



Winter 1994

Was the Ohio Court Well Informed in Its Assessment of the Accuracy of the Contingent Valuation Method

Ronald G. Cummings

Glenn W. Harrison

Recommended Citation

Ronald G. Cummings & Glenn W. Harrison, *Was the Ohio Court Well Informed in Its Assessment of the Accuracy of the Contingent Valuation Method*, 34 NAT. RES. J. 1 (1994).

Available at: <https://digitalrepository.unm.edu/nrj/vol34/iss1/1>

This Article is brought to you for free and open access by the Law Journals at UNM Digital Repository. It has been accepted for inclusion in Natural Resources Journal by an authorized editor of UNM Digital Repository. For more information, please contact disc@unm.edu.

Was the Ohio Court Well Informed in its Assessment of the Accuracy of the Contingent Valuation Method?

ABSTRACT

The District Court of Appeals for Washington D.C. has upheld a Department of the Interior ruling that accepts as a best available procedure for reliably calculating natural resources damages the Contingent Valuations Method (CVM). This paper addresses a number of questions that are relevant for assessing the reliability, or accuracy, of CVM values. Empirical evidence concerning the extent to which CVM subjects behave strategically, i.e., misrepresent their true values, is found to be mixed. While results from received empirical studies do not support unequivocal conclusions one way or another, their results suggest that CVM values may overestimate real economic commitments by a considerable margin. It is then argued that the courts cannot accept CVM values out of hand; they must carefully assess the demonstrated potential for such overestimates of value in their deliberations concerning the use of results from CVM surveys in CERCLA litigation.

INTRODUCTION

In the 1980 enactment of the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) and subsequent amendments¹, Congress instructed the Department of the Interior (DOI) to establish regulations that would identify the *best available procedures* with which Trustees of natural and environmental resources might assign monetary damages to any injury to these resources. Such damage estimates were to serve as a basis for the Trustees' efforts to recover compensation under provisions of the Act. The regulations established by DOI² were subsequently challenged by the State of Ohio and other intervenors in a Federal District Court.³ One important basis of the chal-

* Respectively, Noah Langdale Professor of Environmental Policy, Policy Research Center, Georgia State University, University Plaza, Atlanta, GA 30303, and Dewey H. Johnson Professor of Economics, Department of Economics, College of Business Administration, University of South Carolina, Columbia, SC 29208.

1. 42 U.S.C. Sec. 9651(c).

2. 51 Fed. Reg. 27673, 27720 (August 1, 1986).

3. State of Ohio v. U.S. Department of the Interior, 880 F.2d. at 474.

lenge was the DOI's adoption of the "lesser-of" rule: resource damages were to be measured by the lesser of restoration costs and use values, where use values were essentially equated with market values. In 1989, the *Ohio* court found the DOI's rule too restrictive given the perceived intent of Congress to provide compensative damages for natural resource injuries. In terms of a *statutory* definition of "use value," the court found that other values such as "option" or "existence values" (in general, "nonuse values"), were not to be excluded out of hand.⁴ The court instructed DOI to review their regulations to the end of providing a rule that would allow Trustees to develop values for resources damage "... by summing up all reliably calculated use values⁵, so long as the trustee does not double count."⁶ As a best available procedure for *reliably calculating* statutory use values, the court effectively put its imprimatur on a survey method known as the Contingent Valuation Method (CVM).⁷

The CVM involves the use of surveys describing the good or resource injury to be valued, a rule that relates financial payments to the provision of the good or the avoidance of the injury, and a question that asks the subject to report a maximum willingness to pay (WTP) some amount of money to see the good provided or the injury avoided. Both the "good" and the subject's payment are hypothetical. These features of the CVM, particularly hypothetical payment, are the source of concern over the extent to which values derived using this method actually reflect real economic commitments in any precise way.⁸ Those who assert that the CVM is subject to gross errors, resulting

4. *Id.* at 464. The notion of "nonuse values" refers to a valuation that a person might have for some natural resource even if he or she had no intention to ever use the resource. Although some controversy exists as to what components of "total value" can be attributed to "use value" and "nonuse value", our focus here is on the measurement of "total value." For an extensive discussion of the literature on "nonuse value", see R. Cummings & G. Harrison, *Existence Values and Compensable Damages: Judicial Reliance on Empty Economic Concepts?*, Economics Working Paper B-92-05, Department of Economics, College of Business Administration, University of South Carolina, 1992.

5. Note here that "use values" are of the *statutory* type, and therefore may include what economists define as nonuse values. The *Ohio* court appeared to take the eminently sensible view that *all* values were "use values" in the broad sense of being attributable to the use of some aspect of the resource (e.g., its mere existence). Economists employ the term "use value" more narrowly to refer to current or expected use of the physical resource itself (e.g., for recreation).

6. *Id.*

7. "We sustain DOI in its conclusion that CV methodology is a 'best available procedure.' As such, its conclusion in the Natural Resource Damage Assessment regulations was entirely proper." *Supra* note, at 98.

8. The strengths and weaknesses of the CVM are assessed in several studies. See, for example, R. Cummings, et. al, *Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Method*, Rowman & Allenheld (Totowa, NJ: 1986); R. Mitchell, & R. Carson, *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Johns Hopkins Press (Baltimore: 1989); and R. Cummings, & G. Harrison, *Identifying and Measuring Nonuse Values for Natural and Environmental Resources: A Critical Review of the State of the Art*, American Petroleum Institute (114 pp., Washington, DC: 1992).

in substantial divergences between CVM-estimated values and real economic commitments, sharply criticize the rationale for the *Ohio* court's implicit acceptance of the method as a best available procedure for *reliably* calculating natural resources damages.⁹ For example,

"... even if it were true that contingent valuation measured non-use damages, albeit with gross errors, there would be no principled basis for deliberately choosing to make important social policy on the basis of a method that is known to be grossly inaccurate. Voodoo, after all, is a method of forecasting the future that has at least as many adherents as contingent valuation; no one would suggest that economic forecasting should be based on it merely because other methods have their deficiencies. Lack of alternatives is not a reason for the knowing use of an inaccurate method for assessing damages; if it is the case that non-use damages cannot be measured to any reasonable degree of accuracy, then such damages should not be awarded."¹⁰

The *Ohio* court heard several claims as to the "voodoo" nature of the CVM. In rejecting these claims the *Ohio* court appears to have relied on established legal criteria for determining the admissibility of survey data.¹¹ These criteria are based on "necessity" and the "circumstantial guaranty of trustworthiness."¹² "Necessity" involves the:

"... comparison of the probative value of the survey with the evidence, if any, which as a practical matter could be used if the survey were excluded. If the survey is more valuable, then necessity exists for the survey, i.e. it is the inability to get 'evidence of the same value' which makes the hearsay statement [represented by the survey] necessary."¹³

9. We should note that damage estimates based on the CVM have been used in an actual litigation under CERCLA in only one instance thus far: *Idaho v. Southern Refrigerated Transport Inc.*, 1991 Westlaw 22479 (D. Idaho 1991). The court found that the CVM used by Idaho to estimate existence values was legally insufficient to establish such values in a credible way. It should be noted that this finding was not based on a rejection of the CVM *per se*, but on the "benefits transfer" way that the CVM was used. The study used by Idaho was a CVM study prepared by Battelle Northwest Laboratories for valuing a doubling in the runs of steelhead and salmon in the entire *Columbia River Basin*. Idaho used the results (\$16.97 per non-returning adult steelhead) as an existence value for 1,688 non-returning adult fish in the Salmon River, which is a very small part of the Columbia River's drainage area. Interestingly the court *accepted* results from a U.S. Forestry Service conducted study using the Travel Cost Method, another survey method, for the purpose of valuing recreational values that were at issue in this case.

10. J. Daum, *Some Legal and Regulatory Aspects of Contingent Valuation*, Section 9 in *Contingent Valuation: A Critical Assessment*, Cambridge Economics, Inc. (Cambridge, MA: 1992), at 30.

11. See, particularly, *Zippo Manufacturing Co. v. Rogers Imports, Inc.*, 216 F. Supp. 670, 683 (S.D.N.Y. 1963).

12. *Id.*, at 683.

13. *Supra* note 11, at 683.

Criteria related to the "guaranty of trustworthiness" of a survey include the four dangers of faulty memory, faulty perception, faulty narration and insincerity on the part of survey respondents.¹⁴

The Ohio court explicitly responded to criticism that the CVM can lead to *overestimates* of actual WTP, given that payment in the CVM is hypothetical.¹⁵ Specifically, it found that:

"The simple and obvious safeguard against overstatement, however, is more sophisticated questioning. Even as matters now stand, the risk of overestimation has not been shown to produce such egregious results as to justify judicial overruling of DOI's careful estimate of the caliber and worth of CV methodology."¹⁶

The motivation for this paper is the apparent lack of information available to the Ohio court regarding the risk that CVM values may substantially overestimate real economic commitments on the part of subjects. It is our view that the three studies considered by the court in this regard¹⁷ do *not* reflect the state-of-the-art understanding of the relationship between CVM values and real economic commitments. A comprehensive understanding of this relationship may well result in much less sanguine view than that adopted by the Ohio court regarding the significant potential for overestimation of value by the CVM. The purpose of our paper is to provide this understanding by critically reviewing the current state of the art in the economics literature.

One finds in the economics literature a concern with issues that parallel the court's concern with "dangers of perception" and "dangers of insincerity." The economics counterpart to these concepts are referred to as "preference research" and "strategic behavior," and are

14. See the Zippo case (*supra* note 11). For a set of more technical criteria for assessing the "trustworthiness" of a survey, see Toys "R" Us, Inc. v. Canarsie Kiddie Shop, 559 F. Supp. 1189, 1205 (E.D.N.Y. 1983), particularly at 1205.

15. As an aside, we note the Ohio court's finding (*supra* note 3 at 96) that "Certainly nothing in CV methodology itself shapes the injury inflicted by an environmental disaster." Related to the "dangers of faulty narration" criterion, however, we note that the CVM does indeed shape the injury inflicted by an environmental disaster that is valued in the application of the CVM. For studies that demonstrate the effects on CVM value associated with changes in the way that "injury" is described in the CVM questionnaire, see J. Bergstrom, et. al, *Information Effects in Contingent Markets*, American J. Agricultural Economics, 71, 687-691, August 1989, and J. Bergstrom, et. al, *The Impact of Information on Environmental Commodity Valuation Decisions*, American J. Agricultural Economics, 72, 614-621, August 1990.

16. *Supra* note 3 at 97.

17. The studies cited by the court were R. Bishop, & T. Heberlein, *Measuring Values of Extra Market Goods: Are Indirect Measures Biased?* American J. Agricultural Economics, 61, 926, 1979; E. Yang, et. al, *The Use of Economic Analysis in Valuing Natural Resource Damages: An Overview*, Unpublished Manuscript, Environmental Law Institute, 1983; and R. Bishop, et. al, *Does Contingent Valuation Work? Results of the Sandhill Experiment*, Unpublished Manuscript, Department of Agricultural Economics, Cornell University, August 1984.

discussed in section II. In section III we directly address the question: what do we know about the relationship between CVM values and values that reflect real economic commitments? The answer to this question is relevant for legal assessments as to the claim that CVM estimates might constitute "egregious results." Concluding remarks are offered in section IV.

STRATEGIC BEHAVIOR, PERCEPTIONS AND CVM VALUES

In this section we address two issues that relate to legal criteria for the admissibility of surveys as evidence in litigation for compensable damages. The first concerns the extent to which subjects in an application of CVM might have incentives to behave "strategically," which bears directly upon the "sincerity" of the subjects. The second involves the extent to which CVM subjects perceive and understand the valuation task asked of them, which relates to the dangers of faulty perception.

Strategic behavior *per se* is not undesirable in the CVM. Some types of strategic behavior, most importantly truth-telling¹⁸, are required if values offered by subjects are to be taken as representing real economic commitments. The strategic *bias* issue instead refers to strategic behavior whereby CVM subjects report values that do *not* represent real economic commitments. This type of behavior, suggested by Samuelson¹⁹ as being unavoidable in questionnaires that attempt to estimate values for public goods, as opposed to private goods,²⁰ involves subjects reporting values which they hope will influence the results of a survey in directions that they view as being favorable to them.

18. This is called "non-strategic behavior" in the CVM literature. This is potentially misleading, since one will always expect rational agents to be responding strategically to the game they are confronted with irrespective of whether or not they are telling the truth. Nonetheless, the older CVM literature uses the terms "truth-telling" and "non-strategic behavior" synonymously. We will similarly be using the notion of "sincere" behavior in the legal sense of truth-telling, even though it is perfectly possible for a rational agent to engage in misrepresentation from a sincere game-theoretic perspective. In all cases the context should make clear our meaning.

19. P. Samuelson, *The Pure Theory of Public Expenditure*, Review of Economics and Statistics, 36, 386-389, November, 1954.

20. V. Smith (*Arbitrary Values, Good Causes, and Premature Verdicts*, J. Environ. Economics and Management, 22, 71-89, January, 1992 at fn. 1) suggests that in a later paper Samuelson (P. Samuelson, *Aspects of Public Expenditure Theories*, Review of Economic and Statistics, 40, 332-338, 1958) was more optimistic concerning the prospects for eliciting people's values through questionnaires. Smith refers to Samuelson's statement following reference to the strategic bias issue: "... I do not wish to be too pessimistic. After all, the world's work does somehow get done. And to say that market mechanisms are non-optimal, and that there are difficulties with most political processes, does not imply that we can never find new mechanisms of a better sort [...] as an example regarding strategic bias] Interrogate people for their tastes with respect to public goods in such large homogeneous groups as to give each respondent the feeling that his answer can be a 'true' one without costing him anything extra." [Samuelson, cited above, at *supra* note 19 at 334] In his later considerations as to how one might devise a system of "benefit taxation" which

Strategic bias may reflect any number of motivations. Subjects may report a "value" for the purpose of making a statement concerning strongly held ideological views regarding the good in question. If a subject feels that a particular public good will be provided regardless of the level of payment determined in a survey, she may purposely "free-ride" on the contributions of others by reporting a very low value to minimize the amount she will ultimately pay for the good. Alternatively, she may believe that she will *not* ultimately pay for the public good, but that the decision as to whether or not the good will be provided depends upon the results of the survey. If so, she may be motivated to *over-value* the good in the hope of having the good provided (at no real cost to her). We consider the type of strategic behavior that has been of primary concern to most CVM researchers: free-riding.

One finds in the CVM literature the conventional wisdom that *it is well established* that subject behavior in the CVM will *not* in any substantive sense involve free-riding. For example:

"Compared with the experiments [from the experimental economics literature], which show a low level of strategic behavior, we conclude from the evidence that strategic behavior in properly designed CV surveys should be even lower. The evidence is sufficiently extensive and consistent, we feel, to shift the burden of proof to the shoulders of those who would continue to challenge the validity of all CV studies on grounds of strategic behavior. This judgement should not be taken to imply that the CV analyst can totally dismiss the possibility of bias caused by strategic behavior (. . .) some types of CV studies are more prone to strategic behavior than others."²¹

"However, several types of evidence tend to corroborate the reliability of CVM results. First, CVM results are consistent with preferences revealed by actual choice behavior . . . empirical evidence appears inconsistent with the conventional models of strategic misstatement."²²

". . . [T]he specter of Samuelson's strategic bias proposition remained as a concern . . . until the appearance of Vernon Smith's report of experimental evidence that further belied the strategic bias proposition."²³

would in some sense make people pay for what they get, he concludes "Instantly, you will discover that the same game-theory reasons that compel rational men to hide their desires for public goods will be motivating them to hide their consumers' surpluses from different product configurations." [Samuelson, cited above, *supra* note 19 at 334]

21. Mitchell & Carson, *supra* note 19 at 162.

22. J. Hoehn, & A. Randall, *A Satisfactory Benefit-Cost Indicator for Contingent Valuation*, J. Environ. Economics and Management, 14, 226-247, 1987 at 227.

23. Cummings, et. al, *supra* note 8 at 16.

"Studies by Bohm, Scherr and Babb, and Smith tend to indicate that fears among economists relating to gamesmanship are exaggerated."²⁴

"Empirical evidence thus far does not support the existence of strategic bias among consumers."²⁵

"... survey work with consumers has failed to show any evidence of strategic bias in valuing public goods. This result agrees with the experimental work of Grether and Plott and Smith, who also failed to find evidence of strategic economic behavior in experimental settings."²⁶

Statements of this type invite the reader to accept as "fact" that there is unequivocal evidence that supports one or more of the following statements:

1) the valuation behavior of subjects in the institutional setting used in experimental economics can be assumed to characterize behavior in the institutional setting of the CVM;

and/or

2) notwithstanding substantive differences between the valuation institutions used in experimental economics and that used in the CVM, subject behavior in the CVM is reasonably *similar* to that observed in the institutions used in experimental economics;

and/or

3) within the valuation institution of CVM, it has been demonstrated that subjects do not behave strategically.

Statements one and two presume that experimental studies conclusively show little free-riding. This issue is taken up as a digression below. We next review evidence for each of these statements.

Experimental institutions and the CVM

Can the valuation behavior of subjects in the institutional environment used in experimental economics be assumed to characterize behavior in the CVM institutional environment?²⁷ In experimental economics, two valuation institutions have primarily been used to research free-riding behavior: the Smith (or Unanimity) Auction and the Voluntary Contribution Mechanism.

24. Bishop & Heberlein, *supra* note 17, at 928.

25. W. Schulze, et. al, *Valuing Environmental Commodities: Some Recent Experiments*, Land Economics, 57, 151-172, May, 1981 at 156.

26. W. Schulze, et. al, *The Economic Benefits of Preserving Visibility in the National Parklands of the Southwest*, Nat. Res. J., 23, 149-173, January, 1983 at 153.

27. One might also interpret this statement as simply saying that institutions don't matter. Of course, this statement involves an issue that is basic to all research conducted in the field of experimental economics: *institutions do matter*. This is to say that individual choice behavior depends, in general, on the institutional environment within which choices

The Smith Auction was introduced and developed by V.L. Smith.²⁸ His are the works most often cited by CVM researchers as providing evidence that subjects generally do not free-ride. The salient institutional characteristics of the typical Smith Auction are:

(a.1) Subjects are given an interest-free loan at the outset of the experiment; this loan can only be used to buy shares of a public good.

(a.2) Subjects receive a monetary reward (a payoff) which depends on the number of units of the public good that is produced (how this number is determined is explained below). For example, they may each receive (say) \$10.00 if 6 units of the good is produced, \$12.00 if 7 units are produced, and so on. Thus, in essence, the subject does not buy or acquire a public good. His "shares" in the good are "sold" to the experimenter at the end of the experiment for money. This is his payoff.

(a.3) Subjects face "induced values". At the outset of the experiment they are told the costs of providing one, two, three, or any number of units of the public good. The number of units of the good to be "provided" is determined by the total amount of bids by all subjects. If total bids cover the cost of (say) 5 units, then 5 units of the public good are "provided." If total bids cover the costs of 10 units, 10 units are provided, and so on.²⁹ Subjects are aware of costs for "producing" different levels of the public good. Thus, in combination with (a.2), subjects know the value of producing different numbers of units of the public good. The value question faced by one subject is within the context of what *all other subjects* are bidding. Finally, this valuation context does not require subjects to "research their preferences" for the good in question. Their incentives are rather to research their *strategies* for maximizing their payoffs.

are to be made. We inquire below as to evidence that might support the position that people will actually pay amounts that they say that they will pay in the CVM. This inquiry may be viewed as setting aside questions regarding the relevance of economic theory for assessing the substance of CVM values.

28. V. Smith, *The Principle of Unanimity and Voluntary Consent in Social Choice*, *J. Political Economy*, 85, 1125-1139, 1977; "Incentive Compatible Experimental Processes for the Provision of Public Goods," in V. Smith (ed.), *Research in Experimental Economics* (Greenwich, CT: JAI Press, Vol. I, 1979(a)); *An Experimental Comparison of Three Public Good Decision Mechanisms*, *Scandinavian J. Economics*, 81, 198-215, 1979(b); *Experiments with a Decentralized Mechanism for Public Good Decisions*, *American Economic Review*, 584-599, 1980.

29. In the popular "balanced budget" version of the Unanimity Auction subjects receive rebates of monies collected which exceed the cost of providing the number of units of the good determined in the auction.

(a.4) When a subject offers a WTP for some quantity of the public good, he must *actually pay* his offered WTP out of the loan.

(a.5) Each subject can veto the proposed outcome, in which case all subjects receive no payoffs³⁰ and the public good is not "produced."

The Voluntary Contribution Mechanism has been used by a number of researchers for experiments concerned with the free-riding behavior of subjects.³¹ The salient institutional characteristics of the Voluntary Contribution Mechanism:

(b.1) Subjects are given an interest-free loan at the outset of the experiment; this loan can be used to contribute to the provision of the public good. The number of units of the public good that is "provided" is determined by dividing the total dollar amount of bids from all subjects by the assumed unit cost of providing the good. Thus, if bids from all subjects total \$500.00 and the unit cost of providing the good is \$50.00, 10 units of the good are "provided."

(b.2) Subjects receive a monetary reward (a payoff) which depends upon the number of units of the public good that is produced. It may also depend on the number of subjects participating.

(b.3) Subjects face "induced values" (see a.3 above).

(b.4) When a subject offers a WTP for some quantity of the public good, he must *actually pay* his offered WTP out of the loan.

Given the foregoing descriptions, one must ask if the valuation institution of CVM is *in any substantive way* similar to those used in these experimental institutions. In CVM, subjects are neither given a loan as in (a.1) or (b.1), nor do they actually pay their stated WTP as in (a.4) or (b.4). CVM subjects have no veto power concerning the outcome of the experiment as in (a.5), they do not receive monetary rewards determined by outcomes of the experiment as in (a.2) or (b.2), nor are they typically given cost information as in (a.3) or (b.1).³² Their

30. Apart from a fixed payment of about \$3.00 or so for showing up at the experimental session.

31. Examples include R. Isaac, et. al, *Public Goods Provision in an Experimental Environment*, *J. Public Economics*, 26, 51-74, 1985; G. Harrison, et. al, *An Experimental Evaluation of Weakest-Link/Best Shot Models of Public Goods*, *J. Political Economy*, 97, 201-225, 1989; G. Marwell & R. Ames, *Experiments on the Provision of Public Goods—I: Resources, Interest, Group Size, and the Free-Rider Problem*, *Amer. J. Sociology*, 84, 1335-1360, 1979; G. Marwell & R. Ames, *Experiments on the Provision of Public Goods—II: Provision Points, Stakes, Experience, and the Free-Rider Problem*, 1 *Amer. J. Sociology*, 85, 926-937, 1980; F. Schneider & W. Pommerehne, "On the Rationality of Free Riding: An Experiment," *Quarterly J. Economics*, 96, 689-704, 1981.

32. One may wish to argue, however, that a cost is implied when the dichotomous choice approach is used in the CVM.

values are not couched in terms of the behavior of all other subjects as in (a.3) or (b.1)³³, nor do they face induced values as in (a.3) or (b.3). But most importantly, one *must assume* that CVM subjects are motivated to research their valuation of the public good, whereas subjects in both the Smith Auction and Voluntary Contribution Mechanism have explicit incentives to research strategies that will maximize their payoffs. These observations lead us to conclude that very few *similarities* exist between the valuation institutions used (thus far) in experimental economics and those used in CVM.

The similarity of behavior in CVM and experimental institutions

Notwithstanding the substantive differences between the valuation institutions used in most experiments and that used in CVM, is there evidence suggesting that subject *behavior* in CVM is nonetheless reasonably similar to *behavior* observed in the Smith Auction or the Voluntary Contribution Mechanism?

Unfortunately, there are very few studies that provide empirical results that are pertinent to this question. There are no studies that compare free-riding behavior in CVM and Voluntary Contribution Mechanism institutions. Two studies compare values for a public good derived with the CVM and *variations* of the Smith Auction: Brookshire and Coursey³⁴ and Brookshire, Coursey and Schulze.³⁵

Brookshire and Coursey examined three elicitation institutions: a field CVM, a Field Smith Auction (SAF), and a Laboratory Smith Auction (SAL). In each case they elicited willingness to accept (WTA) and WTP valuations for the public good. The public good was "tree density" in a residential neighborhood in Fort Collins, Colorado. In the WTP exercises, they asked subjects to value increments of 25 and 50 trees from a baseline of 200 trees in a nearby park. In the WTA exercises, they asked subjects to value decrements of 25 and 50 trees from a baseline of 200 trees.³⁶ Thus, their overall experimental design consisted of three value elicitation institutions (CVM, SAF, and SAL), two

33. These contrasts of the CVM with the Smith Auction are nicely set out in the following way. "The CVM procedure outlined previously does not ask the individual to consider the valuation question in the context of what other individuals are bidding. Additionally, the CVM procedure does not present information to the household pertaining to the cost of the alternative (levels at which the public good is provided). Both of these elements are included by necessity in a Smith auction procedure." D. Brookshire & D. Coursey, *Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures*, *American Economic Review*, 77(4), 554-566, Sept., 1987, at 559.

34. *Id.*

35. D. Brookshire, et. al, *Experiments in the Solicitation of Private and Public Values*, in L. Green & J. Kagel (ed.s), *Advances in Behavioral Economics*, Ablex (Norwood, NJ), 1990).

36. Note that the commodities being valued in the WTA and WTP exercises are not the same. One values a decrement from 200, the other an increment from 200. It therefore does not follow from economic theory that WTA should approximate WTP.

valuations bases (WTP and WTA), and two levels of change (25 and 50 trees).

Brookshire and Coursey's analyses focus on WTP and WTA disparities. Although the free-riding question was not central to their inquiry, their raw data allows for an assessment of this question.³⁷ Specifically, we are interested in testing the hypothesis that there is no systematic difference between the CVM values they elicited and the SAL values they elicited. Since the latter values employed financial incentives, they more likely reflect real economic commitments.³⁸ Because the underlying samples exhibit significant skewness and kurtosis, we employ a non-parametric Kolmogorov-Smirnov test procedure to test this hypothesis. Using the data on "final bids" in the SAL institution we derived critical probability values of 0.17 (0.3) for the CVM-SAL WTP comparison and 25 (50) tree increment, and 0.00004 (0.00005) for the CVM-SAL WTA comparison and 25 (50) tree increment. Thus in two of the four possible comparisons these probability values suggest that the CVM and SAL institutions generate *different* valuations.

However, it is not appropriate to use the "final bids" in the SAL experiments, since these did not in fact represent real economic commitments. In the SAL, subjects can always veto the proposed allocation of payments, and indeed they did so in six of the eight SAL experiments.³⁹ This means that the tentative valuations used by Brookshire and Coursey for the final round of these experiments were not what the subjects ended up facing. Instead, they paid zero, or were compensated zero, as per the "rules of the game" for the Smith Auction. Brookshire and Coursey used valuations that the subjects entered in the last round, whether or not they met the group fund requirement or were vetoed. Therefore, the real economic commitment of the subjects in those cases was zero. If one substitutes a zero valuation for all of the SAL experiments that failed to converge,⁴⁰ the tests reported above result in a clear rejection of the null hypothesis in *all* four cases at probability levels of 0.0002 or less. We therefore conclude that these field experiments provide clear evidence that CVM values are significantly

37. We are grateful to David Brookshire and Don Coursey for making the raw data available to us.

38. The actual amount that was contributed by the experimenters (\$1500, as per *supra* note 33, 560) does not seem to bear any relation to the amounts of money that subjects actually paid or said they were willing to accept, as discussed below. It is therefore an open issue whether the subjects viewed the SAL institution as being incredible, in the sense of eliciting values which would not actually have any bearing on the amount of money contributed to the field public good.

39. See *supra* note 33, Table 2, 562.

40. An alternative way to use these data would be to retain only the responses for the SAL experiments that converged to some positive public good supply. However, this approach rejects the observed fact that many subjects had a zero WTP for the public good, as elicited using this institution. We prefer to include all valuation observations.

different from Smith auction values elicited with some financial commitment involved.⁴¹

Brookshire, Coursey and Schulze⁴² examined WTA and WTP for an unfamiliar public good: having to taste a foul liquid substance for 20 seconds. If subjects did not contribute a certain amount to a group fund in the WTP experiments, they all would have to taste this substance. In the WTA experiments, the subjects were asked to say whether they would accept an amount that was less than a certain amount in a group fund. If so, then they would all have to taste the substance.

The data in this study are very difficult to assess because it is in the form of a graph with no information about standard deviations. The hypothetical WTP CVM values (in Part I of their experiment) appear to be about 50 percent higher than their "Smith Auction" counterparts.⁴³ In the case of the WTA valuations, there appears to be a more dramatic difference, with the CVM values being about 100% higher than the "Smith Auction" values. Of course, such "eyeball" impressions have little, if any, weight, but we can do no better without access to the raw data.⁴⁴

In summary, we find no evidence to justify the claim that CVM values and values elicited with a Smith Auction experiment are similar.

Free riding in the CVM institution

Within the valuation institution of the CVM, does evidence exist that demonstrates that subjects do not behave strategically? Four studies are most often cited as providing such evidence: Bohm [1972],⁴⁵ Bohm [1984],⁴⁶ Rowe, d'Arge and Brookshire⁴⁷ and Brookshire, Ives and Schulze.⁴⁸

41. It turns out that using the actual values generated by the subjects in the SAL experiments results in a strengthening of the conclusion of this study (*supra* note 33, at 565) that providing market-like financial incentives results in a *reduction* of loss aversion. In fact it *removes* it altogether! A Kolmogorov-Smirnov test of the hypothesis that the WTA and WTP values for the 25 (50) tree version of the SAL experiments are the same has a probability value of 0.57 (0.31).

42. *Supra* note 35, Figure 2 at 185.

43. The institution here was actually a modification of the Smith Auction introduced in D. Coursey & V. Smith, *Experimental Tests of an Allocation Mechanism for Private, Public or Externality Goods*, Scandinavian J. Economics, 86, 468-484, 1984.

44. We understand from David Brookshire that these data have been accidentally lost.

45. P. Bohm, *Estimating the Demand for Public Goods: An Experiment*, European Econ. Review, 3, 111-130, June, 1972.

46. P. Bohm, *Revealing Demand for an Actual Public Good*, J. Public Economics, 24, 135-151, 1984.

47. R. Rowe, et. al, "An Experiment on the Economic Value of Visibility," J. Environmental Economics and Management, 7, 1-19, 1980.

48. D. Brookshire, et. al, *The Valuation of Aesthetic Preferences*, J. Environ. Economics and Management, 3, 325-346, 1976.

Bohm [1972] is a landmark study that had a great impact on many researchers in the areas of field public good valuation and experimentation on the extent of free-riding. The commodity used in Bohm's experiments was a closed-circuit broadcast of a new Swedish TV program. Six elicitation procedures were used. In each case except one, the TV program was made available and subjects in each group allowed to see it, if aggregate WTP equaled or exceeded a known total cost. Every subject received 50 Swedish Kroner (SEK) when arriving at the experiment.

Bohm employed five basic procedures for valuing his commodity. In Procedure I, the subject paid according to his stated WTP. In Procedure II, the subject paid some fraction (< 1) of stated WTP, with the fraction determined equally for all in the group such that total costs are just covered. In Procedure III, subjects did not know the specific payment scheme at the time of their bid, but did know that it was a lottery with equal probability attached to the payment schemes of Procedures I, II, IV and V. In Procedure IV, each subject paid a fixed amount (SEK 5). In Procedure V, the subject paid nothing.

For comparison, Procedure VI was introduced in two stages. The first stage, denoted VI:1, approximates a CVM since nothing was said to the subject about actually being offered the opportunity to watch the program (i.e., it was purely hypothetical). The second stage, VI:2, involved subjects bidding for the right to see the program against what they thought was a group of 100. This was a discriminative auction, with the 10 highest bidders actually paying their bid and being able to see the program.

No formal theory was provided to generate free-riding hypotheses for these procedures. Procedure I was deemed the most likely to generate strategic *under-bidding*, and procedure V the most likely to generate strategic *over-bidding*.⁴⁹ The other procedures, with the exception of VI, lay somewhere between these two extremes. Note also that explicit admonitions *against* strategic bidding were given to subjects in procedures I, II, IV and V.⁵⁰ Although no theory is provided for VI:2, it can be recognized as a multiple-unit discriminative auction in which subjects have independent and private values. It is well-known that optimal bids for risk-neutral agents⁵¹ can be well *below* the true

49. *Supra* note 45 at 113. All of the statements about predicted directions of over-bidding or under-bidding should be understood as *weakly* predicting the direction of bias. Thus all of the conclusions discussed below must be similarly qualified.

50. See, *supra* note 45 at 119, 127-29.

51. Bohm, *supra* note 45 at 126, discusses the possibility of what he terms "auction fever", in which subjects get caught up in the fight over a few objects (e.g., sports trophies). The upshot would be a form of the "winner's curse", in which the subjects end up paying more for the object than they, themselves, would have liked to pay *ex post*. If this type of irrationality applied to these subjects then their bids might exceed their true valuation, offsetting the logic presented in the text. We interpret the available data on the presumption that agents are rational. A game-theoretic model of behavior

valuation of the agent in a Nash Equilibrium, and will never exceed the true valuation. Unfortunately, there is insufficient information to determine how far below true valuations these optimal bids will be, since we do not know the conjectured range of valuations. Thus, we have relatively clear *a priori* predictions of the expected bias of observed bids in relation to true WTP for Procedures I, V and VI:2.

The major result cited from Bohm's study was that bids were virtually identical for all institutions, averaging between SEK 7.29 and SEK 10.33. Unfortunately, these conclusions are based on parametric test procedures, which are unreliable given the non-Gaussian nature of the samples.⁵² The mean contributions for Procedures I-V, VI:1 and VI:2, respectively, were 7.6, 8.8, 7.3, 7.7, 8.4, 10.2, and 10.3 (all in SEK). The respective medians, 5, 8, 5, 5, 7, 10 and 10, in these cases suggest an even larger disparity between the CVM Procedure VI:1 and the first five procedures. Using a non-parametric Kolmogorov-Smirnov test procedure, we derived critical probabilities that Procedure VI:1 elicited the same values as Procedures I-V and VI:2 of 0.018, 0.24, 0.011, 0.058, 0.13 and 0.99, respectively. Procedures I-V were not significantly different than any other in this group, whereas the values elicited in Procedures VI:1 and VI:2 were virtually identical. Therefore, the test results reported above for Procedure VI:1 *vis-a-vis* Procedures I-V are the same in this case as well.

Given our re-analysis of Bohm's data, what can one conclude as to the extent of free-riding, in hypothetical settings, from his experiments? Procedure VI:1 is the only hypothetical institution studied in these experiments. Comparing Procedure VI:1 and VI:2, we should see WTP values in the former that are larger than the latter, if VI:1 elicited the true WTP and if VI:2 values understated the true WTP. However, our analysis shows that the observed WTPs in each sample are *identical*, suggesting that the observed WTP in Procedure VI:1 *understates* true WTP. This presumes, of course, that WTP elicited in Procedure VI:2 reliably represents a rational bid below true WTP, which presumption is weakened by evidence suggesting that subjects may need substantial experience in auctions of this kind before they respond rationally to their incentives. In any event, we find no support here for the view that CVM values are the same as real economic commitments.

At conventional significance levels we cannot reject the hypothesis that the WTP elicited in Procedures V and VI:1 are different. Since Procedure V was predicted to generate bids that *exceeded* true WTP, this suggests that CVM values *overstated* true WTP. Note that this conclusion differs qualitatively from the conclusion drawn above from

in auctions of this type is presented in J. Cox, V. Smith & J. Walker, *Theory and Behavior of Multiple Unit Discriminative Auctions*, J. Finance, 39, 983-1010, Sept. 1984.

52. We are grateful to Peter Bohm for making the raw data available.

a comparison of Procedures VI:1 and VI:2. However, the subjects in Procedure V were asked in the instructions not to bid strategically, whereas the subjects in Procedures VI:1 and VI:2 were not so instructed. If the subjects actually followed their respective instructions then our conclusion, that CVM values overstated true WTP, breaks down. Unfortunately, there is no way to know how carefully the subjects actually followed their instructions.

Finally, we can reject the hypothesis that WTP in Procedures I and VI:1 are the same. Since Procedure I is predicted to generate bids below true WTP, this result is consistent with CVM generating *true* WTP values.

We conclude that there is no basis to claim dogmatically that Bohm [1972] showed that strategic behavior is absent in hypothetical settings, let alone that his results can be so interpreted concerning CVM. Depending on which pair of experiments one focusses on, we show that one can argue with equal validity that CVM values may understate, overstate, or equal true WTP. Moreover, our conclusions rest on the validity of the priors discussed above as to the "expected biases" in different procedures. Bohm concluded that his results were compatible with the view "... that people respond in an 'irresponsible fashion' ... to hypothetical questions."⁵³ We do not see how such a strong conclusion is warranted, however. These results are important for suggesting a methodology for attacking this problem, but it is premature to draw too strong a conclusion in this respect.⁵⁴

Bohm [1984]⁵⁵ uses two procedures that elicit a real economic commitment from individuals, albeit under different (asserted) incentives for free-riding. The two procedures were used to extract a lower bound and an upper bound to the true average WTP for an actual good. Each agent in group 1 stated his individual WTP. His actual cost was a percentage of that stated WTP, such that costs for producing the good would be covered exactly. This percentage could not exceed 100 per-

53. *Supra* note 45 at 125.

54. These results are used by Mitchell and Carson (*Supra* note 8 at 147) in an effort to generate some numbers on the "percentage of true WTP measured in experimental studies." They use the results from procedure VI:2 as a benchmark, arguing that they come closest to being true WTP since a real economic commitment was required (although this was also the case for procedures I-IV). Of course, as noted above and disregarding the "auction fever" hypothesis of Bohm (*Supra* note 45 at 126) the institution used in this case would lead us to expect these observed bids to understate true valuations, but by how much we cannot easily say. Thus, using the reported data for VI:2 as "true WTP" results in an upward bias in the percentages Mitchell and Carson report (*Supra* note 8 at 147). Further, they compare the average contributions in each procedure to the average for VI:2, resulting in numbers on the propensity to free-ride of 74 percent, 85 percent, 71 percent, 74 percent and 85 percent for procedures I-V, respectively. The raw data are not particularly symmetric, however, and medians tend to be much lower than means in all of these cases. If one uses the ratio of medians instead of means these propensities drop to 50 percent, 70 percent, 50 percent, 65 percent, and 70 percent, respectively. These are also inflated values since the values for VI:2 are biased down from their true values.

cent. Subjects in group 2 also stated their WTP. If the interval estimated for total stated WTP equalled or exceeded the (known) total cost, the good was to be provided and subjects in group 2 would pay only SEK 500. Subjects bidding zero in group 1, or below SEK 500 in group 2, would be excluded from enjoying the good.

In group 1, a subject only had an incentive to understate if he conjectured that the sum of the contributions of others in his group was greater than or equal to the total cost minus his true valuation for the good. Total cost was known by all subjects to be SEK 200,000, but each subject had to speculate as to the contributions of the others in his group. Unfortunately, it is not possible to say what the extent of free-riding was in this case without further information as to expectations that were not observed. In group 2, only those subjects who actually stated a WTP greater than or equal to SEK 500 had an incentive to free-ride. Forty-nine subjects reported *exactly* SEK 500 in group 2, whereas 93 reported a WTP of SEK 500 *or higher*. Thus the extent of free-riding in group 2 could be anywhere from 0 percent (if those reporting SEK 500 indeed had a true WTP of exactly that amount) to 53 percent (49 free-riders out of 93 possible free-riders).

The main result reported by Bohm [1984] is that the average WTP interval between the two groups was quite small. Group 1 had an average WTP of SEK 827 and group 2 an average WTP of SEK 889, for an interval that is only 7.5 percent of the smaller average WTP of group 1. Thus, one can conclude that, if free-riding incentives were present in this experiment, they did not materially affect the outcome.⁵⁵ One can question, however, the extent to which these results can be generalized. The subjects were representatives of local governments, and it was announced that all reported WTP values would be published. This is not a feature of most CVM studies, which often go to great lengths to ensure confidentiality. Thus, while the Bohm [1984] experiment provides evidence that free-riding behavior was absent where WTP involved a real economic commitment, the results do not transfer without qualification to the CVM.

55. *Supra* note 46.

56. The following calculation suggests this conclusion. The average WTP of the 50 percent of the population of 274 subjects who had an incentive to understate their WTP (group 1) was 7.5 percent below the average WTP of the remaining 50 percent who had an incentive to overstate their WTP (group 2). Adjusting for sampling error with a 95 percent one-sided confidence interval, the average WTP of the whole population, if placed in group 1, would be at most 32 percent below the average WTP of the whole population if placed in group 2. Although this 32 percent represents the combined effect of the understatement incentive in group 1 and the overstatement incentive in group 2, the free-rider (or understatement) incentive could still account for (at most) this 32 percent, assuming that nobody responded to the overstatement incentive. We are grateful to Peter Bohm for this interpretation of his data.

Brookshire, Ives and Schulze⁵⁷ (hereafter, BIS) undertake a series of CVM exercises designed, in part, to test for the importance of strategic behavior. They offer an explicit model of the chain of logic that a subject *could conceivably* go through when deciding whether or not to respond truthfully, and then propose to test for it indirectly. They posit:

"If respondents in our example assume the mean bid of the sample survey is used to set the entrance fee and an individual with perfect information has an 'honest bid' greater than the mean bid, there exists an incentive to make a false high bid to bias the sample mean upward." (emphasis added).⁵⁸

These are crucial assumptions that are not obviously applicable to any CVM, unless they are built in by design, as in Rowe, d'Arge and Brookshire.⁵⁹ Given these assumptions, BIS

*"... undertake to demonstrate that strategic behavior, when carefully analyzed in the context of our particular bidding game, can be easily recognized. This, in turn, will enable us to make an evaluation of the extent of the problem."*⁶⁰

The specific bidding game proposed by BIS assumed that each individual had perfect information with respect to the average bid of all others in the CVM.⁶¹ Two further implicit assumptions were (i) that CVM responses actually had some influence on the social decision, and (ii) that none of the other individuals were able to revise their bids after the subject in question gives his bid. In effect, this last assumption means that all subjects were expected to rationally conjecture what the mean bid of the other respondents was, since the assumption, as stated, could only be generally true for the last respondent.⁶²

Thus far in the argument one can appreciate the rationale for the assumptions, if one is to be able to say anything about strategic behavior without a reference valuation institution (such as a Vickrey auction). One can even imagine setting up a CVM in which these assumptions are met. However, BIS added a final assumption for which a rationale was not developed: that true valuations are distributed normally. Specifically:

57. *Supra* note 48.

58. *Supra* note 48 at 327.

59. *Supra* note 47.

60. *Supra* note 48 at 327, 328.

61. *Supra* note 48 at 328.

62. BIS have some discussion (*Supra* Note 48, at 328) about preferences being identical across all individuals, but this is simply an analytical vehicle to work in the assumption that the respondent in question knows the mean bid of everybody else (since he knows his own bid). Identical preferences is surely less palatable than making the direct assumption that the respondent conjectures the mean bid of everybody else, since this allows some heterogeneity of preferences from the outset.

"If we assume that honest bids are distributed normally, bias of the type discussed above will tend to 'flatten' the distribution dramatically. This implies substantial numbers of very high and very low bids relative to the mean bid if dishonest responses constitute a significant fraction of the survey."⁶³

Turning to the actual survey, the respondents were not given the sample mean. Indeed, the WTP question was posed with no qualifying information as to the bids of others.⁶⁴ Hence the authors relied on the respondents behaving as if they correctly conjectured the sample mean.

Evaluating the data, they conclude:

"Examining the distribution of bids [. . .], we note that it is not flat in either the total grouping or the disaggregated cases."⁶⁵

Unfortunately, this conclusion was not based on statistics that might suggest negative kurtosis,⁶⁶ which is the prediction of their strategic behavior hypothesis. One cannot generally rely on "eyeball assessments" of sample data distributions in this manner to draw reliable statistical conclusions, hence the BIS conclusion is not well supported. Thus, while an imaginative effort to investigate the possibility of strategic behavior in CVM, the BIS "distributional model" does not provide a basis for unequivocal conclusions regarding such behavior.

Rowe, d'Arge and Brookshire⁶⁷ (hereafter, RDB), following the suggestion of BIS,⁶⁸ actually gave the sample mean information to respondents to see if they wanted to revise their bid. A subject who revised his bid was presumed to be acting strategically. RDB's approach, unlike that of BIS, is a plausible means by which one might meaningfully extend the assumptions of BIS by "hard-wiring" them into the CVM design. Indeed, RDB correctly recognized the problems with the simple BIS model of strategic behavior:

". . . for the individual to accomplish this goal [of strategic bidding], a great deal of information is necessary. For instance, the sample size, the previous bids, and whether or not they are the last respondent are necessary to bid strategically with an assured outcome."⁶⁹

63. *Id.* at 329.

64. *Id.* at 334

65. *Supra* Note 48 at 340.

66. It is not clear whether or not the data reported in Figure 7 (*Supra* note 48 at 341) of BIS reflects all of the raw data. Rowe, d'Arge and Brookshire [*Supra* note 47 at 15], in discussing their "trimming" procedures, imply that the same procedures were used in BIS. If this is the case, such trimming would bias the results against the proposed hypothesis of strategic bidding.

67. *Supra* note 47.

68. *Supra* note 48 at 345.

69. *Supra* note 47 at 6.

To overcome these informational problems, RDB gave 40 subjects the mean bid of the other (supposed) bidders and then told them that each would ultimately pay the *average* bid, and not their own. However, RDB did not specify whether they gave these subjects information on the full sample size, or whether or not a given subject was the last respondent. Such information is important. For example, if the subject is just one respondent in an arbitrarily large group, then he has little hope of shifting the mean, no matter what he bids.⁷⁰ Similarly, if the subject is included among the first few respondents, and there are many more to follow, he can scarcely affect the average one way or the other.

Turning to the data, we are concerned with the use of the automatic "trimming" procedure used by RDB to eliminate data which could support the hypothesis of strategic bidding:

"First, as in all results reported in this paper, zero bids were analyzed, under criteria suggested by Randall *et al.* and Brookshire *et al.*, where bids greater than 10 SD from the mean were deleted."⁷¹

This procedure could have significantly affected the results of RDB. Although RDB do not report the results with or without the use of the procedure, their conclusion hints that it did make a difference:

"... if zero and large bids are closely analyzed and possibly rejected, strategic bias, if it exists, has a negative effect upon the bid distribution."⁷²

Furthermore, the actual analysis of these data was complicated by the presence of many different treatments in the one design, making it difficult to test directly the hypothesized negative kurtosis of the bid distribution with strategic bidding. Instead, they performed a regression analysis to test for the significance of any interaction of a dummy representing "being environmentally inclined" and a dummy for "provision of the sample mean." They found no significant interaction for either dummy variable, and concluded that significant strategic behavior was absent. However, it is not clear why this interaction term should have a non-zero expected value under the hypothesis of strategic behavior. If, for example, there were as many pro-environmental as anti-environmental respondents in this group of 40, their bid deviations could be offsetting.

The RDB study was an imaginative effort to extend the earlier inquiry of BIS. Like the BIS study, however, it explores the possible use

70. Implicitly this line of argument assumes that the agent feels constrained not to give "incredibly large" bids, perhaps because he rationally fears that such "outliers" will be dropped.

71. *Supra* note 47 at 15.

72. *Supra* note 47, at 15.

of a method for identifying strategic behavior that has no apparent theoretical or empirical rationale. Therefore, one cannot appeal to the results from either study as having established the presence or absence of strategic behavior in CVM.

Free-riding in the experimental literature: A digression

Given that the CVM valuation institution is substantively dissimilar to those used in experimental economics, and that there is no firm basis for arguing that behavior in these two sets of institutions is similar, we have argued that one cannot readily justify drawing inferences from findings in experimental economics to subject behavior in CVM. Regardless of their direct relevance to CVM, the question remains as to what one might conclude from studies of free-riding behavior conducted in experimental economics. In this regard we are interested in the question: are findings of an absence of pervasive free-riding in experimental studies as monolithic as suggested by citations of the type exemplified above?

In terms of the free-riding behavior of subjects, applications of the Smith Auction are mixed and clearly do not support the general conclusion that free-riding is absent. Two factors in particular render such a conclusion false. First, the fact that the collective decision tends to be the efficient one when there is agreement does not mean that each individual has truthfully revealed his preferences, which is what incentive-compatibility or "demand revelation" require. As V.L. Smith points out very clearly:

"... the mean bids differ from the corresponding Lindahl equilibrium bids. Consequently, although the Auction Mechanism provides public good quantities that approximate the Lindahl equilibrium quantity the private good allocations do not approximate the Lindahl equilibrium quantities. [This] is because subjects with low endowment [...] tend to contribute less, while subjects with high endowment [...] contribute more, than is required for a Lindahl allocation."⁷³

These results are quite general to the many other induced-value experiments conducted with the Smith Auction.⁷⁴ What they mean is that some individuals over-contribute and others under-contribute, and that they do not do so at random.

Second, the success rate of the Smith Auction is not high. When the group fails to come to an agreement in the induced-value control experiments, at least one subject has not revealed his preferences truth-

73. Smith [1979(b)], *supra* note 28 at 208.

74. See, for example, Banks, et al., "An Experimental Analysis of Unanimity in Public Goods Provision Mechanisms," *Rev. Economic Studies*, 55, 301-322, 1988, at 314.

fully. Smith [1979(a)] found a failure rate of about 10 percent,⁷⁵ Smith [1979(b)] a failure rate of 20 percent,⁷⁶ and Banks, Plott and Porter⁷⁷ a failure rate of 50 percent. When one allows for these failures the efficiency of the Smith Auction is statistically about the same as a direct contribution mechanism for which free-riding is predicted and confirmed.

The clearest example of this last result was provided in Smith [1979(b)].⁷⁸ The average contribution for 10 experiments was 9.10 units with the Smith Auction. The average contribution was 7.3 units using a different mechanism for which free-riding is predicted was used. However, the average reported for the Smith Auction *excludes* those experiments which failed to reach agreement. The study points out that the average for the Smith Auction drops from 9.10 to 7.9 if the disagreement outcomes are included and given a value of 3.33 (which is the theoretical free-riding prediction). If one includes disagreement outcomes and assigns them their actual value, zero, one obtains an unconditional average provision level of only 6.3 for the Smith Auction, which is *below* the average provision level of the free-rider procedure (7.3). On the other hand, Banks, Plott and Porter⁷⁹ report significantly higher (unconditional) provision levels with the Smith Auction than with a free-rider mechanism. The appropriate conclusion is that the efficiency of the Smith Auction is sensitive to the specific environment in which it is used. Strategic behavior is observed in some, but not all, environments.

The implications of results from applications of the Voluntary Contribution Mechanism for free-riding behavior are also inconclusive.

There exists a set of studies using the Voluntary Contribution Mechanism that report results showing very little free-riding behavior on the part of subjects. Most notable among these studies are Chamberlin,⁸⁰ Marwell and Ames⁸¹, and Schneider and Pommerehne.⁸² For example, Schneider and Pommerehne [1981; p. 702] conclude that "[. . .] individuals did systematically behave as free riders [. . .] but the extent to which free riding occurred was not great."⁸³ Since these studies are given detailed attention in Mitchell and Carson, we will not repeat that detail here.⁸⁴

75. *Supra* note 28.

76. *Id.*

77. *Supra* note 74.

78. *Supra* note 28, Table 5 at 207.

79. *Supra* note 74, Table I at 316.

80. J. Chamberlin, "Provision of Collective Goods as a Function of Group Size," *Amer. Political Science Review*, 68, 707-716, 1974.

81. *Supra* note 31.

82. *Supra* note 31.

83. *Supra* note 31 at 702.

84. *Supra* note 8 at 133-136, 139-143 and 146-148. We note other studies cited in these regards by Mitchell and Carson [at 135] in support of their argument that voluntary demand revelation mechanisms might "outperform" incentive compatible mechanisms:

There also exists a good number of other studies which draw *diametrically opposite conclusions*.⁸⁵ As examples, Isaac, McCue and Plott find pervasive free riding.⁸⁶ Actual contributions by subjects as a percent of true (induced) valuations were only 37, 19, 12, 9 and 9 percent in five separate trials. Harrison and Hirshliefer also found pervasive free-riding behavior in their experiments with the Voluntary Contribution Mechanism.⁸⁷ Subject contributions as a percent of true (induced) values for all trials averaged only 32 percent. Further, Harrison and Hirshliefer's theoretically-based prediction for "perfect" free riding represents a situation in which the subjects still contribute 33 percent of their true values (and not 0 percent).⁸⁸ Thus their findings were consistent with the perfect free riding prediction. Along similar lines, Kim and Walker conclude that "Free riding behavior in the reported experiment was overwhelming, systematic, and very much in accord with economic theory."⁸⁹ Other studies involving applications of the Voluntary Contribution Mechanism that report results similar to those given above include Andreoni,⁹⁰ and Isaac, Walker and others.⁹¹

If one focuses solely on the studies of Chamberlin,⁹² Marwell and Ames,⁹³ and Schneider and Pommerehne,⁹⁴ one might reasonably conclude that there is very little free riding in the Voluntary Contribution Mechanism. However, when these studies are viewed within

Scherr, B. and E. Babb, "Pricing Public Goods: An Experiment with Two Proposed Pricing Systems," *Public Choice*, 35-48, 1975; and Grether, et al., "Economic theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69, 623-638, September, 1979. Sherr and Babb do not use induced values, so they have no basis by which strategic bidding can be identified in either their control group or in the group which participates in a form of a Pivot Mechanism experiment. References to Grether and Plott as providing "evidence" related to free-riding must be viewed as gratuitous: Grether and Plott do not address this issue.

85. Referring to the studies cited above that find little free-riding, Kim and Walker observe that "The outcome furthermore suggests that an explanation of the previous, contrary experimental results [which refute the free-rider hypothesis] is likely to be found in one or more of the invalidating factors that were present in previous experiments." O. Kim, and M. Walker, "The Free Riding Problem: Experimental Evidence," *Public Choice*, 43, 3-24, 1984 at 5.

86. *Supra* note 31, Table 2 at 61.

87. *Supra* note 31, Table 7 at 216.

88. *Id.* at 218.

89. *Supra* note 85 at 4.

90. Andreoni, J., "Privately Provided Public Goods in a Large Economy: The Limits of Altruism," *J. Public Economics*, 35, 57-73, February, 1988.

91. See R.M. Isaac, and J. Walker, "Group Size Effects in Public Goods Provision: The Voluntary Contribution Mechanism," *Quarterly J. Economics*, 103, 179-199, February, 1988; R.M. Isaac, and J. Walker, "Communication and Free-Riding: The Voluntary Contribution Mechanism," *Economic Inquiry*, 26, 585-608, October, 1988; Isaac, R.M. and J. Walker, "On the Suboptimality of Voluntary Public Goods Provision: Further Experimental Evidence," in R.M. Isaac (ed.) *Research in Experimental Economics*, (Greenwich, CT: JAI Press, Vol. 4, 1991); and Isaac, R.M., J. Walker and S. Thomas, "Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations," *Public Choice*, 43, 113-149, 1984.

92. *Supra* note 80.

93. *Supra* note 31.

94. *Id.*

the context of all the relevant literature, it is clear that monolithic "conclusions" one way or the other are simply not justified.

We conclude that the conventional wisdom accepted by many CVM researchers, that "evidence" exists which justifies the general expectation that CVM subjects will not engage in strategic behavior, lacks a substantive foundation. Valuation institutions used in experimental economics are very different from those used in CVM. There is no unequivocal evidence suggesting that subject behavior in CVM is sufficiently "like" subject behavior in experimental institutions. Nor does there exist a body of studies that unequivocally demonstrates the lack of pervasive strategic behavior in valuation institutions used in CVM or those used in experimental economics.

Furthermore, the "incentives issue" is not limited to speculation as to whether or not CVM subjects truthfully report their values. Let us suppose that subjects do *not* behave strategically, and that subjects are "sincere" in their responses in the truth-telling sense of *Zippo*.⁹⁵ Indeed, our experiences with applications of CVM leaves us with the feeling that some subjects do not purposefully lie (we have no proof one way or another, of course). The relevance of the incentives issue nonetheless remains. With the good and the payment hypothetical, we must ask a question that parallels the court's concern with "dangers of faulty perceptions:" what incentives do subjects have to make *good* choices, to undertake the process of researching their preferences to the end of determining what they would *really* be willing to pay for the good in question?

The core of the incentives or perceptions issue may be seen in the following question: do people generally make the same kinds of valuations when asked what they *might* pay for something as they do when they *must* pay? Arrow finds "... the hypothetical bias concerning payment more serious than that about commodities. This is the concern of those who follow the economists' tradition which criticizes hypothetical questions. Verbal answers don't hurt the way cash payments do."⁹⁶

An empirical basis for such concern is suggested by results of Kealy, Montgomery and Dovidio.⁹⁷ They examined the hypothesis that preferences are the same among the two groups of subjects, one that is asked for hypothetical payment and one that is asked for actual payment. They reject this hypothesis, a result which they interpret as suggesting that "... in the hypothetical situation preferences are less well formulated because subjects have *less incentives to seriously contemplate their actual willingness to pay*."⁹⁸ (p. 257, emphasis added)

95. *Supra* note 11.

96. Arrow, K.J., "Comments," pp. 180-186 in Cummings, et al., *Supra* note 8 at 183.

97. Kealy, et al., "Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter?" *J. Environ. Economics and Management*, 19, 244-263, 1990.

98. *Id.* at 257.

As another, perhaps more indirect, example, Smith and Desvousges⁹⁹ find significant differences in marginal valuations of risk with declining risk levels that result from changing information relating to baseline risks of premature death. Economic theory would predict that an individual's values for additional incremental reductions in risk would decline as baseline risk levels decline. This is a simple application of the principle of diminishing marginal utility. However, Smith and Desvousges', in their application of CVM, found *increasing* marginal valuations for risk reductions as baseline risk levels are reduced.¹⁰⁰ Were these CVM subjects "irrational," or did subjects simply have less incentives to seriously contemplate their actual willingness to pay? In our view, one would look to the latter hypothesis to explain the behavior observed in the Smith and Desvousges [1987] experiment.¹⁰¹

CVM VALUES AND REAL ECONOMIC COMMITMENTS

As we argue above, no empirical evidence conclusively establishes whether or not CVM subjects behave strategically, or the extent to which they lack incentives to "sincerely" contemplate their willingness to pay for an environmental improvement. However, one might nonetheless take the position that even though the theoretical basis for CVM is not so "solid," it "works" anyway. Referring to the possibility of strategic behavior, perhaps institutions *don't* matter, or perhaps CVM subjects do not behave strategically. Theory and institutions aside, perhaps CVM *still* yields values which reflect real economic commitments.

Such a methodological position is not without precedent in economics. In response to criticisms of the relevance of value theory, Friedman¹⁰² argues that our models may be valid even if one or more of the assumptions underlying them are not. If behavior predicted by the model reasonably approximates actual behavior, then the model is a legitimate tool for empirical analysis until a better model is developed.

99. V.K. Smith, and W.H. Desvousges, "An Empirical Analysis of the Economic Value of Risk Changes," *J. Political Economy*, 95, 89-114, February, 1987.

100. More precisely, subjects were randomly assigned baseline risk levels. The theoretical expectation would be that subjects assigned low baseline risk levels would value a given reduction in risk less than subjects valuing the same risk reduction but assigned higher baseline risk levels. Similar results are reported in Jones-Lee, M.W., M. Hammerton and P.R. Philips, "The Value of Safety: Results of a National Sample Survey," *Economic Journal*, 95, 49-72, March, 1985. In this study more than 40% of subjects gave the same values for different levels of risk reduction (p. 67, section B.a). Overall, however, responses to some questions reflect consistency with diminishing marginal utility (see, e.g., p. 66, result A.c).

101. We acknowledge a range of other possible explanations for these observations. Any or all of the many other problems with the CVM discussed in this report may be relevant for explaining the observations.

102. Friedman, M., "The Methodology Of Positive Economics," in M. Friedman, *Essays in Positive Economics*, University of Chicago Press (Chicago, 1953), at 15.

To the end of examining this general position, we begin with a brief discussion of studies that compare CVM results with results from indirect estimation methods. We then examine results from studies that directly ask if subjects *actually* pay amounts that they *say* they will in the CVM.

Comparing CVM results with results from indirect estimation methods

One set of efforts designed to assess the validity of CVM values focuses on comparisons with values derived using estimation methods that essentially rely on values indirectly implied by consumer behavior. The methods most often used for such comparisons are variations of the Travel Cost Method (TCM) and the Hedonic Price Method (HPM).¹⁰³ Summaries of these studies and their results are found in Cummings, et al.¹⁰⁴ and Mitchell and Carson.¹⁰⁵

The similarity between value estimates derived using CVM and indirect methods *for some environmental goods* is remarkable. For example, Cummings et al.¹⁰⁶ estimated values associated with the provision

103. The relevance of studies comparing results from a CVM with those from the TCM or HPM for the question of interest here, whether or not CVM values reflect a real economic commitment, is at best indirect. Suppose that one obtains value measures with applications of the CVM and the HPM for a particular environmental good. Suppose further that statistical analyses result in the rejection of the hypothesis that the two values are different; i.e., the CVM value is statistically indistinguishable from the HPM value within some confidence limits. Does this finding establish that the CVM has yielded a "true" or "accurate" value for this particular good, in the sense of a value that reflects a real economic commitment? Of course, the response to this question is: *only* if one is prepared to assert that the HPM value is "true," "accurate," or reflects a real economic commitment. Few, if any, would be prepared to defend such an assertion, however. This is to say that we have no basis for knowing in any precise way the relationship between HPM values or TCM values and real economic commitments that subjects may be willing to make for public goods. Thus V.K. Smith [in Cummings, et al., *Supra* note at 174] correctly notes that "Comparisons of indirect [HPM and TCM] and CVM estimates are largely useless unless we can bound the nature of the errors associated with the indirect estimates." Mitchell and Carson [*Supra* note 8, at 188 and 190] note that "Although suggestive, these comparisons have their limitations...it is well recognized that estimates based on the indirect methods...are themselves prone to error, owing to that fact that their indirect relationship with the good being valued necessarily requires the use of largely arbitrary assumptions to arrive at a WTP amount...convergent validity involves comparing two estimates, neither of which can be assumed to represent the unmeasured variable. Although it is reassuring when the two measures of the same concept are close to one another, the possibility exists that both are inaccurate." All of this is to simply say that a showing that CVM values for a particular good are "close" to those derived with an indirect estimation method obviously demonstrates no more than that the two methods yield similar results. While such a showing may be interesting, it does *not* provide conclusive evidence that the CVM has captured a real economic commitment by subjects.

104. *Supra* note 8 at Chapter 6.

105. *Supra* note 8 at Chapter 9.

106. R. Cummings, et al., "Measuring the Elasticity of Substitution of Wages for Municipal Infrastructure: A Comparison of the Survey and Wage Hedonic Approach," *J. Environmental Econ. Management*, 13, 269-276, 1986.

of municipal infrastructure in boomtowns using HPM and CVM. They failed to reject the hypothesis that mean values for the elasticity of substitution between wages and municipal infrastructure calculated from the two methods are the same. Smith, et al.¹⁰⁷ found a close correspondence between values for water quality derived using CVM and a "simple" TCM model. In 13 of 15 comparison studies considered by Cummings, et al., CVM values were within ± 50 percent of values derived with indirect methods.¹⁰⁸

However, in a number of other cases, CVM values were *not* "close" to values derived from indirect methods. This observation led Cummings, Brookshire and Schulze¹⁰⁹ to ask: are there common characteristics among goods for which CVM and indirect value estimates are "close" and those for which CVM and indirect values are not close? They loosely suggest that such might be the case. Drawing on the paradigm of a perfectly competitive market, they defined as "Reference Operating Conditions" (ROCs) a subject's familiarity with a commodity, their valuation or choice experience with the commodity, little uncertainty as to the attributes of the commodity, and the use of a WTP measure.¹¹⁰ They then suggested that a parallel exists between goods that satisfy these ROCs and those for which CVM values were found to be "close" (within ± 50 percent) to values derived with indirect methods.

Unfortunately, Cummings, Brookshire and Schulze's view of the implications of this parallel does not help to determine the extent to which CVM values reflect real economic commitments. Values from an estimation study in which ROCs are in some (undefined) sense satisfied would seemingly be accepted as "true" or "accurate" by them.¹¹¹ They appear to assume that indirect methods may be taken on their face as having satisfied the ROCs. Thus, they appear to view closeness of CVM values to those derived using indirect methods to imply closeness of CVM values to "true" values reflecting a real economic commitment. We have noted above the absence of any basis for this line of deduction, a "lack" noted by others.¹¹² The "reference accuracy" dis-

107. V.K. Smith, W.H. Desvousges and A. Fisher, "Estimating Environmental Benefits," *Amer. J. Agricultural Econ.*, 68, 280-290, May, 1986.

108. *Supra* note 8, 100-101.

109. *Supra* note 8 at Chapter 6.

110. *Id.* 8 at 102-105.

111. This would certainly appear to be the case in their discussions of the ROCs (Cummings, et al., *supra* note at Chapter 6). However, in Chapter 13 (pp. 230-231), the authors acknowledge the infant stage of our knowledge of criteria for accuracy in CVM measures, discuss other possible ROCs, and call for "...imaginative thinking and research relevant to the specification of precise and defensible ROCs." (p. 230)

112. "...there are no grounds for considering travel cost- or hedonic price-based measures to be more accurate than CV-based measures." Mitchell and Carson (*supra* note at 204). See also the critique of V.K. Smith (in Cummings, et al., *supra* note 8) and R. Mitchell and R. Carson, "Evaluating the Validity of Contingent Valuation Studies," in G. Peterson,

cussions of Cummings, Brookshire and Schulze may be viewed as historically useful as a call for the economics profession's attention to be focused on means by which the accuracy and reliability of CVM measures might be assessed. Much more work is required, however, to determine the nature of possible "conditions" required for any criterion for reference accuracy.¹¹³

For the time being, the results from comparative studies offer little evidence regarding the extent to which real economic commitments are reflected in CVM values.

CVM values and real economic commitments

We now examine the extent to which CVM values *have been shown* to approximate *actual* amounts that people are willing to pay for a good. However, "hard" evidence relevant to this question as it applies to environmental goods is very limited, which reflects the fact that received valuation methods cannot generally be applied to non-deliverable public goods (e.g., we cannot actually auction hypothesized improvements in air quality). This limitation raises an obvious question: can one extrapolate relationships between a CVM and a "real" value that are found in experiments with private goods to public, environmental goods?

Most researchers asserting that CVM values are at least "reasonable" emphasize the importance of the subject's familiarity with the good. Mitchell and Carson argue that "The key problem facing the designer of a CV study, we say, is the novelty of valuing a public good, given the respondents' varying degrees of familiarity with the good being valued and how they currently pay for its provision."¹¹⁴ If one accepts this argument, and set aside other problems with CVM, one could argue that the hallmark of a "good" CVM design is making the subject as familiar as possible with the good being valued.¹¹⁵ A "perfect" design would then be one where the subject is demonstrably familiar with the good. Common, private goods surely meet this criterion.

B. Driver and R. Gregory (ed.s), *Amenity Resource Valuation: Integrating Economics with Other Disciplines* (State College, PA: Venture Publishing Co., 1988), at 188-90.

113. Among others, Mitchell and Carson (*supra* note 8, at 189) argue that a political market model may be preferable to a consumer market model for the valuation of public goods. This preference is justified on the grounds that people value public programs and amenities in referenda. They argue that "Acceptance of a referendum model would imply quite different ROCs; for example voters often make binding choices about amenities with which they have relatively slight familiarity." As discussed above the key descriptor here is "binding," which implies that individuals view the referendum as real. The weight of this argument when applied to a hypothetical referendum where the good to be provided and payment for the good is hypothetical is debatable. See Cummings and Harrison, *supra* note 8, at 59-62, for further critical discussion.

114. *Supra* note 8 at 188.

115. Many commentators assert that CVM values are likely to be more "reliable" and "valid" if the subject is familiar with the commodity. We are aware of no theory, however, that relates the degree of familiarity with the quality of a subject's valuation process.

If the CVM values for such goods do not closely approximate their real values, how can one expect a *better* predictive performance of CVM when the good being valued is unfamiliar? Thus, for the limited purposes of this section's inquiry, results from private-good experiments may be every bit as meaningful as results from studies that involve public goods.

Private Goods

Dickie, Fisher and Gerking¹¹⁶ obtained values for a pint of strawberries by using CVM, and by actually selling the strawberries to households. They concluded¹¹⁷ that they could not reject the null hypothesis of structurally identical demand equations estimated from actual sales and CVM data. However, their results are mixed if one examines them in further detail. Specifically, they found that there were large differences in the actual and hypothetical demand schedules using the predicted demand equations based on their regression results.¹¹⁸ Depending on the price one uses, the hypothetical demand curve can overstate the quantity demanded from 351.4 percent to 68.6 percent, if all interview team results are included. Even if an errant team's results are excluded, the two can differ by as much as 100 percent or as little as 3.5 percent, again depending on the particular price evaluated.

Another problem with the general conclusion of Dickie, Fisher and Gerking's study is that it rests heavily on the deletion of one outlier. Apparently,¹¹⁹ one happy customer said he would buy 10 pints of strawberries at a price of 60 cents, and that he planned to freeze or can 50 percent of them. What makes this an outlier? The question actually asked of subjects did not ask what the current consumption demand was, so there is no obvious basis for calling this an outlier. With it included, the total hypothetical and real expenditures per head are, on average, about 58 percent different. We conclude, therefore, that there is hardly unequivocal support in this study for the view that hypothetical and actual questions generate the same demand schedules.

Furthermore, a number of other studies involving "familiar" goods demonstrate there may be significant differences between CVM values and market values. Bishop and Heberlein¹²⁰ found significant differences between CVM estimates for subjects' WTA for goose per-

116. M. Dickie, A. Fisher and S. Gerking, "Market Transactions and Hypothetical Demand Data: A Comparative Study," *J. Amer. Statistical Ass.*, 82, 69-75, 1987.

117. 116 at 75.

118. 116 at Tables 3 and 4, and Figure 1, at 74.

119. 116 at p. 71.

120. *Supra* note 17 at 929-929.

mits and WTA values based upon actual cash payments.¹²¹ In a later study, Bishop and Heberlein¹²² obtained CVM and actual cash values for deer permits, and found that WTP values were significantly overstated in the CVM relative to the cash market. Referring to the "familiarity" requirement for real economic commitment in CVM values, they note that "Clearly, if contingent valuation is capable of giving unbiased estimates of real values, it should have done so here."¹²³ However, their results:

"... indicate bias. People were more willing to sell their goose hunting permits for real dollars than they indicated they would be in the contingent market. Preliminary results from the deer study indicate that in an auction framework, CVM will overestimate willingness to pay [. . .] money is a powerful stimulus and real money is more powerful than hypothetical money."¹²⁴

Sellar, Stoll and Chavas [1985]¹²⁵ offered evidence concerning the real economic commitment of CVM values for private goods. They used CVM to obtain WTP values for improvements in boat docking facilities in four lakes in Eastern Texas. Values are derived with a TCM model, an open-ended CVM study, and a dichotomous choice CVM study. Two results from this study are of interest for our purposes, although the evidence presented is very weak.¹²⁶ First, both CVM applications yielded a demand relationship with a positive slope for one of the lakes (Lake Somerville). The explanation offered for this anomaly bears on the "familiarity" issue. Specifically, the researchers noted:

"It appears that specifying a contingent market under conditions where boaters were not used to paying a launch fee may have caused problems [. . .] This seems to add to the evidence that the contingent valuation instruments used to collect data for analysis must be designed so that behavior by the respondent is as familiar as possible."¹²⁷

121. Mitchell and Carson, *Supra* Note 8 at p. 195-199, dispute these conclusions. Hanemann, M., "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *Amer. J. Agricultural Econ.* 66, 332-341, 1984, also re-evaluates the conclusions of Bishop and Heberlein's study, demonstrating the extreme sensitivity of results to alternative statistical assumptions.

122. In Cummings, et al., *supra* note 8, at Chapter 8 (130-11, particularly).

123. *Id.* at 134.

124. *Id.*

125. C. Sellar, J. Stoll and J. Chavas, "Valuation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques," *Land Econ.*, 61, 156-175, 1985.

126. Sellar, et al., "Specification of the Logit Model: The Case of Valuation of Nonmarket Goods," *J. Environ. Econ. Management*, 13, 382-390, December, 1986, demonstrate that the anomalous results obtained from data concerning Lake Conroe, discussed below, are attributable to a mis-specification of the estimation model used for analyses reported in the earlier 1985 study. The implications of such model mis-specification for results related to value estimates for the remaining three lakes are unclear, but suggests the need to view all of the results in the 1985 study with healthy skepticism.

127. *Supra* note 125 at 169, 174.

Secondly, CVM values from their open format questionnaire resulted in *negative* surplus measures for two of the lakes (Lake Conroe and Lake Houston). These results, along with similar results from a third lake (Lake Livingston), indicated to Sellar, Stoll and Chavas that "... people reported they were willing to pay less for an annual ramp permit than they already paid in total launch fees over the year on a per visit basis."¹²⁸

Boyce et al.¹²⁹ used experimental methods to elicit WTA and WTP for a Norfolk Pine tree, which is a small house plant. In one set of experiments subjects were told that if they did not buy the tree (the WTP case), or sell it back to the experimenter (the WTA case), the tree would be destroyed "then and there." In another case, subjects were asked WTA and WTP for a tree, but were told nothing about the disposition of the trees left in the hands of the experimenters. They also elicited CVM values from the same subjects. Given the data they report on sample means and medians, one can easily gauge the relationship between hypothetical CVM values and values that represented real economic commitments. Using means from their data, hypothetical CVM values over-stated real economic commitments by 27 percent, 117 percent, 85 percent and 150 percent in the WTP-NoKill, WTP-Kill, WTA-NoKill, and WTA-Kill experiments, respectively. Using median values, which is arguably more reliable given the likely skewness of the data, CVM values over-stated real economic commitments by much larger percentages: 82 percent, 150 percent, 150 percent and 400 percent, respectively. These data suggest that values elicited with CVM can be highly inflated with respect to real economic commitments.

Neill, Cummings, Ganderton, Harrison, and McGuckin¹³⁰ investigated the extent to which subjects will actually pay the amounts they said that they would in a CVM. They conducted two series of experiments. In the first, they compare the valuations elicited for a small, framed painting of a Southwest rural scene by an unknown Navajo Indian painter using a CVM and a Vickrey auction. The subjects were required to actually pay for the painting out of their own pockets.¹³¹

128. *Id.* at 169.

129. R. Boyce, T. Brown, G. McClelland, G. Peterson and W. Schulze, "Experimental Evidence of Existence Value in Payment and Compensation Contexts," in Boyle, K. and T. Heekin (ed.s), *Western Regional Research Project W-133: Benefits and Costs in Natural Resources Planning*, Interim Report 2, Department of Agricultural and Resource Economics, Univ. of Maine, Orono, July 1, 1989.

130. H.R. Neill, R.G. Cummings, P.T. Ganderton, G.W. Harrison & T. McGuckin, "Hypothetical Surveys and Real Economic Commitments", *Economics Working Paper B-93-1*, Department of Economics, College of Business Administration, University of South Carolina, 1993.

131. This qualification is unusual in most experiments, since it is common practice in these settings to endow subjects with some cash with which to bid. The effect of such

Sixteen subjects were given a series of "training experiments" in the use of a Vickrey auction with induced values,¹³² after which they participated in the Vickrey auction for the painting. Forty-one subjects participated in a CVM experiment in which maximum WTP for the art object was obtained. Average valuations were \$37.04 and \$9.49 for the CVM and the Vickrey auction, respectively, with medians of \$30 and \$6. The hypothesis of no significant difference between the distribution of bids obtained from the two experiments was tested and rejected.¹³³

The CVM for the art object did not mimic the rules of provision of the Vickrey auction. Instead, it followed some traditional CVM questionnaires by not specifying explicitly what the rules of provision of the public good would be. This weakens the conclusions that can be drawn from these experiments, since it is *conceivable* that just altering the (hypothetical) rules of provision in the CVM to match those used in the (non-hypothetical) Vickrey auction could alter the elicited WTP in the CVM. In this case one might well conclude that the problem is not the hypothetical nature of the payment, but rather the hypothetical nature of the provision rule.

However, in a second series of experiments, Neill et al.¹³⁴ rejected this hypothesis. They directly compared the valuations obtained from a hypothetical Vickrey auction with those obtained in a Vickrey

endowments has not been examined in a controlled manner. To allow for subjects that might be cash-constrained during the experimental sessions, an interest-free loan for the weekend was arranged. The experiment was conducted on a Friday afternoon and subjects knew that the loan had to be repaid by 5:00 pm the following Monday. A loan contract was signed between the eventual winner and an impecunious research assistant of the experimenter (who had been privately given a large pot of cash by the experimenter for this purpose). The loan was repaid as agreed.

132. Actually there were 22 subjects in this stage of the experiment, but 6 are ignored for reasons explained below.

133. Six of the subjects in these experiments participated in both the CVM and the Auction, but were *not* included in the auction data analyses described above. Their behavior provides a particularly stringent test of the extent to which the CVM constitutes a real economic commitment. There is an obvious danger in using the same subjects in the CVM and the auction since there is a chance that this could cause biases in favor of the CVM as subjects may have felt obliged to actually pay the amount they hypothetically recorded so as to avoid embarrassment. A possibility exists for the opposite bias, of course, if subjects are angered by being asked to actually pay for something that they thought involved an honest and hypothetical survey. This is a "trick" used by many salesman: asking a subject how much they would be willing to pay for a good, and then confronting them with a (smaller) actual price. These six subjects volunteered to participate in the auction after being privately told that it involved them bidding for the art object out of their pocket. In any event, the six in-sample responses are perfectly consistent with the out-of-sample responses analyzed above. The subjects reported CVM valuations of \$65, \$7, \$25, \$25, \$100 and \$5, for an average of \$37.83; their corresponding auction values were \$18, \$6.27, \$2.50, \$12, \$8 and \$5, for an average of only \$8.63. Using a matched-pairs Wilcoxon test one rejects the null hypothesis that these values are the same at a 4 percent critical level, even though the sample is quite tiny.

134. *Supra*, note 130.

auction in which the bid involved a credible financial commitment. Fifty-one subjects in the hypothetical auction had an average (median) bid of \$301 (\$60), and 60 subjects in the real auction had an average (median) bid of only \$12 (\$5). The hypothesis of no significant difference between the distribution of bids obtained from the two experiments was again rejected.

A similar conclusion resulted from an experimental study of the dichotomous choice (DC) format for CVM by Cummings, Harrison and Rutström.¹³⁵ They conducted two series of experiments. The first series attempted to ascertain if the hypothetical DC format was incentive compatible, as claimed by many CVM researchers.¹³⁶ Using three different private goods (an electric juicer, a pocket calculator, and a box of chocolate truffles) they found that hypothetical DC responses significantly overstated real DC responses for the same good. This series of experiments used simple presentations of the hypothetical DC questions to establish, as a baseline, if an incentive compatibility problem existed.

The same qualifications applied for the study of Cummings et al.¹³⁷ as for the open-ended experiments of Neill et al.¹³⁸ Whether or not the hypothetical DC question can be presented in a way that reliably elicits truthful responses is not a matter that can be determined *a priori*. These experiments demonstrate that *some* words do not "work" for *some* subjects.¹³⁹ There is no evidence from these experiments to suggest that the CVM can *reliably* elicit real economic commitments that reflect maximum WTP.

Public Goods

Kealy, Montgomery and Dovidio¹⁴⁰ examine the predictive validity of CVM values for actual cash payment for a private good (a candy bar) and a public good (a deacidification program for lakes in the Adirondack region). For one of these goods, the researchers asked

135. R.G. Cummings, et al., "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?", *Economics Working Paper B-92-12*, Department of Economics, College of Business Administration, University of South Carolina, 1992.

136. See Hoehn and Randall *supra* note at 237, Michell and Carson *supra* note at 151, or Carson, R.T., "Constructed Markets," in J.B. Braden and C.K. Kolstad (eds.), *Measuring the Demand for Environmental Quality*, North-Holland (Amsterdam, 1991) at p. 142.

137. *Supra* note 135.

138. *Supra* note 130.

139. Furthermore, "focus groups" do not obviously help. None of the subjects in a focus group interview found any difficulty saying that they interpreted the hypothetical DC question the way that the researchers interpreted it. Of course, the opinions of subjects in focus groups cannot be viewed as more than potential clues as to what the subject actually thinks the question means.

140. *Supra* note 97.

a group for their WTP with the understanding that the subjects were to actually pay their offered amount in two weeks. They asked a second group for a WTP within a hypothetical (CVM) context. Two weeks later the second group of subjects were asked for an *actual* payment. The authors pose the question: when asked for a WTP, does the subject's knowledge that he must actually pay for the good affect the amount that is reported? They tested two hypotheses. The first was that preferences were the same among the two groups of subjects. They rejected this hypothesis, a result which suggested to them that "... in the hypothetical situation preferences are less well formulated because subjects have less incentives to seriously contemplate their actual willingness to pay."¹⁴¹ The second hypothesis relates to the extent to which WTP amounts initially reported by each of the two groups predicted behavior at the end of two weeks. They found that, for both goods, initial verbal reports were reasonably correlated with actual payment.¹⁴² They also examined the effect of foreknowledge that subjects would be required to actually pay their stated WTP on the degree of correlation between hypothetical WTP and actual payment behavior. They concluded:

... individuals are more likely to overstate than to understate their [hypothetical] WTP when they are not expecting to have to make an actual payment [...] [F]oreknowledge of an obligation to pay in accordance with one's verbal statements of willingness to pay has a positive impact on the predictive validity of contingent values ...¹⁴³

Another experiment with a public-like good is reported by Seip and Strand.¹⁴⁴ A sample of 101 Norwegians were asked in personal interviews whether they would pay 200 Norwegian Kroner for membership in the Norwegian Association for the Protection of Nature (Norges Naturvernforbund, NNV), which is the largest and best established private environmental organization in Norway. Sixty-four subjects responded "yes." A short time later, the 64 subjects that answered yes in the CVM study were sent letters encouraging them to join the NNV at a membership cost of 200 Kroner. There was no reference in these letters to the earlier CVM study. One month later a second mailing was sent to subjects that had not joined the NVA as a result of the first let-

141. *Id.* at 257.

142. They offer the following caveat, however. "We caution the reader that there is a strong possibility that our comparisons of first-period report with second-period behavior could be overly favorable to the predictive validity of the CVM. Subjects may be reluctant to contradict their previous response even if it overstates their true WTP." *Id.* at 258.

143. *Id.* at 259-260.

144. K. Seip, and J. Strand, "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment," *Environmental and Resource Economics*, 2, 91-106, 1992.

ter. Again, reference was not made to the initial CVM study. At the end of the second mailing only *six* of the 64 "yes" respondents in the CVM actually paid the 200 Kroner to join the NNV.

The researchers interviewed by telephone 25 of the 58 "yes" respondents in the CVM study who declined to join the organization after saying they would. Emphasizing the scientific nature of their study, the researchers asked the subjects why they failed to respond to invitations to join the NNV. Twenty-four of the 25 subjects indicated that their WTP expressed in the CVM study was an expression of their WTP for environmental goods in general, not their willingness to pay for the NNV in particular. When asked if they would like to change their original WTP, 17 of the 25 subjects indicated that they would *lower* their response. The authors concluded that the results were ". . . discouraging, by indicating that the CVM can imply quite serious biases of overvaluation, in particular when (like here) the good to be valued is rather abstract and it is difficult to attach concrete environmental values to it."¹⁴⁵

Duffield and Patterson¹⁴⁶ used mail surveys to obtain three sets of values for a fund to be established for the purpose of leasing water rights to be used for the preservation of in-stream flows in a set of Montana rivers. They asked one set of subjects (Cash-TNC) to *actually* make a tax deductible contribution to an *actual* fund, the "Montana Water Leasing Trust Fund," that had been established by the Montana Nature Conservancy. They asked a second group (Hypo-TNC) a hypothetical question: if contacted in the next month with a request to make a tax deductible contribution to the Montana Water Leasing Trust Fund, how much *would* they be willing to contribute? For the third group (Hypo-UM), both the fund *and* payment were hypothetical, although they were also told that a fund *could be* established. The researchers then asked the maximum amount that subjects would contribute if contacted in the next month with a request for a contribution. Comparisons of results from Hypo-UM with those from Cash-TNC and Hypo-TNC are made difficult by differences in the design and follow-up procedures used. Unlike procedures used in the Cash-TNC and Hypo-TNC experiments, subjects in Hypo-UM did not receive a brochure describing the trust fund. Extensive follow-up mailings were used in the Hypo-UM experiment, but not in the Cash-TNC and Hypo-TNC experiments. We therefore only compare results from the two similar experiments: Cash-TNC and Hypo-TNC.

145. *Id.* at 3.

146. J. Duffield, and D.A. Patterson, "Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers," *Unpublished Draft Manuscript*, Department of Economics, University of Montana, Missoula, 1992. We are grateful to John Duffield for permission to use these data.

The results of Duffield and Patterson for residents and non-residents are summarized in Table 1.¹⁴⁷ Column 2 is the number of delivered questionnaires.¹⁴⁸ Column 3 is the number of questionnaires that were returned to experimenters. Columns 4-9 are numbers of respondents reporting WTP between zero and \$250.

A question of immediate importance for our inquiry as to differences between hypothetical values and those that represent real economic commitments arises from the large differences in response rates. Is the response rate of 11 percent for the questionnaire requesting *real* payment, which is half the response rate for the questionnaire requesting *hypothetical* payment, simply a random occurrence, or does it imply a non-response bias? It is difficult to ignore the potential implications of the differences in response rates between these experiments involving real and hypothetical payments, differences seemingly attributed to non-response bias and/or free-riding behavior by Duffield and Patterson.¹⁴⁹ If one ignores these differences, the estimate of the sample average WTP is based on the percentage of returned questionnaires represented by subjects that bid \$10, \$25, \$50, \$100 or \$250, yielding an average household value¹⁵⁰ of \$9.40 for actual cash contributions and \$12.70 for hypothetical contributions. The CVM overestimated real economic commitments by some 35 percent. If instead one takes non-response as being indicative of a zero value, quite different estimates of population WTP are obtained: \$0.98 if estimates are based on real economic commitments, and \$2.97 if estimates are based on CVM results. In this case, CVM overestimated real economic commitments by 203 percent. Admittedly, this is only one example of an alternative treatment of non-respondents, but it does illustrate the possible bias of CVM.

In any case, the results of Duffield and Patterson suggest an overestimate of real economic commitments from values obtained from the CVM. The extent of this overestimation in their study depends critically on how one interprets and then deals with the non-response question.

CLOSING REMARKS

The courts must soon determine the efficacy of CVM for providing "reliably calculated" statutory use values. The decisions will reflect the degree of speculation that can be tolerated in the court's

147. These data are taken from *Id.* at Tables 2, 3 and 4.

148. Excluding questionnaires returned due to change of address or incorrect address.

149. *Id.* at 26.

150. The sum of contributions weighted by the percent of "households" (returned questionnaires) offering the contribution.

determination of the legal "necessity" of CVM survey data in CERCLA cases. Relating both to the court's concern with faulty perceptions and the "sincerity" of subject responses, we have reviewed the extent to which the CVM valuation institution provides market-like incentives for "good choices" and truthful reporting of values.

We maintain that existing empirical results from experimental economics provide little, if any, evidence that CVM subjects generally will not free-ride. The valuation institutions used in experimental economics are patently dissimilar from those used in CVM. Moreover, there is no evidence that unequivocally establishes that subject behavior in CVM is similar to subject behavior in the institutions used in experimental economics. Nor does there exist unequivocal evidence concerning the pervasiveness of free-riding in the CVM valuation institution or valuation institutions used in experimental economics. There simply exists no basis for non-speculative, dogmatic statements regarding free-riding behavior in the CVM one way or another.

We also examined whether people will actually pay amounts they report in CVM surveys. The little evidence that exists regarding this question is at best inconclusive. A few studies show that CVM values may be "close" to values that reflect real economic commitments. However, a number of other studies show that CVM values overstated real economic commitments, and that these overstatements *can be* quite large. The courts must assess the demonstrated *potential* for such overestimates in their deliberations concerning the legal necessity of results from CVM surveys in CERCLA litigation.