NATURAL RESOURCES JOURNAL

Volume 34
Issue 1 *The International Law of the Hydrologic Cycle*

Winter 1994

Lighting Rods, Dart Boards, and Contingent Valuation

V. Kerry Smith

V. KERRY SMITH*

Lightning Rods, Dart Boards, and Contingent Valuation

ABSTRACT

This Article evaluates existing literature on the performance of contingent valuation. It considers: the performance of the methods in natural resource damage assessments; the implications of the Ohio court decision for the use of CVM; the relationship between laboratory experiments on economic valuation and CVM; and the implications of the NOAA Panel's recommendations.

I. INTRODUCTION

This paper was prepared as a companion to the Cummings-Harrison's (CH)[1] evaluation of the use of the contingent valuation method (CVM) for natural resource damage assessments. Their overview seeks to stimulate discussion on what we actually know about CVM's performance. By taking a strong position, often at variance with both the *Ohio* court's decision and the positions of CVM practitioners, Cummings and Harrison want to promote some rethinking of entrenched positions in the CVM debate. Indeed, in early discussions of their research they described the paper's intent as that of a lightning rod for continued dialogue about CVM. Because lightning rods are intended to "draw fire," my evaluation focuses primarily on areas of disagreement and offers five primary conclusions:

(1) CH's interpretation of how CVM values should reflect "real economic commitments" is misleading.

(2) Experimental economics makes its greatest contributions in evaluating institutions—the rules governing social interactions. It cannot claim the same advantages when it ventures outside the setting of induced preferences, controlled information flows, and specifically defined incentive structures.

* Arts and Sciences Professor, Departments of Economics and School of the Environment, Duke University and Resources for the Future University Fellow. Thanks are due Shelby Gerking for providing the primary data from the Dickie, Fisher, Gerking experiments, to Glenn Harrison for numerous helpful discussions about the literature on experimental economics, and to both Glenn and Richard Carson for detailed comments on an earlier version of this paper, though neither is responsible for any of my errors. Thanks are also due Carla Skuce for preparing and editing multiple earlier drafts. Partial support for this research was provided by the University of North Carolina Sea Grant Program Project No. R/MRD-22.

1. R. Cummings & G. Harrison, *Was the Ohio Court Well Informed in Their Assessment of the Accuracy of the Contingent Valuation Method?*, 34 Nat. Res. J. no. 1 (1994).

(3) Contingent valuation surveys used for natural resource damage assessments do not seek to "create markets" or measure "market values."

(4) CH's summary of the evidence available to the "*Ohio* Court" about the performance of CVM is somewhat misleading.

(5) Under one restricted definition, nonuse values can be measured from actual expenditures on market goods, but the results should not be treated as inherently more reliable than CVM estimates.

Section II provides context for the dispute within the evolving literature on the law and economics of natural resource damage assessments. Section III considers the meaning of CH's "real economic commitments" as consistency checks for CVM responses. Section IV raises questions about the relevance of findings from laboratory experiments to CVM. Section V considers an alternative method for measuring nonuse values recently proposed by Larson.[2] Finally, the last section closes on the same general issues as did CH. Because it was prepared after their paper and with the benefit of the NOAA panel report on CVM, it is possible to consider both the decision in Ohio and the panel report in evaluating what we know about CVM.

II. THE DISPUTE IN CONTEXT

Before turning to CH's thought provoking evaluation of the experimental economics literature and the insights it offers for evaluating the validity of CVM, it is important to provide some context. CVM has attracted the attention of lawyers in the public and private sectors because of three interrelated events that are likely to continue to influence federal policy associated with environmental resources.

First, under a variety of legislative mandates a doctrine of natural resource damage liability has evolved as a residual liability for the damages associated with injuries to natural resources from releases of hazardous substances or oil,[3] The Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) of 1980, and the Oil Pollution Act of 1990 have contributed directly to establishing these concepts.

---

2. Most analysts who hold this view would argue that the approach advocated by Larson simply redefines one component of use values to be designated as nonuse or existence values. See D. Larson, *On Measuring Existence Values*, has appeared now *Land Economics* (Vol. 69, Nov. 1993, 377-88).

3. The concept of natural resource damage liability finds its origin in the Trans Alaska Pipeline Act in the mid 1970s. For a description of the progressive evolution of this liability concept see B. Breen, *Citizen Suits for Natural Resource Damages: Closing a Gap in Federal Environmental Law*, 24 *Wake Forest L. Rev.* 851 (1989).

Second, implementation of CERCLA has led to challenges to regulations issued as part of the rule-making process. The most significant of these was State of Ohio v. Dept. of Interior.[4] CH highlight two aspects of this decision: 1) broadening of the types of values to consider in computing the damages associated with injuries to natural resources; and 2) the acceptance of CVM as a "best available procedure". Unfortunately, CH omit another aspect of the decision.

The court ruled that, in general, the cost of restoring the injured natural resource should be the measure of damages. The damages awarded would differ from the cost of restoration only where such costs are "grossly disproportionate" in relation to the future losses of use and nonuse values.[5] It seems clear from the legislation and the court ruling that Congress anticipated that restoration costs would exceed total value of injuries for many cases. As a result, definitions of economic losses excluding nonuse values will increase the likelihood that injured resources will *not* be restored. Because CVM is widely regarded as the only methodology for measuring total values that include nonuse values, decisions to allow nonuse values must address the validity and reliability of contingent valuation. This additional component of the court ruling is important. By re-affirming restoration cost as the general measure of damages the *Ohio* court may have answered CH's closing comment that courts must evaluate the potential for bias in CVM estimates as part of deciding their potential role in litigation. In fact the court decision highlighted the congressional skepticism about all methods for measuring the value of natural resources. In explaining the preference for restoration costs over a "lesser of" provision included in the Department of Interior's rules for implementing damage assessment, the decision suggests that:

> Whether a particular choice is efficient depends on how the various alternatives are valued. Our reading of CERCLA does not attribute to Congress an irrational dislike of 'efficiency;' rather it suggests that Congress was skeptical of the ability of human beings to measure the true 'value' of a natural resource.[6]

The third contextual element concerns the Exxon Corporation sponsored CVM research. In what may be the largest privately spon-

4. Ohio v. United States Dep't of the Interior, 880 F.2d 432 (D.C. Cir. 1989).

5. The decision states:

"Our reading of the complex of relevant provisions concerning damages under CERCLA convinces us that Congress established a distinct preference for restoration cost as the measure of recovery in natural resource damage cases. This is not to say that DOI may not establish some class of cases where other considerations, that is, unfeasibility of restoration or grossly disproportionate cost to use value, warrant a different standard." *Id.* at 55. Here the Court is adopting a broad definition of use value including what the ruling describes as "passive use values" that correspond to all nonuse values.

6. *Id.* at 456-57.

sored evaluation of an economic methodology, Exxon supported research applying CVM to estimate the values for four different experimental commodities. The researchers criticized CVM, arguing that they had demonstrated that the CVM surveys did not measure preferences,[7] were incapable of estimating nonuse values[8] and even that it paralleled "voodoo practices".[9]

While one might expect judgments about the plausibility of these findings to await peer reviews, they have gained some measure of acceptance among economists unfamiliar with past CVM research. Indeed, a recent review of the use of CVM in natural resource damage assessment in the *Harvard Law Review* is one such example.[10] This review bases its critique of CVM exclusively on the Exxon research, following closely the summary prepared by Shavell[11] and concluded observing that: "At least where nonuse values are concerned, CV will always be *the dart board of valuation techniques*—a dart board with numbers so inflated they seriously skew the scoring."[12]

Any discussion of the performance of CVM enters a "highly charged" public debate. In such a setting it is important to scrutinize exactly what can be concluded from available research and to carefully document the relevant qualifications. My criticisms of the CH review are intended to reinforce their primary objective—stimulating a more systematic evaluation of what we know about all of the methods for estimating the values people have for non-marketed resources.

III. REAL ECONOMIC COMMITMENTS

CH argue that methods for measuring people's preferences or values for any commodity should be consistent with the *real economic*

---

7. J. Hausman, Comments to NOAA Panel on Nonuse Damage Assessment Methodology (July 22, 1992) (unpublished comments submitted to National Oceanic and Atmospheric Administration Contingent Valuation Panel, Department of Economics, Massachusetts Institute of Technology).

8. D. McFadden, Comments on Constructed Market Methods for Attributing Nonuse Values to Environmental Resources (July 10, 1992) (unpublished comments submitted to National Oceanic and Atmospheric Administration Contingent Valuation Panel, Department of Economics, University of California, Berkeley).

9. J. Daum, Legal and Regulatory Aspects of Contingent Valuation (April 2-3, 1992) (unpublished manuscript presented at the Cambridge Economics Inc. Symposium on Contingent Valuation: A Critical Assessment, held in Washington, D.C.).

10. Note, *"Ask a Silly Question . . . ": Contingent Valuation of Natural Resource Damages*, 105 Harv. L. Rev. 1981 (1992).

11. S. Shavell, Should Contingent Valuation Estimates of the Nonuse Value of Natural Resources Be Used in Public Decisionmaking and the Liability System? (April 2-3, 1992) (unpublished manuscripts presented at the Cambridge Economics Inc. Symposium, Contingent Valuation: A Critical Assessment held in Washington, D.C.).

12. *See supra* note 10, at 1990.

commitments of individuals.[13] They outline two questions to be answered by an evaluation of CVM:

(1) ". . . does substantial evidence exist that would support the claim that subject behavior within the CVM valuation institution is reasonably similar to behavior assumed in economic theory?" and

(2) ". . . will people actually pay amounts reported in CVM surveys?"[14]

In developing their answer to the first question, CH consider the experimental economics literature findings on the importance of free-riding behavior. More specifically, they suggest that these experimental studies indicate that the record on free-riding behavior is quite mixed.[15]

Before turning to the experimental literature, we should determine how other empirical models intended to describe the behavior of economic agents are evaluated, and whether these evaluation methods hold promise in judging CVM.

---

13. CH credit the original source of this criterion to *Measuring the Demand for Environmental Quality* (J. Braden & C. Kolstad eds., 1991). In their introduction to the volume, J. Braden, C. Kolstad and D. Miltz observe that "[m]any economists are loathe to base economic values-values that will be used to allocate real resources-on information that does not grow out of real economic commitments." *Id.* at 12.

14. R. Cummings & G. Harrison, *Identifying and Measuring Nonuse Values for Natural and Environmental Resources: A Critical Review of the State of the Art* (1992) (published by American Petroleum Institute).

15. This conclusion contrasts with the view of free-riding as unimportant that sometimes has been attributed to Mitchell and Carson in R. Mitchell & R. Carson, Using Surveys to Value Public Goods: The Contingent Valuation Method (1989) (published by Resources for the Future). Actually these authors' conclusions relate to one specific context. That is, Mitchell and Carson (MC) describe the context for their review of the experimental literature and the conclusions they draw from that review in specific terms. They note that they considered "a series of experiments which use **nontrivial rewards** to test the prevalence of the free-riding form of strategic behavior under **realistic fieldlike conditions** designed to produce free-riding." *Id.* at 139 (emphasis added).

While their conclusions are strongly stated, they are also qualified. For example, they stated:

> They [the experimental studies satisfying the criteria identified above] demonstrate that strategic behavior occurs much less often than standard utility maximization assumptions would predict, except where the person is assured that he will get the good no matter what he says he will pay. Even under this condition, free-riding occurs far less than most economists would predict.

Id. at 139-40. The qualification also arises in a footnote where they identify some other studies that they describe as deliberately structuring conditions to motivate people to behave strategically, and MC concede that: "We believe not only that such conditions exist, but that they are largely avoidable." *Id.* at 140 n.21. Because MC's conclusion is not a general verdict on free-riding, it is possible to have two mutually exclusive and seemingly contradictory readings of the experimental literature.

On the one hand, reviewers of the MC summary can criticize the studies they selected, cite old and new evidence not included in their review, and observe that free-riding can be important. Yet MC can agree but argue that in practice, it will not be important because the evidence is "clear enough" to prepare descriptions of CVM questions that control strategic incentives.

A. Background

An examination of the conventional model of consumer behavior reveals how little of what we accept as "verified" by real economic commitments is actually testable without a complete record of total household expenditures.[16] Even in that case, what we can test is conditioned upon important assumptions usually arising from the data available in each application. Moreover, if we examine the record for assumed consumer behavior in these terms, it does not fare well. Deaton's summary of the literature concluded that:

> Although there is some variation in results through different data sets, different approximating functions, different estimation and testing strategies, and different commodity disaggregation, there is a good deal of accumulated evidence rejecting the restrictions [that is, homogeneity of degree zero in prices and income and symmetry of the Slutsky substitution matrix]. The evidence is strongest for [rejecting] homogeneity, with less (or perhaps no) evidence against symmetry over and above the restrictions embodied in homogeneity. *Clearly, for any one model, it is impossible to separate failure of the model from failure of the underlying theory.* [17]

Overall, most of the experience of tests of the hypotheses implied by conventional demand theory even that using data for individual households, indicates that the test results are sensitive to the definition of the commodities involved (and the implied level aggregation); the model specifications used; the treatment of durables; and the assumed separability of labor/leisure from commodity consumption decisions. Despite the largely negative support for demand theory from these system-wide tests, most economists would likely regard the general insights provided by consumer demand theory to have been verified by empirical experience.[18] Understanding the reason for this

16. Indeed, applying a weaker criterion using the axioms of revealed preference to evaluate demand theory, Varian observed that "The sad fact of the matter is that the restrictions [revealed preference derived from utility maximization subject to a budget constraint] only apply when we have observed the entire choice set. Hence, normal sorts of tests of consistency of observed choice must be interpreted instead as tests for separability of the observed choices from other variables in the utility function rather than test of maximization *per se*." H. Varian, *Revealed Preference With a Subset of Goods* 46 *J. Econ. Theory* 179, 184 (1988).

17. A. Deaton, *Demand Analysis, in 3 Handbook of Econometrics* 1767, 1791 (Z. Griliches & M. Intrilligator eds., 1986) (emphasis added).

18. Kiefer, using a micro data set for 3,000 Belgium households, is somewhat more encouraging but not definitive. A test of the joint hypothesis of homogeneity and symmetry would not be rejected at the five percent level. Homogeneity alone would also not be rejected. However, both decisions are sensitive to the p-value selected for the test. Both the joint (homogeneity and symmetry) hypotheses and the homogeneity hypothesis on its own would be rejected at a ten percent level. N. Kiefer, *Microeconomic Evidence on the Neoclassical Model of Demand*, 25 *J. Econometrics* 285 (1984).

conclusion is important because it may provide a practical basis for evaluating the consistency of CVM results with the assumptions of theory.

There are at least three sources of empirical evidence supporting demand theory as it is most often applied with simple, single good demand models. They include:

(1) weight-of-the-evidence judgments based on the response of demand to price and income change;
(2) plausible demographic and attitudinal linkages; and
(3) credible predictive performance for policy uses.

The first of these is largely what the description implies—consistent findings of negative effects for the good's price and positive ones for income in the demand equation. Moreover, for broad commodity aggregates, there is approximately consistent classification of commodities as between luxury (that is, with income elasticities appreciably larger than unity) and staples (income elasticities less than unity). Price elasticities for the goods studied would probably be regarded as somewhat more variable. Nonetheless, in most circumstances it has been possible *a priori* to develop classifications for commodities into groups based on inelastic and elastic responses to price, and then to confirm these relative size expectations implied by these groupings with the independently estimated demand results.

With greater availability of micro-data, there has been more attention to demographic and attitudinal variables as proxy measures for tastes in describing demand. Thus there has been increasing evidence available for the second type of confirmation. In those cases where *a priori* expectations could be formed for these types of influences, the empirical evidence generally supports expectations. While these are confirming findings, it is also important to acknowledge that the results are usually confined to agreement of the sign of the specific variables with prior expectations.[19]

Finally, the estimates of price and income elasticities from simple demand models have generally been regarded as offering plausible predictions of demand impacts for policy initiatives that cause either price or income changes. As with the first two criteria there have been no specific studies that assembled systematic evidence supporting these

19. It should be noted that this conclusion is not the result of a systematic evaluation of the existing empirical literature. For three examples where demographic variables have been introduced into demand systems, see R. Pollak & T. Wales "Demographic Variables in Demand Analysis," Econometrica, 1533-1551 (1981), R. Barnes & R. Gillingham, "Demographic Effects in Demand Analysis: Estimation of the Quadratic Expenditure Function Using Microdata," Rev. Econ. Stat. 591-611 (1984), and with a micro level panel, see K.A. Mork & V.K. Smith, Testing the Life-Cycle Hypothesis with a Norwegian Household Panel, J. of Bus. & Econ. Stat. 7, (1989).

conclusions.[20] Rather it is a judgment based on numerous articles reporting demand applications, textbooks citing specific demand studies as examples, demand analyses used in antitrust and other forms of litigation, demand elasticity estimates role in the setting of prices for regulated firms, and the practice of using elasticity estimates for policy forecasts.

B. Consistency of CVM with 'Real Economic Commitments'

There has been no systematic effort to document uses of CVM estimates. Instead, a "folklore" has evolved based largely on personal evaluations of the empirical record. Most CVM practitioners believe the record shows that CVM credibly estimates people's willingness to pay (WTP), provided the specific application followed "established practices". Of course, the "established practice" is continuously changing as experience with applications involving more difficult environmental commodities increases. Furthermore, because CVM respondents must be given a choice in understandable and realistic terms, the evaluation of its success has a number of subjective elements.

Consider each of the three approaches that have been used to evaluate empirical demand studies with marketed commodities as a potential basis for judging whether CVM responses are consistent with economic behavior (CH's first question). The first approach with marketed goods focused on observed responses to prices and income and their consistency with what economic theory would suggest. Providing a systematic response to this question requires that environmental resources' services be classified into groups of commodities (much as marketed goods have been classified), a set of methods for measuring the amounts of each service be specified, and the available CVM evidence on WTP for each category be reviewed to determine whether it is consistent with the general properties implied by economic theory. This would be an enormous task, well beyond the scope of this paper. However, there have been some small steps taken in this direction. One of these involves a meta analysis (i.e. a statistical summary) of travel cost and contingent valuation studies by Walsh, Johnson and McKean.[21] Their analysis summarized estimates of the consumer surplus per day of recreational activity from each type of study and found that the travel cost and CVM summaries each yield distinctive values for different

20. Systematic statistical summaries have been confined to the marketing literature. One example closely aligned with conventional economic models for consumer demand involves modeling a demand for a firm's branded product using sales or market share data. See G. Tellis, *The Price Elasticity of Selective Demand: A Meta-Analysis of Econometric Models of Sales*, 25 J. Marketing Res. 331 (1988).

21. See R. Walsh et al., *Nonmarket Values from Two Decades of Research on Recreation Demand*, 5 Advances in Applied Micro Economics 167-193 (A. Link & V. Smith eds., 1990) (Consult Table 3).

types of recreation. Their conclusions were based on measuring statistically significant coefficients for the qualitative variables used to take account of the different types of recreation (including salt water and anadromous fishing, big game hunting, and waterfowl hunting). Of course, one might argue that these activities are familiar and what is being measured resembles closely a private commodity.

Nonetheless, there is also some evidence that CVM estimates of WTP for a less familiar and more difficult to measure (and describe) commodity are also consistent with what economic theory would imply. Two sets of evidence provide the basis for this preliminary judgment. One follows from the important initial effort by Chestnut and Rowe[22] to summarize and compare people's WTP for visibility improvements at national parks. The second is an extension to their work that Laura Osborne and I have currently underway.[23]

Using twenty estimates derived from four CVM studies for improvements in the visibility at recreation sites, they adjusted the estimates to constant dollars and fit a response surface relating the willingness to pay (WTP) per visitor party per day (in 1988 dollars) to the logarithm of the ratio of the revised relative to the initial visibility (in miles) together with this same variable interacted with a qualitative variable identifying whether the studies involved Eastern National Park sites. Both variables were found to be significant determinants of the deflated WTP estimates. Laura Osborne and I confirmed and extended their analysis by expanding the set of estimates included in the summary, controlling for the how visibility change was presented, and adjusting for the non-spherical nature of the error structure. The columns in Table 1 report some of the preliminary results from our analysis. Our summary is confined to willingness-to-pay estimates for improving (or avoiding deterioration in) visibility conditions at recreation sites. The studies were limited to those that present the change in visibility conditions as taking place with certainty. They generally used photographs that depict how specific vistas would change as different air quality conditions alter the visible range.

Our analysis differs from Chestnut and Rowe in several ways. The sample was composed using all available estimates from each CVM analysis. Our measure for the visibility change approximates the proportionate change in the miles of visible range. The lowest level of visibility is treated as the base. We include a separate qualitative variable

22. L. Chestnut & R. Rowe, Economic Valuation of Changes in Visibility: A State of the Science Assessment for NAPAP, *in Acidic Deposition: State of Science and Technology,* Report No. 27 *Methods for Valuing Acidic Deposition and Air Pollution Effects* (1990).

23. V. Smith & L. Osborne, "Do Contingent Valuation Estimates Pass a 'Scope' Test? A Preliminary Meta Analysis" paper presented at American Economic Association Meetings, Boston, Mass. January 1994.

(labeled as the direction of change) to indicate whether the question posed improvement from a deteriorated state or avoiding a deteriorated state (that is, compensating versus equivalent measures of the Hicksian surplus associated with the willingness to pay definition).

A second difference arises in the consideration given to other potential determinants of CVM estimates of WTP. We included qualitative variables to adjust for how the valuation was asked--as an entrance fee or a charge in the next visit. Third, models were estimated with sub-samples that deleted specific studies or locations to evaluate whether the overall conclusions were influenced by the composition of the sample. Finally, the estimated standard errors used in testing the relationship between CVM estimates of the willingness to pay for visibility improvements used Huber's adjustment for the non-spherical errors.[24]

A number of models and tests were evaluated with these and other CVM analyses of visibility changes. The principal issue of interest here is whether larger improvements in visibility lead to larger monthly WTP estimates (in constant 1990 dollars). Our summary of CVM estimates across these studies is clear in its support for a positive relationship. Models (1) and (5) report the ordinary least squares (OLS) estimates of WTP for improved visibility using the full sample with a partial versus a complete specification.

To evaluate the effects of sample composition three sub-samples were evaluated. Equation (3) deletes the estimates for Eastern sites; (4) deletes the Decision Focus pilot study for small visibility changes at the Grand Canyon;[25] and (2) deletes estimates from both the Eastern sites and the Decision Focus pilot study. None of these changes affect the statistical significance or relative size of the estimated coefficient for visibility improvements.

Of course, finding consistent relationships between estimates of WTP and different recreational activities or between WTP and one aspect of visibility improvements at recreation sites like the Grand Canyon does not imply that the CVM estimate of the willingness to pay in each case is accurate. Moreover, it does not imply that this level of consistency with theory would be found in comparable summaries with other types of environmental resources. In many situations it may

24. The Huber adjustment used here is for heteroscedasticity and parallels the White proposal for a consistent estimate of OLS estimates' covariance structure in these cases see: P. Huber, *The Behavior of Maximum Likelihood Estimates under Non-Standard Conditions*, *in* Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 221 (1967); H. White, *A Heteroscedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity*, 28 Econometrica 817 (1980).

25. W. Balson et al., Development and Design of a Contingent Valuation Survey for Measuring the Public's Value for Visibility Improvements at the Grand Canyon National Park (Sep. 1990) (Draft Report by Decision Focus, Inc., Los Altos, CA).

be difficult to formulate a single quantitative scale for measuring changes in the amount of the commodity that is offered. Nonetheless, it is important to recognize that this type of consistent record replicated across all environmental commodities would exceed the level of systematic analysis and testing that underlies our acceptance of conventional demand studies for marketed commodities.

Turning to the second source approach for developing confirmatory evidence (that has been used for demand models based on marketed goods),—plausible relationships between the demographic and attitudinal variables and demand or valuation responses—here the record of CVM may be more detailed than with marketed commodities. This is at least partially due to the need for primary data collection to undertake CVM studies. While initial applications did not collect attitude and perception measures, most recent studies have collected this information and successfully used it in explaining people's responses to CVM questions involving risk, air and surface water quality, groundwater contamination, pesticide residues on food, aesthetic features of the landscape, drinking water quality and availability, and numerous other applications.[26]

---

26. For examples of these types of studies by resource or pollution type, see: R. Carson et al., A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill, report to the Attorney General of the State of Alaska (Nov. 10, 1992) (Natural Resource Damage Assessment, Inc.) (one of the most extensive large scale contingent valuation studies based on in-person, state-of-the-art interviews done for the State of Alaska); V. Smith & W. Desvousges, *Risk Communication and the Value of Information: Radon as a Case Study*, 82 Rev. Econ. Stat. 137 (1990) (environment risk); W. Evans & W. Viscusi, *Estimation of State Dependent Utility Functions Using Survey Data*, 83 Rev. Econ. Stat. 94 (1991) (also environmental risk); P. Jakus & V. Smith, Measuring Use and Nonuse Values for Landscape Amenities: The Case of Gypsy Moth Control (Jan. 1992) (paper presented at Association of Environmental and Resource Economists, New Orleans, LA) (aesthetic dimensions of landscape); Y. Eom & V. Smith, "Calibrated NonMarket Valuation" January 1994 (unpublished paper, Resource and Environmental Economics Program, North Carolina State University); E. van Ravenswaay & J. Hoehn, Contingent Valuation and Food Safety: The Case of Pesticide Residues in Food (1991) (Staff Paper No 91-13, Dept. of Agricultural Economics, Michigan State University); G. McClelland et al., Methods for Measuring Nonuse Values: A Contingent Valuation Study of Groundwater Cleanup (Oct. 1992) (on file at U.S. Environmental Protection Agency, and the Center for Economic Analysis, University of Colorado); J. Powell, The Value of Groundwater Protection: Measurement of Willingness to Pay Information and Its Utilization by Local Government Decisionmakers (1991) (unpublished Ph.D. Thesis, Dept. of Agricultural Economics, Cornell University); R. Carson & R. Mitchell, Economic Value of Reliable Water Supplies for Residential Water Users in the State Water Project Service Area (1987) (report prepared for the Metropolitan Water District of Southern California, Washington, D.C.); D. Whittington et al., *Giving Respondents Time to Think in Contingent Valuation Studies: A Developing Country Application*, 25 J. Envtl. Econ. Mgmt. 205 (1992). A detailed bibliography of over 1,600 references related to contingent valuation can be found in A Bibliography of Contingent Valuation Studies and Papers (1994) (published by Natural Resource Damage Assessment, Inc., La Jolla, CA).

On the last element cited as a component used in judgments of conventional demand models—successful policy uses—there has been little scope to use CVM estimates to predict behavioral outcomes for policy purposes and then evaluate the outcomes. For the most part, CVM estimates associated with environmental resources have been used in benefit-cost analyses (or more recently, as CH note, in damage assessments) where there is little or no opportunity to observe what has been estimated or predicted.

Overall, considering CH's first question, it would be possible to apply the same types of standards used in evaluating conventional demand studies to CVM estimates. To date, there have been few systematic efforts to undertake this task. Based on some preliminary results from meta analyses of CVM studies for valuing visibility improvements, it appears that CVM's WTP estimates are broadly consistent with the implications that can be developed from economic theory.

CH's second question in evaluating whether CVM represent real economic commitments concerns whether people actually pay what they state. Here CH describe some of the past evidence and their own (with Rutström) findings using simulated markets (i.e. markets constructed under experimental conditions for real commodities). In these situations, a reward system is not used to induce subjects' preferences. Instead, the analyst must rely on structuring the incentives so that truth-telling is the incentive compatible response. Because it may take time for subjects to have confidence that the incentives actually work this way, the experiment must either: (a) provide them opportunities to acquire experience with the institutions (or rules) providing these incentives, or (b) "teach" them what they would have learned from that experience. The analyst's knowledge of actual preferences (and therefore, real economic commitments) is conditional upon how successfully these tasks are accomplished. Of course, this is true of any test. There are always maintained hypotheses. This is not my point. Rather, my concern is whether our ability to adequately exercise these controls and thereby establish conditions comparable to an experienced choice is itself influenced by whether private and public good is involved in the experiment.

To illustrate my point, consider first the strategies used to construct these types of simulated markets with private goods. Two types have been used—auctions and direct sales. In the first case, an incentive compatible process is designed. Each individual may be uncertain about whether he (or she) will acquire a specific commodity. The rules are designed to induce truthful responses.[27] Direct sales are also a pos-

27. Many of the experimental studies allow participants to engage in "practice" rounds with "instruction" to assure that they understand what a rational response to the incentives implies for their behavior. Even with practice rounds, differences in the behavior of experienced versus inexperienced subjects can be pronounced.

sible method for comparing hypothetical and actual responses with private commodities. However, using this approach to recover valuation information requires that the analyst supply prior information about how participants' decisions relate to their demands or preferences (that is, demand functions must be specified).

Implementing both approaches inevitably involves analyst judgment. In the case of auctions, one might ask how the experimenter gives subjects sufficient experience to understand the implications of the incentives without providing what might be interpreted as clues about other participants' responses or values. Does the experience transfer across different types of goods? This argument parallels the selection process Mitchell Carson report they used in evaluating the free-riding experiments (see note #15). That is, there are circumstances where we can control the rules (and experience) so the experiment is credible. There may be others (involving public goods) where the change in rules is so great as to reduce the control and therefore the confidence in the experiments. Reduced variation in the bids in actual versus hypothetical markets does not in itself mean that these are more accurate reflections of their true values. It may mean people have learned what they believed (from their past experience or instruction) to be the "correct" response. When we consider direct sales, the process of specifying and estimating demand (or other behavioral models) can lead to large variations in the estimated values.[28] This is not simply a criticism

---

28. In J. Hausman & G. Leonard, Contingent Valuation and the Value of Marketed Commodities (July 24, 1992) (unpublished paper, Dept. of Economics, Massachusetts Institute of Technology) the re-analysis of the strawberry experiment, see M. Dickie et al., *Market Transactions and Hypothetical Demand Data: A Comparative Study*, 82 J. Am. Stat. Ass'n 69 (1987)(hereafter known as DFG for this discussion), found the conclusions about a correspondence between actual sales and purchase intentions were sensitive to how the data were treated in the estimation. Indeed, replicating the Hausman-Leonard preferred estimator with a count data Poisson model using the original DFG data indicates the key parameter estimates (from an economic perspective) are more sensitive to the model specification than to the use of an actual versus a hypothetical sales context. For the specifications reported here, the sensitivity is greatest for actual sales and is not as pronounced using the hypothetical sales data. This raises the basic issue-which model-original DFG or the reduced specification to use in any comparisons between them.

Applying a less mechanical perspective and comparing the range of parameter estimates across specifications, there is a remarkably close correspondence between the demand models estimated based on real and hypothetical sales. The following table illustrates this pattern for the parameters of price and income, using the Poisson count estimator with linear in independent variables models:

	Actual Sale		Hypothetical Sale	
Model Specification	Price	Income	Price	Income
Original DFG specification-	1.958	32×10^{-3}	-1.904	$.16 \times 10^{-3}$
	(-3.42)	(2.24)	(-4.28)	(1.19)
Deleting Dummy Variables	-1.941	33×10^{-3}	-2.042	$.27 \times 10^{-3}$
for Income Stratum	(-3.47)	(2.45)	(-4.56)	(2.33)
Price, Income, Household	-1.375	$.10 \times 10^{-3}$	-1.937	$.27 \times 10^{-3}$
size, interviewer teams	(-3.00)	(1.09)	(-4.91)	(2.71)

(numbers in parentheses are Z statistics for the null hypothesis of no association) of contingent valuation data. It is a reflection of the role of analyst judgment in the implementation of most microeconomic models.

As this research has changed the commodity from a private to a public good, there has been a subtle but important change in what was offered to people. It has not been a change in a public good, but instead a lottery with some individuals making payments (in the "real" component of these experiments) that did not assure them of any change in the public goods involved. This feature characterizes work by Seip and Strand, Kealy et al., and Duffield and Patterson.[29] Neither the real nor the CVM versions of these surveys indicate how their respective lotteries would be resolved. As a consequence, it is impossible to evaluate how these mechanisms might be expected to perform using a simple public good. The commodity that is offered and understood by people could well be different for each participant, depending on how each person perceives these lotteries. There are theoretical and design-related reasons for questioning many of the earlier experiments supporting criticisms of the expected utility (EU) framework. It seems that most of the experiments testing preference reversals (and other violations of EU behavior) can themselves be questioned. Some involve compound lotteries (constructed to avoid income effects; reduce experimental costs; or meet other objectives). In these cases, Holt [1986] has

---

These conclusions contrast with the Hausman and Leonard judgments which seem to imply the observed differences are inconsistent with applied econometric experience across model specifications in other contexts involving actual decisions and because of this variability, imply we should reject contingent valuation. In contrast with their conclusions, this level of sensitivity should not be surprising given the sample sizes involved, the use of micro data, and the presence of a number of zero consumption choices.

Indeed, some time ago in another context Berndt, see E. Berndt, *Reconciling Alternative Estimates of the Elasticity of Substitution*, 58 Rev. Econ. Stat. 59 (1976), used the variation in assumptions about the treatment of capital measurement and the definition of rental prices for capital to explain the wide differences in estimates of the elasticities of substitution between labor and capital. More recently, Hazilla and Kopp, see M. Hazilla & R. Kopp, *Systematic Effects of Capital Service Price Definition on Perceptions of Input Substitution*, 4 J. Bus. Econ. Stat. 209 (1986), performed a more detailed empirical evaluation of how these assumptions influence estimates of the elasticities of substitution using a common data set for 36 sectors.

A similar message emerges for studies involving the indirect methods for nonmarket evaluation. For example, my meta analysis of travel cost recreation demand estimates considering both the consumer surplus per unit of use and the price elasticity of demand with Y. Kaoru, see V. Smith & Y. Kaoru, *Signals or Noise? Explaining the Variation in Recreation Benefit Estimates*, 72 Am. J. Agric. Econ. 419 (1990); V. Smith & Y. Kaoru, *What have We Learned Since Hotelling's Letter: A Meta Analysis*, 32 Econ. Letters 267 (1990), found the estimates were sensitive to modeling assumptions comparable to those raised by Hausman and Leonard.

29. K. Seip & J. Strand, *Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payments*, 2 Envtl. Resource Econ. 91 (1992); M. Kealy et al., *Reliability and Predictive Validity of Contingent Valuation: Does the Nature of the Goods Matter?*, 19 J. Envtl. Econ. Mngmt. 244 (1990); J. Duffield & D. Patterson, Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers (Jan. 1992).

noted that their tests of transitivity rely on a strong form of the independence axiom. This axiom assures the lottery choice and selling price elicitation will be separable.[30] As a result, the test of transitivity should not be separated from the maintained assumption of the independence axiom, because the latter is central to the test. Once this connection is recognized, we cannot attribute rejections of transitivity to that feature of the experiment. It may be the maintained assumption of independence that is inconsistent with behavior.

Equally important, Harrison has demonstrated that the nature of the financial incentives influences the ability (i.e. the power) of experiments in detecting violations of EU behavior.[31] The absence of distinctive incentives (as between the null and alternative hypotheses) was an important factor in his explanation of violations of the expected utility model—whether preference reversals, the Allais paradox, or prospect theory.

This digression is important to my argument for two reasons. The framing of all public goods involved in past experimental comparisons of real and hypothetical valuation tasks has actually offered respondents lotteries. Interpreting their findings on the consistency between CVM and actual responses (or lack of it) requires that we describe how people perceived the lotteries and interpreted them in formulating their contingent and actual responses.

Equally important, the past twenty years of testing for preference reversals reveals how extraordinarily difficult it would be to satisfy the criteria that CH have recommended for evaluating CVM. Judging whether responses are consistent with the predictions of economic models and if the payments would be made, each requires maintained assumptions to implement. With them, we find a cascading set of qualifications that must themselves be evaluated as part of our interpretation of their relevance for the performance of CVM.

IV. EXPERIMENTAL ECONOMICS AND CVM

CH use two types of experimental evidence in their evaluation of contingent valuation. The first is associated with "conventional" economic experiments. These are generally assumed to involve situations where the participants' preferences are controlled by the analyst. This control is accomplished by defining a payoff or reward scheme in monetary units that relates to the amount of the experimental commodity each participant receives or sells.

30. C. Holt, *Preference Reversals and the Independence Axiom*, 76 Am. Econ. Rev. 508 (1986).

31. See G. Harrison, Expected Utility Theory and the Experimentalists (Sept. 1990) (Dept. of Economics working paper B-90-04, University of South Carolina); G. Harrison, Theory and Misbehavior of First-Price Auctions: Reply, 82 Am. Econ. Rev. 1426 (1992).

The second type of experiment involves situations where the analyst does not control participants' preferences. Demand revealing mechanisms are assumed to induce each participant to reveal his (or her) true value for the commodities involved. Labelled as experiments involving "homegrown values" in contrast to the "induced values" of the conventional approach, these studies are more diverse in the control exercised over experimental conditions.

When the conventional experiments have involved public goods, they have used one of two types of institutions to determine how the amount of the public good is determined from participant interactions—a Smith (or unanimity) auction and a voluntary contribution framework.[32] As CH note, there is little to be learned from either decision rule about how contingent valuation surveys will perform. Because of this incompatibility between the rules governing each approach to understanding people's behavior, CH reconsider the evidence from conventional experiments on free-riding and use experiments involving "homegrown values" to comment on whether CVM accurately measures people's willingness to pay.

A. Free-riding

CH's discussion of free-riding is organized to assess whether the Mitchell and Carson summary of the evidence accurately represents current views on strategic behavior. CH (and Plott in a separate critique of Mitchell and Carson) argue it does not, observing that the mixed evidence from both Smith auctions and voluntary contribution schemes indicate that there are circumstances when people undertake strategic behavior.[33]

A central question that is not addressed in their summary (or in Plott's evaluation) concerns whether components of this mixed record are more likely to be relevant to the circumstances one might encounter in applying CVM to measure natural resource damages. One feature of the design of these experiments may be especially relevant. It concerns the commodity specification.

Natural resource damage assessment seeks to use nonmarket valuation in two different tasks—valuing the services lost to measure interim lost values (including both use and nonuse values or "passive use") and estimating future losses that arise because a restoration plan

---

32. V. Smith, *The Principle of Unanimity and Voluntary Consent in Social Choice*, 85 J. Pol. Econ. 1125 (1977).

33. C. R. Plott, Contingent Valuation Methods as Applied to Nonuse of Natural Resources: Evidence from Experiments (July 21, 1992) (testimony for National Oceanic and Atmospheric Administration Contingent Valuation Panel, Division of Humanities and Social Sciences, California Institute of Technology).

does not return the resource to baseline conditions. In both situations the commodity specification seems likely to involve services that are not readily perceived as divisible.[34] This is relevant to the experimental literature because most experiments unnecessarily introduce a perception of divisibility and a unit "price" to participants.

This approach has been used in experiments involving the unanimity and voluntary contribution frameworks. Experimental instructions state that the good is available at a fixed marginal cost (or price) whose value is known to participants.[35] This structure treats the public good as divisible. While it is consistent with the usual conditions for efficient levels of provision for public goods, it does not parallel the circumstances generally encountered in the contingent valuation analyses for natural resource damage assessments. In these cases, the definition of the commodity and how it would be impacted by any restoration activities are central questions in framing a CVM analysis.

Two sets of experiments have investigated situations where divisibility and a constant unit cost were not part of the design. The Schneider-Pommerehne study (originally cited by Mitchell and Carson) avoided the unit cost and divisibility issue by stating a total cost and by offering an examination copy of a textbook.[36] However, selecting the text as the commodity also converts this study from an induced preference to a "homegrown value" experiment.

The second study involving some experiments without unit cost and divisibility issues (by Marwell and Ames) is also not an ideal basis for a judgment.[37] Their evaluation of the effects of divisibility involved two sets of incoming undergraduates. One group was given a schedule of payments in discrete intervals, where positive provision and individual payoff required a group contribution of at least 2000 to-

34. Indeed, an important implication of the Exxon-sponsored CVM research is that surveys relying on artificial definitions of divisible changes in commodities that people do not perceive in that way are unlikely to provide plausible estimates of their willingness to pay.

35. Examples of these types of instructions include most of the studies evaluating public goods that were cited by CH. See, e.g., J. Banks et al., *An Experimental Analysis of Unanimity in Public Goods Provision Mechanisms*, 40 Rev. Econ. Stud. 301 (1988); R. Isaac et al., *Public Goods Provision in an Experimental Environment*, 26 J. Pub. Econ. 51 (1985); G. Harrison & J. Hirshleifer, *An Experimental Evaluation of Weakest Link/Best Shot Models of Public Goods*, 97 J. Pol. Econ. 201 (1989). But R. Isaac & J. Walker, *Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism*, 103 Q. J. Econ. 179 (1988), appears to be a notable exception to this framing of constant marginal cost, divisible public goods in many of the laboratory public goods experiments. It is impossible to separate the effects of this specification from other changes in determining the overall implications of the change for strategic behavior.

36. F. Schneider & W. Pommerehne, *Free Riding and Collective Action: An Experience in Public Microeconomics*, 96 Q. J. Econ. 689 (1981).

37. G. Marwell & R. Ames, *Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods, IV*, 15 J. Pub. Econ. 295 (1981).

kens (each person received 225 tokens to be divided between a private and a group or public good). The indivisible group were given the same instructions, but told the payoff "... had to be spent on a group project. They could choose anything they wanted on which to spend the money, such as a party or a hi-fi for their floor, so long as there was something purchased collectively."[38]

They report dramatic differences in the percentage of resources allocated to the group or public good. Forty-three percent was invested by the subjects receiving a divisible commodity with private returns based on the provision point restriction and 84 percent for the group receiving the indivisible case.

Unfortunately, by failing to describe a decision rule for selecting the group project, they have offered what could easily be regarded as a compound lottery with greatly varying perceptions across participants. Thus, while this experiment provides evidence consistent with a conclusion that there is limited scope for free-riding when the commodities are indivisible public goods, there are good reasons to question whether divisibility alone is responsible for the outcome. Moreover, it is not clear this experiment satisfies Harrison's criteria for distinctive incentives, allowing analysts to effectively isolate the behavior implied by null and alternative hypotheses.

Overall then, this review of the literature from conventional experiments agrees with CH that interpreting Mitchell and Carson's (MC) conclusions as a general precis of all experiments investigating free-riding would not adequately represent the diversity of current evidence on the free-riding hypothesis. However, it is also important to note that MC did not claim to be offering such a general evaluation. Instead, they used experimental evidence available at the time to argue it was possible to frame problems involving public goods so that free-riding responses would be minimized. CH do not address this issue. Plott's criticism of the Mitchell-Carson summary also avoids the question of whether we know enough to "frame our way out of free-riding problems." Instead, he suggests that free-riding in conventional settings can be serious, noting that:

> The first instinct of people faced with a public goods/free-riding/prisoner's dilemma situation is to produce outward signs of cooperative behavior. In the context of public goods in which people face incentives, they tend to make first period (of a several period process) voluntary contributions. The level falls short of that suggested in the quotation from Mitchell and Carson but it could be as high as 25 percent to 50 percent of an objectively known level of incentives ... Cooperation usually erodes rapidly in subsequent periods

38. *Id.* at 306.

choices. While complete free-riding seldom occurs, the level of contributions are near the Nash equilibrium levels with a bias toward the cooperative options. The nature and speed of the decay process is a function of the structure of the payoff. Environments with a high marginal payoff to cooperation decay more slowly.[39]

To the extent the propensity to undertake free-riding is influenced by divisibility in the public good, as might be implied by Marwell and Ames, then the available experiments cited here and in CH do not allow a judgment on whether free-riding will be present in these situations. Moreover, there have been no experiments to evaluate the degree to which analysts can avoid free-riding with variations in the framing used to elicit respondents' WTP.

B. Homegrown Values and CVM

The second type of experiment discussed in CH's evaluation of CVM concerns the use of simulated markets. However, in contrast to Bishop and Heberlein or Dickie, Fisher and Gerking, Cummings et al. use a "laboratory" setting.[40] In an attempt to avoid the influence of individual preferences and constraints, some of their experiments used "paired" hypothetical and real sales. Two commodities—an electric juicer and a box of chocolate truffles—were considered for these paired (or "in sample") experiments. That is, separate groups of respondents for each commodity are offered hypothetical and actual purchase decisions.[41] Their evaluation of the performance of CVM is based on the fraction of stated choices in comparison with the actual purchases.

Four issues are relevant to these comparisons. First, the wording of their hypothetical question (in the "in-sample" experiments) emphasized before offering each commodity that "We are not actually offering you the opportunity to buy the juicer." Then participants were asked whether they would pay at a stated price. Those subjects participating in the in-sample experiments proceeded to the "real sale" component and were told "We would now like to give you the opportunity to actually buy the juicer (the chocolate truffles) at the price of (the commodity involved in the experiment). **Note that you do not have**

39. Plott, *supra* note 33, at 12-13.

40. R. Bishop & T. Heberlein, *Measuring Values of Extra Market Goods: Are Indirect Measures Biased?*, 61 Am. J. Agric. Econ. 926 (1979); R. Bishop & T. Heberlein, *Does Contingent Valuation Work?*, *in* R. Cummings et al., *Valuing Environmental Commodities* (1986); Dickie et al., *supra* note 28.

41. Based on the descriptions in R. Cummings et al., Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible? (Oct. 1992) (working paper B-92-12 Division of Research, College of Business Administration, University of South Carolina), the experiments for the calculators involved independent samples of respondents for the hypothetical and actual sales.

to say the same thing that you did on the previous question." (emphasis added)[42] The process then proceeded to ask if each individual would purchase the relevant commodity at the same stated price.

While one must distinguish between hypothetical and real sales as part of eliciting participants' decisions, all of the literature in CVM suggests that the wording can matter. It is impossible to evaluate whether it did in the cases they considered. Research available after their experiments were completed suggests that investigation of the reasons why respondents made particular decisions can be important. For example, Carson et al., used this strategy to design sensitivity tests for their CVM estimates of a household's willingness to pay to prevent another Exxon Valdez type oil spill.[43] Respondents to the CH and R experiments could have interpreted the first question as requesting a judgment about whether the juicer was worth a certain amount, not as a purchase intention. In these cases, it is conceivable that those participants owning juicers (or with no preference for fresh squeezed juice) could answer "yes" and yet not purchase the juicer. Similar explanations could be offered for the chocolates.

A second concern arises with the metric used to evaluate CVM. CH and R rely on the fractions stating a purchase choice. Their testing strategy involves comparing the aggregate demand functions implied by individuals' stated and actual purchase decisions at a point.[44] By using in-sample experiments, they control for differences in the mix of

42. *Id.* at 41. This statement was not included in their experiments involving independent subjects, where solar calculators were sold and in the reference experiments involving only one treatment for the juicer and chocolates.

43. See R. Carson et al., A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill (Nov. 10, 1992) (report to Attorney General of the State of Alaska, published by NRDA Inc., San Diego, CA).

As I note in the last section, the NOAA Contingent Valuation Panel also highlighted the need to include these follow up questions along with checks on understanding and acceptance.

There is an interesting parallel in proposals to use these follow up questions in screening CVM responses or improving statistical models for them, and the proposal by M. Blackburn et al., Statistical Bias Functions and Informative Hypothetical Surveys (Feb. 1992) (Economics Working Paper B90-02, Div. of Research, College of Businss Administration, University of South Carolina), to model the inconsistencies observed between contingent and actual choices for the paired samples.

In their case, they find that using a statistical model for the inconsistencies and conclude on a cautiously positive note that some calibration (analogous to what is done in marketing research) may be possible. The former analysis could well yield possible calibrating adjustments to CVM responses.

44. There are a variety of maintained assumptions that would imply a test of consistency in individual demands could be conducted at an aggregate level, given an assumption of interior solutions at the individual level. For example, Lau's extension, see L. Lau, *A Note on the Fundamental Theorem of Exact Aggregation*, 9 Econ. Letters 119 (1982), of the Gorman aggregation conditions could be used to formulate consistent aggregate demand specifications, see, e.g., D. Jorgenson et al., Two Stage Budgeting and Exact Aggregation, 6 J. Bus. Econ. Stat. 313 (1988). The paired in-sample comparisons of CVM and actual purchases assure control over the demographic, preference related and other constraint related variables.

preferences, constraints, and other factors that might be argued to influence the aggregate quantity demanded across different samples.[45]

This does not evaluate how either source of data performs in estimating people's WTP for the commodities being sold (recall the two questions noted earlier that comprise the CH criteria for gauging real economic commitments). Moreover, it is not the way these data would be used in practice. Analysts would assume that people have different preferences and/or constraints. As such, they would seek a model that recognized the decision as one taking place at the extensive margin of choice and model the responses within that framework.[46]

Estimates of willingness to pay would be conditional upon the maintained assumptions used to describe respondents' circumstances and preferences. In such a micro context, the analyst's ability to recover a (WTP) estimate depends upon the extent to which the experiments recover estimates of respondents' responses to price, income, or both.[47]

At the aggregate level, the issue is different because we are not evaluating the ability to estimate WTP. The objective is to gauge the consistency of the quantity estimates provided by the two frameworks. This evaluation reveals nothing about the ability of either framework to estimate WTP. This follows from the discrete nature of each individual's choice. At the respondent level, each person's actual choice will indeed be based on his (or her) preferences and constraints. But suppose the individual already has a juicer or calculator. Price and income are irrelevant to the decision for actual choices. By contrast for hypothetical choice, the same respondent may well be inclined to report whether the object would be worth the stated price (perhaps because he had paid that amount or more in making a past purchase). Such responses could provide a credible estimate of WTP, but not of actual demand at the time of the survey because some of that demand had already resulted in purchases. To evaluate whether these differences are important requires both variation in prices and an analysis of demand at the individual level. Unfortunately, the CH and R experiments focused on tests using the proportion agreeing to purchase the commodity and therefore are addressing hypotheses that parallel the aggregate demand question. Such evaluations do not address the issue of how accurately each method measures individual WTP. Only

45. For two of the three commodities, the CH and R experiments include in-sample comparisons and out-of-sample comparisons to gauge the order effect of asking the hypothetical questions before actual purchase questions.

46. See S. Pudney, Modeling Individual Choice: The Econometrics of Corners, Kinks and Holes (1989) for an overview of the econometric issues involved with modeling individual choice at the extensive margin.

47. The differences depend upon how the cost of the item is posed, whether multiple unit purchases would be feasible, and the prior theoretical restrictions maintained in the specification and estimation of the models.

the calculator experiments included cases where there was some price variation across experimental groups.[48] It is possible to offer some preliminary evidence that suggests these issues may be important.

Consider two different uses of results from the CH and R calculator experiments.[49] They report seven experiments—four groups with hypothetical purchases and three real purchases, varying prices from $3 to $7. Using their aggregate data for either hypothetical or real with the six experiments,[50] we would conclude that there is **no relationship** between the price charged and the purchase decision. Moreover, after accounting for the price, the distinction between real and hypothetical purchases would **not** be judged to be a significant influence on the purchase decisions. Table 2 summarizes these estimates for the six experiment analyses based on a simple log-odds framework. [A no decision is coded as one and yes as zero in these estimates.]

Of course, analyzing these experiments in a pooled format at an aggregate level is likely to be considered unfair to the authors because the no test would be powerful with such small samples. A discrete choice model re-estimated using the individual responses addresses this issue and offers the prospect for evaluating whether the models provide a plausible basis for estimating individual WTP. Table 3 summarizes these results. The first column indicates that participants in the real sales had statistically a significantly greater probability of responding **no** in the actual sales, as Cummings and Harrison observed comparing the estimated sample proportions with their paired sample tests.

However, what their testing strategy misses is that neither of the samples, real or hypothetical, would have been accepted as a plausible basis for estimating the "typical" student's willingness to pay for a calculator. To develop estimates of the Hicksian consumer surplus requires that the analysis recover estimates of a price effect on the purchase decision. Neither estimated equation indicates price was a significant determinant of the participants' responses. Indeed, the only factor that was found to be a consistently significant determinant of stated and actual purchases—whether the participant owns a calculator—had an estimated influence that agreed with our *a priori* expecta-

48. The juicer experiments involved two prices ($8 and $12), but the results are summarized together and the data appendix includes only the results for the $8 experiments. See McKinley, *supra* note 43 (Table 1), for a summary of the composite results. When compared with CH and R, the rates for the higher price can be inferred.

49. This comparison would not be possible without complete access to the experimental data which CH and R kindly made available along with all the papers describing their findings. The calculator experiments were selected because they are the only ones with complete data available that varied the prices for the commodities.

50. This choice was selected because it parallels the tests used in their evaluation of actual versus hypothetical sales.

tions in both samples. Moreover, the estimated coefficients from the two models were not significantly different.[51]

Finally, there is an important lesson in four sets of evidence that have been used to gauge the performance of CVM.[52] As we move from commodities with private components and readily divisible services to those that are best treated as indivisible, purely public goods (and are the ones most likely to be encountered when CVM analyses are developed for natural resource damage assessments), we cannot assume CVM's performance (whether good or bad) will be independent of the commodities involved.

V. OBSERVED BEHAVIOR AND NONUSE VALUES

Revealed preference, as it has been applied in measuring the values people place on nonmarket public goods, is a misnomer. Behavior alone tells us nothing about how much people value a nonmarketed public good. To recover information about people's willingness to pay, we must be prepared to accept either unverifiable assumptions about preferences in the form of weak complementarity or to observe a site-specific link between that commodity and some other marketed good.[53] With this second approach, it is also necessary to assume those purchasing the marketed good observe the nonmarketed commodity in terms closely related to what can be measured at the site-specific level. This logic has provided the basis for the estimated use values associated with the indirect valuation methods.

As noted at the outset, Larson has proposed adapting another set of unverifiable restrictions on preferences to measure nonuse values.[54] Two aspects of his proposed method are important. First, it redefines nonuse values to be monetary measures of the Marshallian adjustments observed in a commodity that is known in advance to be Hicks-neutral to the environmental resource (that is, the Hicksian demand for the commodity does not change with changes in the environmental resource). By definition there is no use value underlying any

51. Because the estimated coefficients are normalized by the standard error for the model's error, our test involves evaluating whether the ratio of the coefficient to the standard error would be judged to be different between the two models.

52. The four sets of evidence evaluating CVM include: (a) comparisons of indirect methods with CVM, (b) use of simulated markets and CVM, (c) conventional laboratory experiments and CVM and (d) experiments based on homegrown values and CVM.

53. J. LaFrance, Incomplete Demand Systems, Weak Separability, and Weak Complementarity in Applied Welfare Analysis: Expanded Version, (Dec. 1992) (working paper No. 77, Dept. of Agricultural Economics, University of Arizona) develops a related argument in his recent critique of the value of weak complementarity in recovering estimates of the use value of nonmarketed resource.

54. It originated in Neill's, see J. Neill, *Another Theorem on Using Market Demands to Determine Willingness to Pay for Non-Traded Goods*, 15 J. Envtl. Econ. Mgmt. 224 (1988), proposed method for developing bounds for the marginal values of nonmarketed goods.

observed changes in the demand for this commodity in response to the environmental resources (as measured using its Hicksian demand). Under these conditions, an observed change in the Marshallian demand for this commodity in response to a change in the resource must be the result of nonuse value for the resource.

Second, and equally important, the method assumes that there exist measures of the quantity of the resource that serve equally well in indicating how its services contribute to use-related adjustments in commodities that are Hicksian substitutes or complements as well as for goods that are Hicks-neutral to the environmental resource.

Larson's approach is an ingenious re-definition of the types of preference-based relationships that can link marketed goods to nonmarketed environmental resources. While it does not capture situations where there is no link possible (that is, Hanemann's definition of nonuse values),[55] there is a more basic question to be raised. How does one identify the marketed good(s) that is (are) Hicks neutral to specific environmental resources? Hicksian responses are unobservable. As Bockstael and McConnell have suggested, we can only identify negative lower bounds for the value of a nonmarketed resource using Hicksian substitution and complementarity relationships and these convey no information.[56]

The point to be emphasized is that the strategy used in implementing conceptual methods for recovering valuation information matters to the estimates. As one attempts to extend these approaches from use to nonuse values, the results should not be regarded as reflecting "real economic commitments" simply because they are associated with real choices. They are equally likely to be the result of analysts' assumptions (or perhaps less kindly, "guesses") that must be treated as articles of faith to recover measures of use and nonuse values.

VI. GENUINE ECONOMIC VALUES, CVM, AND NATURAL RESOURCE DAMAGE ASSESSMENT

Tardiness occasionally can have its rewards. The final version of this paper was completed after the report of NOAA's Contingent Valuation Panel became available. As a result, it is possible to reflect on their recommendations as part of a summary of the issues raised by the Cummings and Harrison paper. Four issues will be considered: (1) the valuation tasks required by natural resource damage assess-

55. W. Hanemann, Three Approaches to Defining 'Existence' or Nonuse Values under Certainty (1988) (unpublished paper, Dept. of Agricultural and Resource Economics, University of California).

56. N. Bockstael & K. McConnell, Public Goods as Characteristics of Non-Market Commodities *Economic Journal*, Vol. 103, Sept 1993 1244-1257.

ments; (2) CVM information used by the *Ohio* court; (3) Cummings-Harrison-Rutström experiments and CVM; and finally, (4) the NOAA Panel's recommendation and the evidence on CVM's ability to measure "Genuine Economic Values".

The monetary values sought in a damage assessment are rather specific and depend upon the injury to the natural resources involved. The use of CVM in this context does not necessarily seek a market value for environmental resources involved in these assessments. Because these resources are assets that can provide both private and public good services, the values must reflect both types of services. Market outcomes for these natural assets would not necessarily reflect the values for public good services. Equally important, even if there existed markets for these assets (and public good services were unimportant to these assets' values to society), it is not clear the asset price would be the correct measure of damage. The law requires consideration of how the injuries affect the services from the asset and therefore how injury related changes in services would alter the asset's price. Thus, even in this most extreme example (without public good services), the valuation task depends on the injuries and their relationship to the private services provided by natural resources.[57]

Good practice in all fields requires that we periodically update conventional procedures as research increases our knowledge base. However, CH seem to imply more than this concern to learn from experience in describing current views about contingent valuation. For example, CH's stated motivation for re-evaluating our conclusions about CVM's performance suggests that: "the motivation for this paper derives from the apparent lack of information made available to the *Ohio* Court for its assessment of this critically important aspect of damage estimates obtained by applications of the CVM . . . It is our view that the three studies considered by the court in this regard do not constitute the state of the art of our understanding of the relationship between CVM values and values that reflect real economic commitments."[58]

While the ruling specifically identifies three studies, it also cites the DOI 301 Project review of "Techniques to Measure Damages to Natural Resources." One of the three cited was a review article by Cross[59] with fairly detailed summaries of CVM research. The DOI report specif-

57. The complexity of the valuation task is not avoided by drawing analogies to what would happen if there were markets for the resource's services. There is no reason to believe the services are perceived as available in readily divisible and separable units. These perceptions must be considered in defining the methods used to respond to injuries to natural resources and any proposed (for the purposes of CVM analysis) means of payment for these activities. See Valuing Natural Assets: The Economics of Natural Resource Damage Assessment (R. Kopp & V. Smith eds., 1993) for discussion of these issues.

58. Cummings & Harrison, *supra* note 14, at 3.

59. F. Cross, *Natural Resource Damage Valuation*, 42 Vand. L. Rev. 269 (1989).

ically evaluated the implications of 23 contingent valuation studies for valuing CERCLA-related resource services. Moreover, the *Ohio* court's ruling also noted that the same Department of Interior Technical Information Document included an annotated bibliography with ". . . 323 articles and studies related to natural resource assessments, including many treatises addressing CV [contingent valuation] methodology."[60]

Indeed, the Desvousges and Kahen component of the DOI report provides a summary of the needs, assumptions, and limitations of contingent valuation that emphasized a format that is inconsistent with the Exxon sponsored studies. They highlighted the need for in-person interviews, the importance of taking steps to ensure the commodity presented to respondents is "consistent with economic principles yet credible to respondents," and the need for careful statistical analysis of CVM results.[61] Thus, the issue is not, as CH imply, whether the court was ill-informed, but rather whether there is sufficient new information to warrant a change in their evaluation of CVM. Here my evaluation of the new research is different from theirs.

The only new study with a commodity that resembles a situation where CVM is applied in a context that is similar to that relevant to damage assessment is the recent Duffield-Patterson study.[62] Ignoring the nonrespondents in both the hypothetical and actual donations, these authors found, as CH report, that CVM estimates are 35 percent larger than average cash contributions. However, Duffield and Patterson also found that there is no significant difference in the means or the frequency distributions between CVM and actual donations when the analysis focuses on respondents. Differences between CVM and actual donations arise when different analysts make different assumptions about the values to impute to nonrespondents. Developing a framework that consistently deals with these nonresponses is among the challenges facing research designed to evaluate the conditions when CVM accurately measures "genuine economic values."

CH's penetrating review of the experimental literature suggests that it is important to consider how the findings of these experiments are being used. The existing literature tells us little about the importance of strategic behavior for CVM surveys for several reasons. First, as they note, the institutional context for experiments and CVM surveys is very different. Equally important, the commodities involved, opportunities to engage in repeated decisions, specified technologies for delivering the public goods (along with implicit divisibility of their

60. 880 F.2d at 95.

61. W. Desvousges & V. Kahen, Measuring Natural Resource Damages: An Economic Perspective (Sept. 1985) (prepared for the CERCLA 301 Task Force, U.S. Dept. of the Interior, Research Triangle Institute).

62. *See* Plott, *supra* note 39 at 1.

public goods), and the incentives themselves are quite different from the primary focus of CVM surveys.

CH's review clearly identified that Mitchell and Carson's conclusions about free-riding should not have been interpreted as a dismissal of free-riding in all circumstances. The overall literature on free-riding (including the more recent experiments) offers a mixed record of findings about when it can be important. What remains unanswered is whether it is possible to test (in an experimental setting) the approaches used in CVM surveys to attempt to control strategic behavior.

Controlled experiments linking CVM and actual sales are an important extension to the simulated market approach introduced by Bishop and Heberlein. However, to offer an effective basis for evaluating the correspondence between preferences estimated based on CVM in comparison to actual sales of commodities, they will require greater attention to demand modeling, taking account of the differences in people's characteristics and constraints. Blackburn, Harrison and Rutström recent attempts to estimate the bias in hypothetical purchase intention questions in comparison with actual choices is an attempt to develop an empirical model of how people may make mistakes in responding to hypothetical questions.[63]

The NOAA Panel report was published on January 15, 1993. It clearly accepts CVM studies as capable of producing: ". . . estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use [nonuse] values." The report expands upon this evaluation, qualifying how it is to be interpreted by noting that:

> The phrase 'be the starting point' is meant to emphasize that the panel does not suggest that CV estimates can be taken as automatically defining the range of compensable damages within narrow limits . . . . The Panel is persuaded that hypothetical markets tend to overstate willingness to pay for private as well as public goods. The same bias must be expected to occur in CV studies. To the extent that the design of CV instruments makes conservative choices when alternatives are available . . . this intrinsic basis may be offset or even over-corrected . . . . The judicial process must in each case come to a conclusion about the degree to which respondents have been induced to consider alternative uses of funds and take the proposed vehicle seriously . . . . The Panel's conclusion is that **a well-conducted CV study provides an adequately reliable benchmark to begin such ar-**

63. Blackburn et al., *supra* note 43.

guments. It contains information that judges and juries will wish to use, in combination with other evidence, including the testimony of expert witnesses.[64]

The panel's report suggests fairly broad guidelines for evaluating whether a CV study has been "well conducted," but in commenting on the issue of reliability of CV estimates, the report clearly places the burden of proof with survey designers.

The proposed guidelines seem to represent a formidable "barrier-to-entry" for CVM research used in damage assessments.[65] Nonetheless, caution and use of the best practice methods based on today's information, would seem to be prudent. Our prospects for success will be influenced by clear communication with respondents, analyst understanding of what are the important (but often non-economic) dimensions of the problem context from respondents' perspectives, respondents' knowledge and willingness to deal with the proposed choices seriously, and the details of survey implementation and data analysis.

Their guidelines for designing a CVM study offer a good start for a general CVM protocol. Some aspects of their recommendations will likely be regarded by CVM practitioners as unnecessary. However, it may be difficult to offer clear-cut evidence (as opposed to expert judgment) to support their objections. Likewise, the CVM critics may also find fault with Panel recommendations that are rendered as if they were based on extensive experiments. The evidence cited to support the specifics of their guidelines is sparse.

Several aspects of their report were disappointing and seem likely to cause problems. First, the Panel's report suggests that adherence to the Guidelines will increase the reliability of CVM estimates, but they do little to identify which elements are most important.[66] Second, the Panel's report provides no discussion of their views about the limitations (or lack of them) in the other methods available for measuring people's demands for nonmarketed resources in comparison with CVM. While it is true that these approaches cannot estimate passive use values, when use value offers the major source of losses from in-

64. Report of NOAA Panel on Contingent Valuation, 58 Fed. Reg. 4610-11 (1993) (emphasis added). The NOAA panel included (in the order listed on their report): K. Arrow, R. Solow, P. Portney, E. Leamer, R. Radner and H. Schuman.

65. The report does appear to offer some leeway in CVM surveys designed for pure research by acknowledging opportunities to economize on survey costs or to combine multiple research objectives.

66. Their comments on the contribution of their proposed Guidelines to reliability provide very few hints as to how they feel it will be affected by departures from their recommendations. The most one can find is a statement that: "A CV survey does not have to meet each of these guidelines fully in order to qualify as a source of reliable information to a damage assessment process. Many departures from the guidelines or even a single serious deviation would, however, suggest unreliability *prima facie*." 58 Fed. Reg. 4601, 4608 (1993).

juries, a choice of methods will need to be made and some appraisal of the relative performance of CVM versus indirect methods would be exceptionally helpful.

Finally, the Panel treats the extent of the market (including both the geographic and commodity dimensions) as a legal issue. This approach is misleading. Within the "legally defined group" distinguishing who would demand restoration for injured services encompasses the issues of substitution and embedding. It is therefore not simply a procedural matter.

There can be no ultimate resolution to questions that require we learn how people are affected by activities that cannot be subjected to private individualized controls and choice. Understanding how they feel about such issues will continue to be difficult unless we trust what they tell us. This does not mean we uncritically accept CVM. Rather it suggests that interviews to learn stated preferences like the various types of choices we can observe, represent forms of social interaction. Until they are adequately incorporated into our models of people's behavior, there will remain skepticism about using CVM for valuation information. The NOAA Panel's report has taken a first step at charting a course of research and "best practice" methods for litigation to begin that process. At this stage in the accumulated research record, there is nothing in the CVM results to date that suggests the method provides estimates of per household willingness to pay that would be less reliable than those available using the indirect methods. Each approach must be evaluated on a case-by-case basis.

Table 1. Smith-Osborne Meta Analyses of Contingent Valuation Estimates of WTP for Visibility Changes[a]

Indepdendent Variables	Mean Values	Models[c]				
		(1)	(2)	(3)	(4)	(5)
Intercept		-1.098 (-0.75)	1.536 (5.26)	1.367 (4.39)	1.568 (6.66)	1.521 (4.23)
Percent change in visibility	.700[b]	.688 (2.53)	.639 (50.81)	1.120 (2.32)	.684 (53.95)	1.130 (2.42)
East*percent change in visibility	.155[d]	-.356 (-2.09)			-.516 (-18.42)	-.822 (-2.53)
Interative bidding (=1)	.052			.195 (1.81)	.277 (7.34)	.108 (1.14)
Direct question (=1)	.086			-4.650 (-10.85)		-4.613 (-14.25)
Direction of change (=1)	.302	1.335 (1.32)	.341 (1.14)	.402 (1.38)	.327 (1.38)	.395 (1.74)
Asked region (=1)	.655	1.826 (1.57)	-.404 (-1.39)	-.451 (-1.46)	-.460 (-1.86)	-.462 (-1.86)
Park interview (=1)	.095			-.694 (-9.07)	-.673 (-9.20)	-.912 (-7.06)
Question framed for visibility change at revisit to site (=1)	.172	1.665 (1.48)	-.915 (-1.97)	-1.017 (-1.88)	-.728 (-2.78)	-1.304 (-2.95)
Entrance fee	.198					.289 (5.36)
Residents in sample	.707					-.200 (-1.23)
n		115	88	97	106	115
R^2		.337	.362	.729	.450	.728

[a] Numbers in parentheses are ratios of coefficients to White consistent variance estimates

[b] This variable is scaled by 100 for these estimates so it is the proportionate change in the miles of visibility in absolute magnitude - [revised - initial]/initial.

[c] The original Chestnut-Rose [1990] model specified constant value WTP per visitor per day in 1988 dollars = 6.02 in [revised/initial visibility] = -3.57 East *in [revised/initial visibility], with R^2 = .71
(-2.54)

[d] Fraction of observations for visibility changes at eastern sites

Table 2: Cummings - Harrison - Rutström Calculator Experiments: Logit Models[a]

	n = 6		
Independent Variables[b]	(1)	(2)	(3)
Intercept	2.65 (0.57)	.76 (0.18)	2.65 (0.65)
Price	.08 (0.09)	.08 (0.10)	-.28 (-0.34)
Actual sale = 1		3.78 (1.52)	
Actual sale* price			.70 (1.47)
R^2	.002	.44	.42

[a] The data for this correspond to the ratio of no responses to total respondents in each of the seven samples described in Cummings, Harrison, and Rutström (CH and R) [1992]. Samples 1 through 6 received the same basic instructions with 1 to 3 the hypothetical sales and 4 to 6 the actual sales. These are ordinary least squares based on a log odds model.

[b] The numbers in parentheses are the ratios of the estimated coefficient to its estimated asymptotic standard error for the null hypothesis of no association.

Table 3: Micro Analysis of the CH and R Calculator Sales Experiment[a]

Independent Variables	Full Sample	Hypothetical Treatment	Actual Sale Treatment
Intercept	1.132 (3.30)	1.335 (3.37)	1.369 (2.00)
Price	.061 (0.81)	.066 X 10^3 (0.00)	.269 (1.53)
Actual Sale = 1 (Otherwise = 0)	.807 (3.03)		
Don't Own Calculator = 1 (Otherwise = 0)	-1.582 (-6.53)	-1.475 (-.5.23)	-1.992 (-3.83)
n	241	124	117
pseudo R^2	.31	.22	.38

[a] The dependent variable is discrete with 1 = no and 0 = yes. Estimates were computed using Probit and the number in parentheses under each estimated coefficient is the ratio of the coefficient to its estimated asymptotic standard error.