in the following indirect and inferential manner. First, we note the literature that suggests that, for ordinary demand studies based on "hard" market data, estimates may involve errors (the range for accuracy might be) on the order of $\pm 50\%$ or more. V. Kerry Smith (Chapter XI) expands on this argument, arguing that much of economist's "hard" data may be subject to the same type of criticism concerning, e.g., hypothetical and reporting biases as those leveled at the CVM. Such errors are generally attributable to such things as assumptions concerning the distribution of error terms and functional forms. Secondly, from these data we then infer that econometric value estimates based on indirect market methods would involve ranges of error no less than those in ordinary demand analyses, i.e., one can argue, at most, that indirect market methods yield value estimates which would encompass "true" values within the range $\pm 50\%$. Thirdly, appealing to the concept of "reference accuracy," we note that received studies demonstrate that biases associated with starting points, payment vehicles, information and iterative bidding could result in errors as large as $\pm 50\%$ in CVM studies.

In retrospect, we might well have stopped our arguments here: available evidence suggests that either the CVM or indirect market methods may yield estimates of "true" preference revealing values within a range no better than $+ 50\%$. We carried these arguments a step further, however, in addressing the following question. Noting -- uncritically, it must be acknowledged -- cited instances wherein economists quite comfortably impute accuracy to market-based estimates of value, we implicitly construct the following strawman: suppose that indirect market methods yield accurate results -- "accurate" within the range $\pm 50\%$; are value estimates from indirect market and contingent valuation methods different? We continue by positing that if they are not different, then the accuracy of indirect market values implies the accuracy of CVM values. Referring to the fifteen CVM-Indirect Market study comparisons given in Table 6.12 (Volume I.B), and noting that ranges ($\pm 50\%$) for accuracy of CVM values overlap with those for indirect market methods in 13 comparisons, we then conclude that, for commodities which are amenable to application of indirect

market methods (a caveat then used to form ROC's), the CVM may yield value estimates that are as accurate as (the assumed accurate) values derived from indirect market methods. It should be noted that any specification for the magnitude of errors associated with the use of the CVM is premature at this time. We choose $\pm 50\%$ as a means for focusing attention on what is, in our view, an interesting approach for assessing the accuracy of CVM measures.

In many ways our discussions of accuracy achieved their intended purposes: they certainly received the attention of Conference participants; most importantly, they succeeded in initiating a dialogue focused on how future research might address calibration and accuracy issues. Constructive criticism of our discussions of accuracy offered by Conference participants may be seen as involving the following three sets of issues.

2. What is Accuracy? The first set of issues involves the question as posed by Arrow: what do we mean by "accuracy" and what level of accuracy is it reasonable to expect from applications of the CVM? In response to these questions, Arrow offers four observations: (i) referring to hypothetical issues, the reality with which economics (and other social sciences) deal, involves counter-factual lines of deduction -- statements comparing actions with states that "would" hold, but in fact do not. Our concern is with questions of the form: what would we do if reality were marginally different (e.g., if income were one unit higher)? In virtually, all cases, the "truth" relevant for these questions can never really be known; (ii) inaccuracies in real-world efforts to estimate individual preferences via demand analyses based on "hard" data are probably best seen in the fact that half of the "new" products put on the market fail. (iii) our colleagues in medical and engineering sciences consider, as a matter of course, estimates producing errors on the order of one to ten (one order of magnitude, see VI.D) to be normal; (iv) therefore, it is not clear that we should be disturbed if our value estimates are thought to be within $\pm 50\%$ of true values, or $\pm 100\%$. Ranges of error of 3:1 or 5:1 may pale in significance when compared to those reflecting technical ignorance in most environmental fields.

V. Kerry Smith also stresses Arrow's point that we can never know "true" valuations. Indeed, in our general scientific inquiry we never prove hypotheses, we fail to reject them. Arrow's reminder of the limitations of "hard" data vis-a-vis their use in estimating value is expanded by Smith along interesting and provocative lines. In Table 11.1 (Chapter XI), he demonstrates the potential for strategic and hypothetical biases (broadly defined) in

various sources which are generally thought to produce "hard" -- accurate -- data.

As an aside, we are compelled to note the contrast between Arrow's and Smith's arguments and the framework for considering the question of accuracy offered by Freeman (Chapter X.E). Define B as a subject's response to a CVM question and assume that B is a random variable with mean, B' ; B^* is the individual's true valuation. Freeman's suggested approach for analyses of accuracy is then one which focuses on $B'-B^*$ and on the variance of $e = B-B'$. He distinguishes between "biases" -- $B'-B^*$ differences attributable to starting points, information, etc. (the topic of Section A above) -- and random errors reflected in $B-B'$ differences, where random errors result from the hypothetical character of the CVM (the substance of Section C above). With biases eliminated by questionnaire design, and assuming that e is normally distributed with zero mean, large samples (which would result in $e = 0$) may result in $B' = B^*$. In the light of our earlier discussions, the application of Freeman's approach involves two major questions, satisfactory responses to which elude the authors. First, on what basis does one argue in a compelling way that the many sources for hypothetical biases are random and, particularly, normally distributed with zero mean? Most importantly, and directly related to Arrow's and Smith's arguments, how does one divine the "truth" -- whence comes B^* which critically serves as the basis for assessing the effectiveness of questionnaire design in eliminating "biases"? In the scientific literature, the concept of measurement accuracy rejects the notion that "true" valuations can be known, the result of which is a focus on removing demonstrable errors.

Finally, Bishop-Heberlein's arguments have implications for the question: what is accuracy? In terms of the accuracy of values derived from the CVM, their discussions would seem to imply that accuracy turns on the correspondence between attitudes and behavior, wherein such correspondence might be in some sense measured by Azjen-Fishbein criteria (vis-a-vis questionnaire design) and by cognitive and affective variables. In passing, we note their second (tongue-in-cheek) criterion for accuracy which was suggested at the Assessment Conference: "good enough for government work", which might (quasi-seriously) be taken to mean that order of magnitude estimates may be regarded as "accurate" for some applications of the CVM.

3. Reference Accuracy and public good values. In our efforts to couch the accuracy issue in terms of "Reference Accuracy" -- accuracy is defined in terms of biases resulting from deviations from Reference Operating

Conditions -- the approach per se was well received by Conference participants; our exposition of a numerical application of the approach was not. In this latter regard, our $\pm 50\%$ argument was seen as "weak" by Freeman, as being "ad hoc" by Rosen, and unconvincing by Mitchell and Carson. Referring to the CVM as well as indirect market methods, V. Kerry Smith questions the extent to which any error range can be imputed to estimated value measures given the present state of our knowledge. The basis for much of the expected criticism of our (no better than) $\pm 50\%$ reference accuracy range for CVM measures reflects several related arguments which, we of course concede, are well made. Mitchell-Carson, Bishop-Heberlein and Rosen point to the fact that well-designed CVM studies need not include biases resulting from starting points, payment vehicles, information and/or iterative bidding. Indeed, our discussions above in Sections H and I suggest that payment cards can be structured so as to mitigate or eliminate starting point biases; payment vehicle bias may be a misnomer -- mode of payment may be inextricable from the commodity; and, particularly for "familiar" goods, information issues may be amenable to control by questionnaire design. Thus, these individuals argue, demand studies using the CVM (or indirect market methods) are not of equal quality, as is implied by our general statement that reference accuracy for the CVM may be no better than $\pm 50\%$. To these arguments Mitchell and Carson add the observation that sampling errors, discussions of which were excluded from our assessments of the CVM, must also be considered -- sampling errors alone could result in errors of $\pm 50\%$.

4. The need for accuracy or calibration research. In the physical sciences, Reference Accuracy, based on ROC's, is the accepted practice for evaluating the precision of instruments for measurement. Generally speaking, Conference participants were supportive of our efforts in Chapter VI which were designed to initiate thought and research concerning means by which ROC's might be defined and by which we might measure the error implications of CVM applications wherein one or more of the ROC's are not satisfied. Thus, Arrow calls for more field and laboratory experiments designed to establish conditions under which reasonably defined accuracy in the CVM might obtain, a call echoed by Rosen who, in addition, feels that replications of CVM studies might be useful in these regards. Bishop-Heberlein appeal for research designed to calibrate errors with the extent to which ROC's are satisfied. V. Kerry Smith's insistence on the need for modeling efforts is joined with his observation of our lack of knowledge as to how violations of ROC's affect subjects' valuations.

Of course, the need for standards against which the accuracy of CVM values might be assessed underlies our suggested ROC's. Given the critical importance of ROC's for the use of Reference Accuracy, attention is now turned to an evaluation of those conditions.

5. The Reference Operating Conditions. There are at least two requirements for estimation and use of Reference Accuracy for the CVM: the specification of Reference Operating Conditions -- the conditions or circumstances which limit the accuracy of a measurement tool; and the magnitude of errors which result from failure to satisfy any given ROC.

Consider, first, the problem of specifying ROC's relevant for the CVM. That ours is not the last word on ROC's relevant for the CVM is made clear by ROC's explicitly or implicitly suggested by Conference participants. Referring to Table 3, ROC's 1 through 14 are those suggested by us in Table 6.13; ROC Number 8 was implied in our discussions of the $\pm 50\%$ Reference Accuracy range for the CVM but, for reasons which now escape us, was not explicitly included as an ROC. ROC's 5-7 are those suggested by Kahneman -- note the overlap with ROC's 4 and 5. Mitchell and Carson suggest, based on referenda and psychological research, ROC 9 (and concur with ROC's 1, 3 and 14). A choice for an ROC Number 10 is implied by the apparently contradictory positions of Randall, who would require subjects to view the CVM process as a real opportunity to influence policy, and Arrow, Freeman (1979) and, we should add, Rosen, who would view a subject's perception of the CVM process in such a real, nonhypothetical way as possibly inviting strategic responses. Finally, Bishop-Heberlein's discussions imply ROC 11.

TABLE 3
ALTERNATIVE REFERENCE OPERATING CONDITIONS

Reference Operating Condition	Measurement Error When ROC is not Satisfied
1. Subjects must understand, be familiar with, the commodity to be valued.	?
2. Subjects must have had (or be allowed to obtain) prior valuation and choice experience with respect to consumption levels of the commodity.	?
2. Subjects must have had (or be allowed to obtain) prior valuation and choice experience with respect to consumption levels of the commodity.	?
3. There must be little uncertainty.	?
3. There must be little uncertainty.	?
4. WTP, not WTA, measures are elicited.	? ± 300%
4. WTP, not WTA, measures are elicited.	? ± 300%
5. (Kahneman) Valuations must involve transaction structures, not compensation structures.	? ± 300%
5. (Kahneman) Valuations must involve transaction structures, not compensation structures.	? ± 300%
6. (Kahneman) CVM values obtained must relate to use, with minimum ideological content.	?
6. (Kahneman) CVM values obtained must relate to use, with minimum ideological content.	?
7. (Kahneman) Payment vehicles must be well defined and credible vis-a-vis the CVM the CVM commodity; values obtained with one vehicle may not be interpretatively	?

"transferred" to those which we would obtain with other vehicles.

8. CVM applications must involve: ?

8. CVM applications must involve: ?

(i) No basis for starting points or anchoring; ?

(i) No basis for starting points or anchoring; ?

(ii) "appropriate" information concerning the commodity and the valuation process; ?

(ii) "appropriate" information concerning the commodity and the valuation process; ?

(iii) initial, noniterated valuations. ?

(iii) initial, noniterated valuations. ?

9. (Mitchell-Carson, from referenda/ psychological research):

9. (Mitchell-Carson, from referenda/ psychological research):

(i) Subjects must be given as simple a choice as possible; ?

(i) Subjects must be given as simple a choice as possible; ?

(ii) outliers should not unduly influence research; ?

(ii) outliers should not unduly influence research; ?

(iii) subjects should be permitted to abstain from the valuation process. ?

(iii) subjects should be permitted to abstain from the valuation process. ?

10. (Implied by Randall, Chapter VIII): Subjects must view the CVM process as a meaningful opportunity to influence policy via their responses; ?

or

- (Arrow, Rosen and Freeman, 1979): ?
Subjects must view questions as
being sufficiently hypothetical so
as not to provide incentives for
strategic behavior.
11. (Bishop-Heberlein):
11. (Bishop-Heberlein):
- (i) Azjen-Fishbein criteria for the ?
structure of valuation questions
must be satisfied.
- (i) Azjen-Fishbein criteria for the ?
structure of valuation questions
must be satisfied.
- (ii) "close" correspondence between ?
attitudes and behavior is required.
- (ii) "close" correspondence between ?
attitudes and behavior is required.

It must be acknowledged that the rationale for including
any of the ROC's in Table 3, as well as the rationale for
excluding other possible ROC's, is weak or nonexistent at

this point in time. For example, our suggested ROC's 1 and 2 are justified by, first, the "familiarity" argument and secondly, our observation that in several comparison studies, $\pm 50\%$ accuracy ranges for CVM values overlap with $\pm 50\%$ ranges for indirect market methods for valuations of commodities which we assert are commodities with which subjects are probably familiar and have some degree of indirect market experience. Obviously, neither argument is immune to challenge. As a further example, in Section I.4 above we argue for the abandonment of the "information bias" rubric (ROC 8 (ii)). As a final example, we note that at this stage of the state of the arts, we are unable to even give precise definitions for many of the limits on CVM measures that we believe to be important; e.g., in 9(i), what is a "simple" choice?; in ROC 10, what is a "meaningful opportunity" or a "sufficiently hypothetical" choice?

Thus it is hoped that the combined discussions in this book concerning the potential role of ROC's in providing means by which ranges of Reference Accuracy may be attributed to CVM measures will provoke imaginative thinking and research relevant to the specification of precise and defensible ROC's; in any state of the arts assessment, of course, the immediately preceding discussions establish the infant stage of this process at this point in time.

As is obvious from Table 3, while we at least can see a place to begin in terms of specifying ROC's, our knowledge is virtually nil in terms of the error implications of not satisfying an ROC. Referring to ROC 8 in Table 3, Rowe and Chestnut's (1980) error estimates can be of very limited usefulness for our purposes given our inability to assess the quality of studies used in their samples vis-a-vis other relevant ROC's. Of course, this virtual void in our knowledge is the motivation for the insistence on "calibration" research by almost all of the participants (see, particularly, the Comments by Arrow and Rosen in Chapter XII, and those by Bishp-Heberlein (Chapter IX) and by V. Kerry Smith (Chapter XI)).

K. THE STATE OF THE ARTS OF THE CONTINGENT VALUATION METHOD

In Chapter I we noted the need for a "reflective pause" in CVM research wherein concerned researchers can take stock of the progress that has been made in the development of the method, and of the major issues which require resolution for further developments. The need for such a pause was made manifest by our review of the myriad "criticisms" of the CVM, all of which pointed to the disarray and confusion amongst CVM researchers attributable to two central facts. First, there has been a lack of consensus among researchers as to the priority issues and hypotheses that warrant empirical focus. Research efforts appeared scattered and diffuse as we repeatedly addressed asserted "biases" in the CVM (e.g., starting point, information, vehicle biases, etc.) in the "heuristic" manner described in Chapter III, with seemingly but one basis for accepting or rejecting a "bias": some ill-defined "preponderance of evidence." In large part, this lack of a well-defined, prioritized research agenda for the CVM reflects the ad hoc, "chemistry set" approach to CVM research noted by V. Kerry Smith, Bishop-Heberlein, and other Conference participants. Empirical applications of the CVM have outstripped intellectual inquiry -- via formal models or otherwise -- as to how individuals may behave within contingent market settings and implications for questionnaire design and implementation practices. Secondly, following perhaps from the preceding observations, CVM researchers have been apologetic, or defensive, vis-a-vis the "rest of the profession" due to the pervasive feeling that interrogated responses by individuals to hypothetical propositions must be, at best, inferior to "hard" market data or, at worst, off-the-cuff attitudinal indications which might also be expected to reflect efforts by individuals to manipulate the survey to their selfish ends.

The difficulties involved in efforts to provide some state of the arts context for the controversies surrounding the viability of the CVM for estimating values for public goods are made manifest by the assessment process seen in Parts I and II of Volume I.B. Thus, many of the positions and conclusions presented by us in Part I were later altered or retracted as a result of the focused dialogue concerning priority issues in CVM research between the authors, four other prominent CVM researchers, and leading economics and psychology scholars whose interest in public goods valuation is a step removed from the CVM per se. Of course, the reader will judge the success of this process in providing a

state of the arts context for the CVM. In this regard, our general view of this context is described as follows:

1. The CVM Without Apology. It is surely time for replacing apologies for the CVM with a positive research agenda to be described below. As a first step in this direction, we must eschew the joys of self-flagellation over our lack of knowledge of the "truth": we don't and won't know it, nor will our colleagues in the "rest of the profession" vis-a-vis their value estimates, nor will scientists in other disciplines. Following Arrow's exhortations, we must directly address the question, what is accuracy, and then look to calibration methods which provide us with a means to achieve accuracy levels that are reasonable and consistent with those levels obtained in other areas of economics and in other disciplines.

What is accuracy in a CVM estimate? It is a subject's valuation of a commodity which "reasonably" reveals his/her preference for the commodity. What does "reasonable" mean? "Reasonableness" is established by criteria -- Reference Operating Conditions -- which allow us to measure the magnitude of probable errors in any given application of the CVM. Thus, whether resulting ranges for Reference Accuracy associated with applications of the CVM are never better than $\pm 50\%$ or $\pm 500\%$, our focus is on defining the reference accuracy range. As with any other estimates, the "usefulness" of estimates with any range of error is determined by the purposes to which the estimates are to be put.

2. Conclusions concerning accuracy. While perhaps useful in pointing to needed research, the above is little more than a definition of accuracy. Given, as was argued above in Section J, that efforts to develop ROC's relevant to the CVM have just begun, and that we are almost totally ignorant of the error implications associated with the few ROC's that seem palatable at this time, must we then agree with V. Kerry Smith's judgement (Chapter XI) that no conclusions about the accuracy of CVM measures can be drawn based on research accomplished to date? We think not. At this point of reflective pause in the development of the CVM, one fails to see implications for the accuracy of CVM measures from received research only if one's view of "acceptable" implications is limited to evidence that demonstrates some degree of precision -- narrow ranges of error. This is to say that while we cannot build the case for ranges of Reference Accuracy for the CVM of magnitudes that would make CVM value estimates of practical use in many cases, at this point in the method's development a "useful conclusion" in the sense of V. Kerry Smith's assertion might well be that

the method produces order of magnitude estimates -- but we think one can argue that error ranges are much smaller.

Before continuing this argument, it is relevant for our purposes to recall V. Kerry Smith's demonstration (Chapter XI) of the wide range of potential for hypothetical and reporting errors in "hard" data commonly used, without apology, in economic analyses. Such data are seemingly accepted in total ignorance of ROC's relevant for their collection and the resulting ranges of Reference Accuracy. This observation, when combined with Coursey and Nyquist's findings of potential errors in ordinary demand analysis and Mitchell-Carson's general comments regarding sampling errors, should serve -- to paraphrase Freeman (Chapter X.E) -- as a chilling reminder of the limitations of empirical analysis/models in most areas of economic analysis. It seems fair to say that, in the general economics literature, questions of accuracy are not prominent. This is not to suggest a nihilistic approach to CVM research: the whole world is wrong (inaccurate), so why should we be concerned with accuracy. We mean to suggest the perspective: economists' typical preoccupation with such things as standard errors, etc., may have misled us into viewing value estimates as "precise" in terms of narrow error ranges, \pm 5%, 10% or even 20%. Couched in the broader terms of Reference Accuracy, such "precision" in general economic value estimates may quickly dissipate. Again, that such broader views of accuracy are generally ignored in economics is made manifest by V. Kerry Smith's provocative discussion in XI.B.

Returning to our discussion of what one can conclude regarding the accuracy of CVM measures, we begin by recalling an earlier discussion of the "truth". We do not and will not know it. But something analogous to "truth" may be attributed to values derived from, as examples, actual cash trades in Bishop-Heberlein's Sandhill study and in Vernon Smith's laboratory experiments. Eschewing arguments as to how Bishop-Heberlein's auction formats might have been improved in one way or another, their cash offers/payments are certainly the "truth" vis-a-vis preference revelation in the sense that folks clearly paid (were paid) for a well-defined commodity and then used the commodity. For the limited, most likely nonequilibrium, "simulated" market used by them, we can surely attribute preference revelations to these values. The differences between mean cash and CVM WTA values was roughly 42%; between cash and CVM WTP values, differences ranged from about 38% to 124% across their four auction formats (Tables 9.2 and 9.3 in Volume I.B). Do these differences imply nothing vis-a-vis conclusions as to the accuracy of CVM

measures? If accuracy is viewed as involving "small" ranges for Reference Accuracy, one would lament the "large" differences, as do Bishop and Heberlein, and concur with V. Kerry Smith that nothing (positive) can be concluded. If orders of magnitude are relevant, one might find Bishop-Heberlein's results startling: CVM and cash offers are virtually the same (see Figure 6.1 in Volume I.B). Our colleagues in environmental engineering may well envy such accuracy. In these regards, we note Bishop-Heberlein's later "surprise" (IX.F) at how well the CVM does work -- cash-CVM differences were not "outrageous".

Questions of the transferability of laboratory results to real-world conditions aside, hypothetical responses in Vernon Smith's experiments were consistently within 10% of actual market outcomes. In the Coursey et al. laboratory experiment, differences between values derived from final Vickery auctions and hypothetical questions were less than 20% for WTP and approximately 100% for WTA. The central point in all of this is apparent, however. In terms of the standard for comparisons of CVM values, we can continually argue as to how well preference revelations are made manifest by Bishop-Heberlein's cash offers, Vernon Smith's securities values, Coursey et al.'s measures related to tasting sucrose octa-acetate, or, moving to public goods, TCM and HPM values derived by the eight sets of authors given in Table 6.12 (Volume I.B). But however well any of these measures reflect meaningful revelations of preferences by individuals, every piece of evidence that we have demonstrates that the CVM yields value estimates that are indistinguishable from those standards in order of magnitude terms. Indeed, and herein lies the relevance of our $\pm 50\%$ arguments, in most instances CVM values are within $\pm 50\%$ of values derived from alternative methods for estimating preference revealed values.

3. Final Remarks. Thus, our final (c.f. our stronger, pre-Conference, reservations in Chapters I - VI of Volume I.B, ad passim) assessment of the state of the arts of the CVM is generally positive. We find impressive the accuracy of CVM measures inferred by the available evidence at this stage of the method's development. We find encouragement in the Conference results, particularly those reported by Arrow, Kahneman, and Bishop-Heberlein, which suggest that breaking the "hypothetical barrier" in the CVM may not be as hopeless as we and others earlier believed.

"Promise" is not "performance," however, and our assessments given above refer only to the potential promise of the CVM as a viable method for estimating values for public goods. The realization of that promise implies real challenges for theoretical and

empirical research for those involved with the method's further development. In concluding this Executive Summary, we now focus attention on critical issues for any research agenda which are relevant for guiding future CVM research.

L. CRITICAL ISSUES FOR FUTURE CVM RESEARCH

In the most general terms, it must be hoped that greater focus can be achieved in future research with the CVM. Both Bishop-Heberlein and V. Kerry Smith emphasize the ad hoc character of the bulk of CVM research to date -- a characterization aptly described by Bishop-Heberlein as reflecting a "chemistry set" approach. To a large extent, the ad hoc quality of CVM research has resulted from the emphasis or priority given empirical results -- necessitated in many cases by data needs of the entities providing research funding -- as opposed to theoretical and design issues. Results from this empirical emphasis are made manifest by the profession's preoccupation, without resolution, with such operational "biases" as starting point, information and vehicle issues as noted in Chapter III, issues some of which, upon reflection by Conference participants, may now be viewed as not implying biases per se but rather implying limits on questionnaire design and the manner in which CVM values are interpreted. Thus, the first critical issue for future CVM research is the metaphorical realignment of the empirical cart and the theoretical horse. There is a critical need for modeling efforts focused on individual behavior in contingent market settings which may serve as a basis for formulating hypotheses for empirical testing. This need for modeling efforts underlies virtually all of the additional issues for further CVM research discussed below.

A second critical issue for future research involves the specification and measurement of Reference Accuracy for CVM measures. In this regard, imaginative and innovative thought is required for defining relevant ROC's (e.g. Table 3 above) and for calibrating errors with deviations from ROC's. Thus, we must ask questions exemplified by: What is "familiarity" or "experience" vis-a-vis a CVM commodity; what is "uncertainty" and what constitutes "ideological content"; what variables may perform best as measures of cognition and/or affectation and how are attitudinal variables calibrated with measures of attitude-behavior correspondence; how can we better structure value questions so as to enhance a priori our expectations that preference revelations are obtained which are at least consonant with incentive-compatible revelations in market contexts? In addressing these issues we will need to profit from and exploit the lessons learned in laboratory and field experiments, as well as in research in other disciplines.

A final critical issue for future CVM research involves our need to resolve the "incentives" question. In this

regard, our concern extends beyond the hypothetical payment question. We concur with Arrow's suggestion that question settings that are sufficiently pseudo-real may be expected to result in satisfactorily pseudo-real responses and we are not convinced as to the extent to which one can distinguish between payment effects and those attributed to familiarity and experience questions. Of interest in these regards is the threads of an argument, seen implicitly in Randall's paper, as well as in Kahneman's Comments, and explicitly in Mitchell-Carson's paper (Appendix), that valuations of contingent changes in provision levels of public goods might be better obtained via processes which attempt to simulate results from institutions other than the market institution. Their examples specifically suggest the referendum institution. In terms of familiarity and experience, the provision of public goods via reliance on market-like transactions valuations is, at best, tenuous vis-a-vis the referendum process which is actually used in this regard. Some sort of preference revelation must surely be inferred by the act of an individual's signing a petition which requests a public/social action which the individual generally knows will result in his/her payment of higher taxes. Thus, a la Randall, the subject may indeed be motivated by the opportunity to influence policy. Whether such motivation would lead to strategic" signings of a cost-specific referendum is an important empirical question. Here we simply note the potential appeal for such a variation in CVM applications in dealing with many of the sources of familiarity/experience problems, when market analogies are used in the CVM and its possible use in resolving (or re-casting) the incentives problem.