Making the Sale on Contingent Valuation

Sameer H. Doshi[*]

Scholarship and jurisprudence have not seriously considered the question of whether the contingent valuation (CV) technique of monetizing preferences for nontradable public goods is consistent with the Daubert *standards for scientific evidence. The greatest difficulty is in establishing that CV is testable and has measurable error rates; this problem is consonant with criticisms that economists have leveled at the CV method more generally. Additionally, the "state of the art" of contingent valuation practice has recommended the use of the willingness-to-pay question format for CV, rather than willingness-to-accept. This recommendation is misplaced in many cases, particularly in calculating damages in environmental tort cases. This Article shows why a willingness-to-accept question format is more desirable on theoretical and practical grounds, and how microeconomic theory may be used to construct an operational test of the fit between CV estimates of willingness-to-accept and actual revealed valuations—thus satisfying the* Daubert *requirements. Finally, this Article discusses how CV evidence should be treated in actual litigation, considering the superiority of the willingness-to-accept format.*

I.   INTRODUCTION

    How much would you pay to keep your drinking water free of sewage? That may seem like an odd question, because chances are that your tap water is already, at the least, sludge-free. Unless you have lived in poorer parts of the world, it is difficult to imagine a scenario in which you feared for your safety before stepping underneath the shower. It might make more sense to imagine how much bribe money you would require to allow a small risk of fecal matter now and then in your lemonade. For my mother, though, enjoying a sewage-free faucet is not enough; she insists on installing a fancy reverse-osmosis filter before she would consider preparing her afternoon tea from the tap. For someone with my mother's preferences, it makes sense to ask her how much she would pay for such a gadget (and indeed, we can just inspect her credit card bill to find out). Also, now that she has enjoyed the filter for a while to wash fruits with some serenity, she could probably place a reasonable dollar estimate on her selling price for the filter. The average individual, not quite so fastidious, may have never seen this device, though, and will be unable to value its selling price properly.

    These questions are not idle games, but very important empirical tools for calculating damages and making regulations in the environmental space. Contingent valuation—essentially, asking people to report their own valuations in imagined scenarios—is one of the few empirical methods used in economics for making a quantitative estimate of individuals' benefit from nontraded goods or services. Because explicit market prices are not available for clean air, safe schools, or the existence of rare plant species, analysts can use econometric methods to infer valuation from market data,[1] or else survey people and ask them how much they value these items. The latter technique, contingent valuation (CV), has been criticized for various infirmities that make it allegedly unreliable as an estimator. Because of these flaws, we should carefully consider whether CV belongs in the legal process of fact-finding, and if so, what kind of CV questions would be the most appropriate. Federal courts and federal agencies have allowed the use of

---

    1.   Indirect market-based estimation techniques include the travel cost method, which uses natural variation in the time and cost of visiting an environmental amenity for different population segments in order to estimate a demand curve; and the hedonic pricing method, which uses multiple regression techniques to find environmental quality premia, or risk premia, in property values or labor wages.

CV in calculating natural resource damages, but no commentator has directly considered the question of whether CV evidence is admissible under the prevailing legal doctrines for scientific expert testimony.

In this Article, I discuss the concerns raised by economists about the CV technique, show how CV fits into the neoclassical theory of public goods, and propose a new framework for reconciling CV with the doctrinal requirements for scientific evidence. Specifically, I argue that a willingness-to-accept question format in CV studies, rather than the popularly used willingness-to-pay format, is the correct way to elicit economic valuations of nontradable, public environmental goods. I analyze the testability of the technique and "knowability" of its error rate. Finally, I consider the optimal evidentiary treatment of CV studies, given the contested discourse over the validity of the method.

II.   CV IN A NUTSHELL

Let us first glance quickly at the method of contingent valuation in order to understand why it poses such difficulty for a judicial gatekeeper considering CV evidence as expert testimony. Neoclassical economics has generally looked at a decision-maker's revealed behavior in order to infer a preference relation over choice objects and to assign a money valuation to those objects. However, for situations where goods are nontradable, as with public goods like environmental quality that cannot be directly purchased, market prices are unavailable and thus revealed-preference techniques for inferring valuations are extremely difficult. By contrast, stated-preference techniques ask respondents to state their numerical valuation for a particular object verbally. "Contingent valuation," one type of stated-preference valuation method, is so called because it asks respondents to imagine a hypothetical or contingent scenario in which they might personally be required to finance the provision of a public good, or in which they might be required to trade away their right to that public good.

A typical CV survey samples random respondents by phone, by postal mail, over the internet,[2] or in person. The survey presents a good or amenity (sometimes a private or excludable good, but usually with a "public" character like environmental quality) to the respondent and asks the respondent how much he would be willing to pay to acquire the good, or to secure a policy of protecting the good. The basic question can ask

---

2.   Internet surveys are far cheaper than the other methods, but they are prone to nonrandom selection bias, and do not allow the questioner to probe the respondent's thinking, as in the interactive survey formats.

for an open-ended valuation, or else it can ask the subject to vote yea or nay on a explicit, specific price (the latter is termed the "dichotomous choice" method). With the dichotomous-choice method, the price posed to respondents varies randomly, and the frequency of a "yes" answer at different prices can be used to construct a demand curve, or an expected willingness-to-pay (WTP) value. Another way to implement dichotomous choice is by the "iterative bidding" method, which starts by posing a particular price and asking the respondent to say whether she would buy at that price. If the respondent votes yes (no), the questioner then poses a higher (lower) price and asks if the respondent would be willing to pay that. These questions continue, according to an incremental schedule, until the respondent votes no (yes), thus establishing the ceiling of the respondent's willingness to pay. The same techniques can be imple-mented with a different question frame: rather than asking how much a respondent would pay for environmental quality (willingness to pay), the questioner can ask how much a respondent would accept to cede a claim on, or allow destruction of, environmental quality (willingness to accept, or WTA). I will explore this dichotomy *infra*.

      CV can focus on a respondent's "use value" of an environmental object: that is, the value that a user of a lake places on the ability to fish, or the value that a hiker places on a clean mountainside. CV questions can also seek to estimate "nonuse values" that do not involve any physical connection of the citizen to the environmental object. These values include option value (a positive value placed on the right of possible use of some environmental object in the future), altruistic value (satisfaction that a citizen gets from the knowledge of others using the environmental amenity), existence value (a citizen's satisfaction due to the very existence, or nondestruction, of the environmental object), and bequest value (a good feeling from the thought of future generations being able to enjoy the environmental amenity).[3] Other indirect methods for measuring valuation of nontradable goods like environmental quality rely on revealed preference, i.e., observations of market behavior, but almost by definition these methods can capture only use value, not nonuse value. Individuals hardly show their regard for nonuse value of environmental amenities by any real-world behavior, save perhaps for contributions to conservation organizations.[4] Donations to the Sierra

---

    3.   *See* ROBERT C. MITCHELL & RICHARD T. CARSON, USING SURVEYS TO VALUE PUBLIC GOODS: THE CONTINGENT VALUATION METHOD 61-65 (1989).
    4.   Donations to environmental and wildlife organizations comprised $3.53 billion (nominal dollars) in 1994, meaning that (using GDP as a measure of aggregate income) individuals on average gave only about four cents of every $100 earned to environmental causes.

Club and the like are a very imperfect vehicle for an individual to buy environmental quality: the probability that the organization will try to preserve a particular environmental amenity is less than certain and difficult for the individual to estimate, and the probability that the organization's advocacy will be pivotal in preserving the particular amenity is similarly complicated.

In general, and even when capturing use values, the CV method may be preferable to revealed-preference methods for a variety of reasons. One advantage of contingent valuation over other indirect techniques is that it can draw a representative sample of the population's preferences, whereas any indirect market-based technique for eliciting environmental quality valuation may place institutional barriers to the participation of low-income people. Also, CV allows respondents to express their intensity of preferences for environmental goods, which is better than a ballot referendum can do.[5] However, the income effect inherent in willingness-to-pay estimates (whether based on revealed or stated preference) may be seen as unfair by giving richer people a greater weight in the social benefit calculation.[6]

On its face, the CV technique offers an effective means of placing quantitative value on environmental quality, addressing several problems of the modern administrative state: How should agencies calculate the benefits of particular policies or projects in undertaking Kaldor-Hicks-style efficiency calculations? How should courts set damages in environmental tort cases, for deterrence or compensatory purposes? CV

Jerrell Richer, *Green Giving: An Analysis of Contributions to Major U.S. Environmental Groups*, RESOURCES FOR THE FUTURE DISCUSSION PAPER 1 (1995), *available at* http://www.rff.org/documents/RFF-DP-95-39.pdf. There is no *a priori* way to estimate how much contribution seems "enough," but a value of this magnitude can serve as a lower bound for aggregate valuation of environmental nonuse values. Aggregate valuations implied by contingent valuation far exceed the $3.53 billion mark; for example, one CV survey of non-Alaskans found that they would be willing to pay $31 per household to avert a similar disaster (to the Exxon Valdez oil spill) in the future, which makes $2.8 billion when summed over all U.S. households. *See* Robert R.M. Verchick, *Feathers or Gold? A Civic Economics for Environmental Law*, 25 HARV. ENVTL. L. REV. 95, 104 (2001). It is important to note that the calculated $2.8 billion relates to just one oil spill, not to the dozens of other environmental problems around the country. Presumably the aggregate valuation of environmental quality using CV would be at least a couple orders of magnitude greater than the level of donations to conservation organizations.

    5.   Verchick, *supra* note 4, at 105.

    6.   There are anecdotes of CV respondents who report, "I would be willing to pay $1 million to protect the Grand Canyon," only to be reminded that their annual budget constraint is $25,000. In theory at least, CV studies, alone among valuation techniques for nontradables, can correct for this income effect by instructing all respondents to imagine that their income is a standard number, say $50,000. *Id.* (noting that efficiency-based demand aggregations "necessarily overemphasize the preferences of the wealthy and underemphasizes the preferences of the poor").

data can also be used for more primitive applications besides policymaking: In calculating "green national product" or in using natural resources as a production input for economic growth calculations, how much value should be imputed to the environment? CV overcomes the problems of other valuation techniques for nontradables and appears to present an elegant way to fill in important gaps in the regulatory process. I will focus mostly on CV as a damages estimator in judicial settings, but much of the following can be extended to other applications including growth accounting and administrative penalties.

III. THE FLAWS OF CV

While CV is generally established as a tool for regulators and environmental tort lawyers, the technique has been criticized by economists on a variety of fronts—some claiming that CV fails to capture actual valuations in the context of rational decision-making, some alleging that CV creates incentives for strategic or dishonest answers, and some arguing that the whole exercise lacks meaning because respondents do not have "true" valuations to report. Each of these allegations is problematic for the validity and thus admissibility of CV evidence into judicial and regulatory proceedings. In this Part, I will quickly review these critiques and show why these problems are not devastating to the logical coherence of CV as a technique for measuring value.

A. Artificial Choices

As many economists have noted, any acceptable CV method should be "consistent with the logical requirements of rationality . . . and at least broadly consistent with sensible features of economic preferences."[7] Despite challenges in the last two decades from various findings of behavioral economics,[8] the neoclassical account of economic decision-making, involving maximization of a utility function over affordable bundles of goods, is still the dominant paradigm for decision theory

---

        7.   Daniel McFadden, *Contingent Valuation and Social Choice*, 76 AM. J. AG. ECON. 689, 690 (1994).
        8.   Two relevant texts are Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263 (1979) (establishing "prospect theory," which shows how decision-makers behave in conflict with expected utility theory when assessing choices under uncertainty), and RICHARD THALER, THE WINNER'S CURSE: PARADOXES AND ANOMALIES OF ECONOMIC LIFE (1994) (explaining departures from rationality such as status quo bias and the endowment effect).

within economics. To be meaningful and consistent, any model of economic valuation should be consistent with this axiomatic approach.

The first main area of concern is the hypothetical nature of the payments asked for in CV studies. Most of the goods under consideration, like clean air, are nontradable, and never could be the subject of a private purchase. Even if I wanted to spend money to purchase an increase in clean air, I could not, because Coaseian-type bargaining would likely be impossible with so many polluters contributing to smoggy conditions. (An additional reason commonly given for nontradeability is the lack of property rights over clean air, but this is actually less salient: although nobody owns the air, emissions and ambient pollution levels can be monitored, and I could enter into private contracts, enforceable in regular courts, with each polluter over its level of emissions.) So because CV questions are only hypothetical, and thus do not force the respondent to think in a real decision-making mode akin to everyday spending decisions, the technique may not be able to elicit a "true" valuation.[9] In this light, one meta-study showed that generally CV valuations diverge from market-based valuations.[10]

Critics have identified other flaws that expose the artificiality of the method: for example, the results of a single CV study may be internally inconsistent between different question formats, in that WTP estimates using the dichotomous-choice CV method tend to exceed WTP estimates for the same good using the open-ended valuation method.[11] A variety of studies have also shown that WTA estimates from CV studies tend to exceed WTP estimates systematically.[12] Given that citizens will likely never explicitly face the "How much would you pay for extra

---

    9.   *See* Frank B. Cross, *Restoring Restoration for Natural Resource Damages*, 24 U. TOL. L. REV. 319, 329-30 (1993).

    10.   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. RESOURCES J. 1 (1994).

    11.   Thomas C. Brown, Patricia A. Champ, Richard C. Bishop & Daniel W. McCollum, *Which Response Format Reveals the Truth About Donations to a Public Good?*, 72 LAND ECON. 152 (1996) (finding the discrepancy between dichotomous-choice answers and open-ended answers, for goods with use value and also for nonuse value).

    12.   *See* Elizabeth Hoffman & Matthew L. Spitzer, *Willingness To Pay vs. Willingness To Accept: Legal and Economic Implications*, 71 WASH. U. L.Q. 59 (1993); Richard C. Bishop, Thomas A. Heberlein & Mary Jo Kealy, *Contingent Valuation of Environmental Assets: Comparisons with a Simulated Market*, 23 NAT. RES. J. 619 (1983) (finding that estimated WTA for hunting permits in a CV study is at least three times the estimated WTP for the same hunting permits); Jack L. Knetsch & J.A. Sinden, *Willingness To Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value*, 99 Q. J. ECON. 507 (1984) (showing that WTA significantly exceeded WTP in experiments wherein lottery tickets were allocated to student participants).

environmental amenity?" or "How much would you accept to lose some environmental quality?" questions in real life, it is not immediately clear which question format should be seen as correct. However, I will show *infra* why there is a difference in the normative and practical desirability of these formats.

B.   Constructive Preferences

In addition to concerns that CV does not capture a "real" valuation, several authors have also found that CV responses are susceptible to influence by external cues and parameters of the questions, raising worries that the results of a CV survey are not likely to reflect a "true" valuation. Most notable is an "anchoring effect" or "starting-point bias" in iterative-bidding or dichotomous-choice question modes, whereby the respondent's final answer is correlated with the initial number presented by the questioner.[13] Boyle, Bishop, and Welsh explain this "starting-point bias" by positing that respondents may lack information about how to value the environmental good, and they may thus use the posited hypothetical price as a clue to market value.[14] Consistent with this hypothesis, one study found that starting-point bias is more pronounced when querying respondents about their valuation of hypothetical new beaches vs. existing beaches.[15]

Anchoring might show that respondents have no preformed preference relation, or, at the least, that they have a range of valuation rather than a point estimate.[16] Perhaps they lack a good sense of the value of the good, but they are Bayesians who interpret the initial price given as a clue to the true valuation.[17] In any case, the anchoring effect seems

---

13.   *See, e.g.*, Jonathan Silberman & Mark Klock, *The Behavior of Respondents in Contingent Valuation: Evidence on Starting Bids*, 18 J. BEHAVIORAL ECON. 51 (1989); McFadden, *supra* note 7; Kevin J. Boyle, Richard C. Bishop & Michael P. Welsh, *Starting Point Bias in Contingent Valuation Bidding Games*, 61 LAND ECON. 188 (1985); Ian J. Bateman, Ian H. Langford & Jon Rasbash, *Willingness-to-Pay Question Format Effects in Contingent Valuation Studies, in* VALUING ENVIRONMENTAL PREFERENCES 511 (Ian J. Bateman & Kenneth G. Willis eds., 1999); *see also* Robin Gregory et al., *How Precise Are Monetary Representations of Environmental Improvements?*, 71 LAND ECON. 462, 470 (1995) (asking respondents to state their open-ended valuation answer in terms of a "budgetary unit," either $1, $2, $5, $10, or $20, and finding that the mean WTP is positively correlated with the budgetary unit used in the question); Donald Green et al., *Referendum Contingent Valuation, Anchoring, and Willingness To Pay for Public Goods*, 20 RESOURCE & ENERGY ECON. 85, 110 (1998) (arguing that "psychometric" anchoring effects are more salient than rational strategic behavior in explaining CV responses).

14.   Boyle, Bishop & Welsh, *supra* note 13, at 190.

15.   Silberman & Klock, *supra* note 13.

16.   Gregory et al., *supra* note 13, at 470.

17.   David Aadland, Arthur Caplan & Owen Phillips, *A Bayesian Examination of the Interaction Between Anchoring and Cheap Talk in Contingent Valuation* (July 2007),

inconsistent with the model of a primitive preference relation that exists apart from external parameters. Relatedly, it has been noted that a respondent may accept a given bid $X in a dichotomous-choice format, but could alternatively accept a lower bid $Y, and then reject $X, in an iterative bidding format.[18] This variance might be because the respondent, after noting Y, would feel that the government would "waste" the extra (X-Y).[19] By undermining the existence of a valuation independent of the study, these phenomena make it more difficult to conceive of CV as representing a "true" valuation that can be plugged into regulatory gaps. It is important to see that these anchoring effects have largely appeared in WTP studies, rather than WTA studies; it seems that respondents have a difficult time estimating the value of an environmental amenity that they have not previously possessed. We might not expect that the same effects would obtain in a WTA-type study, quizzing people about the potential loss of a property right or environmental amenity that they already held.

The problem of measuring CV against true utility is further highlighted by another quirk of the CV technique: the alleged failure of CV respondents to take note of the quantities involved correctly. Critics such as Diamond and Hausman[20] have noted the "embedding effect" anomaly, in which a CV respondent might report "$X" when asked his valuation for cleaning up one lake, and then might report the same $X when asked his valuation for cleaning up five lakes (including the

---

http://uwadmnweb.uwyo.edu/econfinance/phillips/ACP_final_version_(3).pdf. Another author (John K. Horowitz, *A New Model of Contingent Valuation*, 75 AM. J. AG. ECON. 1268 (1993)) suggests that Bayesian respondents might interpret the very fact that the questioner is asking about that particular public good (as opposed to any other possible goods) as a clue to the paramount social value of the good. A similar Bayesian story could explain why willingness to accept payment to destroy a (currently existing) project might exceed willingness to pay to create a (currently uncreated) project.

   18.   Robert T. Carson et al., *Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez*, 25 ENVTL. RES. ECON. 257 (2003).

   19.   This latter phenomenon could be seen as anchoring on fairness, rather than rational Bayesian anchoring.

   20.   *See* Peter A. Diamond & Jerry A. Hausman, *Contingent Valuation: Is Some Number Better Than No Number?*, 8 J. ECON. PERSP. 45 (1994); Daniel Kahneman & Jack L. Knetsch, *Valuing Public Goods: The Purchase of Moral Satisfaction*, 22 J. ENVTL. ECON. & MGMT. 57 (1992). Related to the "scope fallacy," some critics point to "sequencing errors" in the results of CV surveys, in which respondents considering a whole sequence of environmental commodities appear to value a specific environmental commodity more highly when it comes earlier in the sequence. The critics say that a consumer should value an environmental commodity equally, regardless of where it appears in a sequence. But if anything, this sequencing effect shows that the respondent is *correctly* updating and considering his tighter budget constraint each time he answers a new question; the phenomenon is consistent with some notion of diminishing marginal utility for environmental improvements.

identical lake from the first question). The most notable example is a study by Desvousges et al. in which respondents were asked for their WTP to prevent the deaths of migratory waterfowl in oil fields.[21] Desvousges finds that WTP did not vary significantly when respondents were asked in separate versions of the questionnaire about saving 2000 birds, 20,000 birds, or 200,000 birds.[22]   Schkade and Payne rerun essentially the same survey about the migratory birds, this time requiring respondents to think out loud about their answers, and again find no significant scope effect upon WTP.[23]

      Still, the scope criticism is not completely devastating to the integrity of CV vis-à-vis traditional decision theory; the critique may presuppose too much about the shape and arguments of utility functions. By a nonsatiation axiom, rational citizens should derive more existence value when more birds are saved, if saved birds loom as separable in the psyche. However, what if the utility-giving good is the knowledge of saved birds, rather than the existence of a bird? If the unit of account is "acts of saving birds," we could still have nonsatiation and transitivity, except that saving 100 birds at a single site in Boston would be treated the same as saving 1000 birds at the same site in Boston. The point is that the testability of CV depends crucially on our prior belief about what the utility function looks like, which may be wrong in the case of nontradables. In any case, the "saving birds" story fundamentally requires an inquiry of how much citizens would pay to gain environmental benefit, whereas natural resource damage assessments, as I will show, involve compensating citizens for ceding their property right to something. For damage assessments, a very different approach is required. However, even if the coherence of CV can be saved, it is important to know whether the data gleaned thereby can be used at all in the legal process.

IV.   LEGAL TREATMENT OF CV

A.   *CV Per Se*

      In order to understand why both scholarship and doctrine teach us so little about the admissibility of CV evidence, we must trace the

      21.   William H. Desvouges et al., *Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability*, *in* CONTINGENT VALUATION: A CRITICAL ASSESSMENT (J.A. Hausman ed., 1993).
      22.   *Id.*
      23.   D.A. Schkade & J.W. Payne, *How People Respond to Contingent Valuation Question: A Verbal Protocol Analysis of Willingness To Pay for an Environmental Regulation*, 26 J. ENVTL. ECON. & MGMT. 88, 97 (1994).

methodology from its origins, and consider how evidence law evolved along the same time path. Later, in Part V, I will begin to show how the logical requirements of both intellectual systems may be satisfied.

CV was used beginning in the 1960s to estimate values for nontradable goods,[24] and began to receive growing support and application from economists beginning in the 1980s. Meanwhile, regulatory authorities began to recognize it as a valid tool for environmental damage assessment contemporaneously. Under the CERCLA statute for cleanup of contaminated land, the Department of Interior (DOI) issued regulations allowing the use of CV for determining use values and nonuse values of environmental objects.[25] Recognition and employment of CV grew throughout the 1990s; the Clinton administration also allowed regulatory use of CV data in cost-benefit analysis by agencies.[26] Under the Clinton administration, both DOI and NOAA called CV the "only method available" for estimating nonuse values of environmental goods.[27]

As an evidentiary matter, there might be some concern that CV is hearsay (by definition, survey data like CV is evidence of statements made outside a courtroom, proffered to prove the truth of such statements), but this was largely obviated by a 1963 federal case, *Zippo Manufacturing Co. v. Rogers Imports*, which held that survey results are admissible under the hearsay rule because respondents' answers are "expressions of presently existing state of mind, attitude, or belief"—one of the historical hearsay exceptions.[28] The judge complicated this rule by stating that the necessity of survey evidence should be weighed against the "circumstantial guaranty of trustworthiness surrounding the making of the statement."[29] It is not clear whether any judge has ever considered CV under the *Zippo* rubric. There are many cases in which CV data may be the only available evidence for establishing use or nonuse value of an

---

24. In the earliest days of CV, the technique was also known variously as the "Davis approach," "questionnaire approach," or "hypothetical valuation." *See* Richard C. Bishop & Thomas A. Heberlein, *Measuring Values of Extramarket Goods: Are Indirect Measures Biased?*, 61 AM. J. AG. ECON. 926 (1979).

25. The relevant regulatory rule was in 43 C.F.R. § 11.83 (d) (1987), although this subsection has since been deleted.

26. *See* REGULATORY WORKING GROUP, OFFICE OF MGMT. & BUDGET, ECONOMIC ANALYSIS OF FEDERAL REGULATIONS UNDER EXECUTIVE ORDER 12866 § III(B)(4) (Jan. 11, 1996), *available at* http://www.whitehouse.gov/omb/inforeg/riaguide.html#iii.

27. *See* U.S. DEPARTMENT OF INTERIOR (DOI), NATURAL RESOURCE DAMAGE ASSESSMENTS, 59 Fed. Reg. 52,749, 52,751 (1994); NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA), NATURAL RESOURCE DAMAGE ASSESSMENTS, 59 Fed. Reg. 1062, 1074 (1994).

28. 216 F. Supp. 670, 683 (S.D.N.Y. 1963).

29. *Id.*

environmental amenity; but on the other hand, the "guarantees of trustworthiness" could be open to question, as discussed *supra*. At any rate, the case law around the admissibility of CV grew richer beginning in the 1980s.

Two major events brought CV into public view in 1989. First, the Exxon Valdez oil spill off the Alaskan shoreline resulted in lengthy litigation in which CV evidence was brought forth by the plaintiffs to show lost nonuse values from the environmental damage.[30] For the first time, CV was a star witness in a major environmental dispute. Contingent valuation also received a major endorsement in a 1989 United States Court of Appeals of the District of Columbia Circuit case, *Ohio v. Department of Interior*. The court held that under the CERCLA statute, nonuse values should be included in damage assessments; and also that DOI erred when it issued regulations stating that "option and existence values may be estimated in lieu of use values only when use values cannot be determined."[31] Anticipating some of the concerns later raised by the *Daubert v. Merrell Dow* court (a seminal case for evidence law, discussed *infra*), the court also upheld the allowance of CV in the CERCLA regulations, deciding that CV is a "best available procedure," per the statutory charge, for assessing lost use values when market-based methods are unavailable. The court considered arguments that CV results have too much variance and are (upwardly) biased, but rejected both criticisms: the extent of natural resource damage is expected to have high variance, said the court, and "more sophisticated questioning" can cure the bias problem.[32] Finally, the court upheld the rebuttable presumption given to damage assessments (including CV) by the statutory text[33]; without the presumption, said the court, there would "loom the specter of prolonged battles of experts."[34]

Despite the D.C. Circuit's endorsement of DOI's regulatory scheme, federal trial judges have not been as friendly when directly dealing with CV evidence. Judges' attitudes toward the admissibility of CV were

---

30.   *See, e.g.*, RICHARD T. CARSON ET AL., A CONTINGENT VALUATION STUDY OF LOST PASSIVE USE VALUES RESULTING FROM THE EXXON VALDEZ OIL SPILL, 1992 REPORT TO THE ATTORNEY GENERAL OF STATE OF ALASKA (Nov. 1992); Carson, *supra* note 18. The Exxon Corporation later funded a research symposium in 1992 made up almost entirely of economists who were critical of the CV technique. In this Article, I quote from some contributions in the resulting book, CONTINGENT VALUATION: A CRITICAL ASSESSMENT (J.A. Hausman ed., 1993). For a more recent CV study of the damages from the Valdez oil spill, see Carson, *supra* note 18.

31.   Ohio v. U.S. Dep't of Interior, 880 F.2d 462, 464 (1989).

32.   *Id.* at 477-78.

33.   Comprehensive Environmental Response, Compensation, and Liability Act, 42 U.S.C. § 9607(f)(2)(C) (1980).

34.   *Ohio*, 880 F.2d at 480.

exemplified in *Mercado v. Ahmed*,[35] an early-1990s wrongful death case predating *Daubert*. The federal trial judge in *Mercado* opined that "there is no basic agreement among economists as to what elements ought to go into the life valuation. There is no unanimity on which studies ought to be considered. There is a lack of reliability."[36] Thus the judge rejected an economist's testimony, based on a secondary survey of CV studies, as to the value of an injured plaintiff's lost pleasure of life. This judge, perhaps applying the (pre-*Daubert*) *Frye* standard of general acceptability in the field, decided that CV was not sufficiently accepted among economists to comprise admissible evidence.[37]

Another pre-*Daubert* CV controversy came in a CERCLA cost recovery action, *Idaho v. Southern Refrigerated Transport*,[38] in which the federal trial judge rejected a CV survey of the existence value of steelhead fish, labeling CV as "conjecture and speculation," lacking "reasonable certainty" sufficient to establish existence value.[39] The court's concern largely focused on the discrepancy between the original purpose of the survey—to plan changes in the Northwest hydropower system—and the proffered evidentiary purpose, to establish the value of injury to wildlife, in litigation that arose a few years after the survey. The survey had asked respondents to place a value on doubling the size of the salmon and steelhead population from 2.5 million to 5 million. In essence, the judge was rejecting this survey data because it used a WTP question mode, rather than the more appropriate WTA. The judge did not articulate a rule for CV admissibility quite so explicitly, but as I will show *infra*, WTA is a more theoretically appropriate and empirically tractable measure of damages. It is plausible that the judge might have accepted the old evidence if it had appropriately measured citizens' willingness to accept compensation to redress their loss of the fish.

B.   *The Rules of Evidence*

Given the problems that have been identified with CV, and trial judges' distaste for it, is the technique appropriate for use in a judicial

---

35.   Mercado v. Ahmed, 756 F. Supp. 1097, 1101 (N.D. Ill. 1991).

36.   *Id.* at 1103.

37.   It seems likely that the judge might reach the same decision in 2007 under the *Frye* standard, given that many top economists continue to question the validity of the technique. Still, CV does have its supporters: two frequent authors in the field have written that CV is "clearly a method that many experts routinely rely on as a base for their judgments." Richard T. Carson & Robert Cameron Mitchell, *Contingent Valuation and the Legal Arena*, *in* VALUING NATURAL ASSETS 231 (Raymond J. Kopp & V. Kerry Smith eds., 1989).

38.   No. 88-1279, slip op. (D. Idaho Jan. 24, 1991).

39.   *Id.* at 19.

proceeding, under the usual rules of evidence?   Currently there is no well-articulated rule or standard for resolving this question: the early-'90s decisions mentioned above have been supervened by the later *Daubert* ruling. From 1923 until 1993, the federal evidentiary standard for expert testimony was given by the teaching of *Frye v. United States*, which barred expert opinion based on a scientific technique unless the technique is "generally accepted" as reliable in the relevant scientific field.[40]   *Frye* was overturned in *Daubert v. Merrell Dow*, a pharmaceutical products liability case, wherein the United States Supreme Court established that the rule for admissibility of scientific evidence should come from *Federal Rule of Evidence* 702, which superseded the common law doctrine of *Frye*.[41]   FRE 702 allows a qualified expert witness to testify if (and only if) her "scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue."[42]   The *Daubert* court laid out four factors to operationalize the rule of FRE 702: the theory or technique should be (1) testable; (2) subject to publication or peer review; (3) characterized by known or potential error rates, along with standards controlling operation of the technique;[43] and, like in *Frye*, (4) accepted widely within a relevant scientific community.[44]

    Clearly, many economists believe CV can be a useful tool and have published papers on CV in peer-reviewed journals. The "accepted widely" standard under Daubert is more liberal than the "general acceptance" standard under the earlier *Frye* rule, so the views of economists, though not unanimous, are almost surely not a bar to the admission of CV evidence.[45]   The requirement that CV have clear standards is easy to meet by a brief brush through the literature. Previous attempts at standardizing a proper CV methodology have focused on reducing the bias and variance of results, as a response to critics who claim that the technique is unreliable. For example, Cummings, Brookshire, and Schulze list four "reference operating conditions"

---

    40.   Frye v. United States, 54 App. D.C. 46 (1923).
    41.   Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993).
    42.   FED. R. EVID. 702.
    43.   Presumably error rates are useful for juries to evaluate probabilities of guilt, or probabilities of particular propositions being truthful, conditional on certain evidence. However, it is not clear that juries are expert users of Bayes' Rule, so the utility of data on error rates may be limited. *See, e.g.*, Brian C. Smith et al., *Jurors' Use of Probabilistic Evidence*, 20 L. & HUMAN BEHAVIOR 49 (1996); William C. Thompson, *Are Juries Competent To Evaluate Statistical Evidence?* 52 L. & CONTEMP. PROBS. 9 (1989).
    44.   *Daubert*, 509 U.S. at 593-94.
    45.   *Id.* at 588. The *Daubert* court contrasted the "rigid" *Frye* standard with the "liberal thrust" of the newer and superseding Federal Rules of Evidence. *Id.*

including subjects' familiarity with, and their prior experience with valuing, the environmental good; low uncertainty; and a WTP rather than WTA approach.[46] Mitchell and Carson present similar guidelines, including reminder of the budget constraint; specification of the exact good being valued; identification of any resulting price changes in related goods; and specification of the method of payment.[47] Similarly, economists Kenneth Arrow and Robert Solow prepared a report for the NOAA in 1993, in which they laid out necessary guidelines for CV studies to produce reliable estimates.[48] These included many of the above factors, plus large sample size; face-to-face interviews; pilot surveying before the full study; dichotomous-choice question format; questions about "why" the respondent decided as she did; distinction between "steady-state" vs. interim losses of environmental quality; and sensitivity to the time path of damage vs. restoration.[49] Finally, and most notably for this Article, the report recommended (like the earlier Cummings recommendations) use of WTP rather than WTA, based on the idea that the systematically lower WTP estimates are more "conservative" and thus more desirable.[50]

The Arrow-Solow panelists concluded that conditional on their recommended guidelines, CV studies do "convey useful information" and should be seen as "reliable by the standards that seem to be implicit in similar contexts, like … the assessment of other damages normally allowed in court proceedings."[51] Of course, it is difficult to meet all of these factors in practice: in-person surveys cost more than telephone or internet surveys; pilot programs cost time and money, and large sample sizes drive up the cost of surveys. Nonetheless, the literature has at least identified state-of-the-art standards for the technique, and this is enough to satisfy that element of the *Daubert* test.

Though the Arrow-Solow report was released in the same year as the *Daubert* decision, very few observers have attempted to evaluate CV in light of the *Daubert* factors for admissibility of scientific testimony directly.[52] This is not a trivial exercise, because the points that are not

---

46. R.G. Cummings et al., *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, 63 ECON. GEOGRAPHY 358, 358-59 (1987).

47. MITCHELL & CARSON, *supra* note 3, at 301-03.

48. NOAA, Proposed Rules: Natural Resource Damage Assessments Under the Oil Pollution Act of 1990, 58 Fed. Reg. 4601 (1993).

49. *Id.* at 4610.

50. *Id.* at 4612.

51. *Id.* at 4610.

52. Ghosh suggests that all nonmarket valuation techniques, including hedonic pricing (discussed *supra* note 2) and CV, should fail the *Daubert* test because "social science is not testable because all relevant variables cannot be controlled for in a systematic manner and . . . the

clear under *Daubert* are whether CV is testable and whether CV has a known (or even measurable) error rate. Addressing the testability and error rate issues, the Arrow-Solow panel advised comparing the results of real-life referenda on paying for public goods with CV studies of the same question.[53] Johnson argues in a single paragraph that "contingent valuation satisfies the *Daubert* test easily," based on an inspection of the four key *Daubert* elements.[54] However, Johnson elides the third of these factors, the known error rate; instead he avers to the standards for "proper implementation" of CV studies, as in the Solow/Arrow report.[55] In order to satisfy *Daubert*, we must show theoretically that the error rate of CV is estimable, and then go out and gather the data to do just that. If we can find a "natural experiment" in which citizens reveal their valuation for environmental goods in an uncontrolled, quasi-market setting, we may be able to estimate the accuracy of the CV method, in order to satisfy the *Daubert* strictures. My goal in this Article is to show a potentially promising way to implement and test CV in a way consistent with evidentiary rules and the purposes of environmental protection.

C.   *Judicial Treatment of CV Since* Daubert

Since the *Ohio* and *Daubert* decisions, CV and related techniques have not fared well in the federal courts. Judges have generally disallowed nonmarket valuation evidence on grounds of its "prejudicial" value, without reaching the question of whether it meets the prongs set forth in *Daubert*. It may be that judges have avoided the question of CV's admissibility under *Daubert* precisely because of its difficulty.

In the post-*Daubert* world, much of the relevant case law on environmental valuation has related to hedonic damages evidence, rather than CV. Hedonic valuation methods[56] have been popular because they require relatively little expenditure of time and effort; once a relevant data set is located and an appropriate regression model specified, a statistics package can calculate the desired estimates in a couple of

---

underlying model will always have assumptions that can never be tested." That this claim is too broad to speak directly to the admissibility of CV only highlights the paucity of commentary on this issue. *See* Shubha Ghosh, *Fragmenting Knowledge, Misconstruing Rule 702: How Lower Courts Have Resolved the Problem of Technical and Other Specialized Knowledge in* Daubert v. Merrell Dow Pharmaceuticals, Inc., 1 CHI.-KENT J. INTELL. PROP. 1, 55 (1996).

  53.   NOAA, Proposed Rules, 58 Fed. Reg. at 4607.

  54.   Gordon J. Johnson, *Paying the Piper: Comments on Liability for Natural Resource Injury: Beyond Tort*, 6 ALB. L.J. SCI. & TECH. 265, 282-83 (1996).

  55.   *Id.* at 283.

  56.   *See supra* text accompanying note 2.

minutes. One federal trial judge in a 1996 products liability case[57] held that hedonic damages fail to pass the conditions of reliability and helpfulness under FRE 702, and additionally, that hedonic damages evidence is unduly prejudicial under FRE 403: given the complications of WTP calculations, "the Court does not believe the jurors would accept the method so much as they might latch onto the first expert figure lobbed their way." Other federal judges have ruled similarly.[58]

Hedonic valuation evidence is useful for calculating the value of nontradables (such as clear view; workplace health) that are indirectly tradable through other markets, such as through prices in the real estate market or wages in the labor market. However, there are many times where personal preferences over environmental goods are not recoverable through real-world inspection, especially in cases of existence value for remote objects. Where human behavior has generated no relevant and useful information on valuation, CV evidence is the next best thing. Yet the received doctrine tells us very little about whether CV evidence may be used in litigation.

A couple of post-*Daubert* federal cases have addressed the admissibility of CV evidence, and favorably so. In *General Electric v. U.S. Department of Commerce*,[59] the corporate defendants challenged NOAA's natural resource damage assessment regulations under the Oil Pollution Act, which required only that damages assessment procedures be "reliable and valid for the particular incident."[60] The regulation neither required nor proscribed use of CV evidence for nonuse values, but instead gave discretion to the trustees bringing suit.[61] The defendants argued that NOAA acted arbitrarily and capriciously by failing to codify the recommendations for CV contained in the 1993 Arrow-Solow report explicitly. Holding that NOAA's regulation was acceptable, the D.C. Circuit said that the "reliable and valid" requirement was consistent with the specific safeguards for CV that were spelled out in the Arrow-Solow report—and thus, "if performed correctly, contingent valuation can produce both useful and reliable results."[62] Finally, in *National Ass'n of*

---

    57.   Kurcnz v. Honda N. Am., Inc., 166 F.R.D. 386, 390 (W.D. Mich. 1996).
    58.   *Id.* In *Hein v. Merck & Co.*, 868 F. Supp 230, 231 (M.D. Tenn. 1994), a federal trial court held, applying the *Daubert* standard, that expert testimony on hedonic damages is speculative, unreliable, and invalid. In *Smith v. Ingersoll-Rand Co.*, 214 F.3d 1235, 1245-46 (10th Cir. 2000), the Tenth Circuit upheld, under FRE 702, a federal trial judge's decision to deem testimony on hedonic damages as relevant, but exclude quantification of such damages.
    59.   128 F.3d 767 (D.C. Cir. 1997)
    60.   15 C.F.R. § 990.27(a)(3) (2007).
    61.   *Id.*
    62.   *Gen. Elec.*, 128 F.3d at 773-74.

Manufacturers v. U.S. Department of the Interior, the D.C. Circuit upheld DOI's rule permitting use of relatively old contingent valuation studies, citing Section 301(c)(1) of CERCLA, which requires the DOI to identify the "best available procedures" to determine natural resource damages.[63] The three-judge panel failed to comment on the merits of CV, citing only the *Ohio* case, and deferring to DOI's "expert judgment."[64]

    In short, the federal courts have yet to articulate a standard for dealing with contingent valuation evidence in the *Daubert* world, besides saying that they will defer to agencies' judgment under their statutory mandates. Such a doctrine of deferral, apparently following the rule of *Chevron v. NRDC*,[65] does little to guide agencies in deciding how to meet their own statutory requirements of reliability; nor does it say what federal courts should do under *Daubert* when parties in a nonstatutory environmental damages action seek to bring CV evidence into the judicial process.

    And what of federal agency policies on CV evidence? Today, the key CERCLA statutory framework for natural resources damages assessment authorizes two types of regulations: "Type A" regulations for simplified assessments (requiring little empirics, based on units of emissions or units of affected area), and "Type B" regulations for the "best available" alternative procedures for determining direct and indirect injury.[66] At a general level, the regulations allow only valuation methodologies that are "reliable for a particular incident and type of damage."[67] Today, the key CERCLA regulation for CV allows employment of the CV methodology to determine use values, option values, and existence values for injured natural resources; however, option and existence values may be estimated "only if no use values can be determined"—the very provision that was struck down by the *Ohio* court.[68] Executive attitudes to CV have seemingly become more skeptical in the eighteen years since *Ohio*. The parallel regulation under the Oil Pollution Act, a similar statute regulating damage assessments for maritime oil spills, discusses the quantification of injury assessment, but

---

    63.   134 F.3d 1095 (D.C. Cir. 1998).
    64.   *Id.* at 1116.
    65.   Chevron U.S.A. v. Natural Res. Def. Council, 467 U.S. 837 (1984) (establishing that a federal court should always defer to an agency's statutory interpretation unless there is some ambiguity in the statutory language).
    66.   42 U.S.C. § 9651(c)(2) (2000).
    67.   43 C.F.R. 11.83(a)(3)(i) (2007).
    68.   *Id.*; *see also* Ohio v. U.S. Dep't of Interior, 880 F.2d 462, 464 (1989) ("Estimation of option and existence values shall be used only if the authorized official determines that no use values can be determined.").

does not discuss specific valuation techniques, as the CERCLA regulations do.[69]

Meanwhile, section 1006(e) of the Oil Pollution Act and its associated regulation, 15 C.F.R. § 990.13, promulgated by NOAA, gives any natural resource damages assessment by a trustee under the statute a rebuttable presumption in any judicial proceeding. A statutory provision of CERCLA gives a similar rebuttable presumption to any natural resources damages assessment made by a CERCLA trustee.[70] By shifting the burden of production to the defendant who hopes to challenge the CV evidence, the rebuttable presumption forces the defendant to show that CV is inadmissible. This rebuttal might require, under *Daubert*, a showing that CV is a nontestable technique. However, as I will discuss below, this showing is difficult given the range of validity and reliability studies available for both WTP and WTA.

V. WILLINGNESS TO PAY AND THE THEORY OF PUBLIC GOODS

A. *Strategic Effects*

Measuring the error rates of contingent valuation is almost certainly a nontrivial exercise, because it is not clear what "true" valuation we should compare the CV estimates against. In order to understand how CV might be subject to systematic error rates, it is necessary to examine the microfoundations of willingness-to-pay calculations. The textbook theory of public goods typically assumes that we should vertically sum demand curves in order to find the aggregate social demand for the public good: at any given quantity, social value is the sum of all individual values. Due to the nonrivalrous character of the public good, in order to determine the aggregate losses from some environmental injury, we need merely to add up all values for the involved quantity of environmental amenity.

However, there are several problems with this formulation. The theory expects to find a granular WTP for each individual, as though she were purchasing the environmental good for herself—independent of any other person's WTP. In real life, citizens usually do not pay for public goods, and if they do, their valuation is probably influenced by the institutional form of payment, and their perception of their peers' level of payment.[71] If these discount factors form part of the true valuation that

---

69. 15 C.F.R. § 990.52 (2007).
70. Comprehensive Environmental Response, Compensation, and Liability Act, § 107(f)(2)(C), 42 U.S.C. §§ 9601-9675.
71. *See* Schkade & Payne, *supra* note 23, at 99.

CV estimates should be verified against, then testing CV becomes more difficult because many of these nuances cannot be captured by the CV question format. If these discount factors are not part of the "true" valuation (one that would obtain if the environmental good could be made an excludable private good) then finding a true valuation by looking at market behavior is nearly impossible.

Consider the strategic incentives facing a CV respondent. When respondents are told that the survey is simply intended to determine aggregate valuation, and respondents will face no resulting real-life cost, then the respondents have an obvious incentive to overinflate their bid, in order to increase the probability of provision. (Doing this has strictly positive expected value for the respondent, assuming that she likes the public good.) When respondents are told that in case the public good is provided, they will be required to pay their full stated WTP, they have an obvious incentive to understate their bid. (If they do not, then they have a 100% chance of getting zero surplus from the transaction.)[72]

The "implementation frame"[73] of the question—that is, the apparent role of the respondent in deciding the probability of provision of the good—also affects respondents' incentives to state their true preference. Assume that a CV respondent is told that he will be required to pay some fraction of his stated WTP. The respondent is then likely to alter his response based on whether he judges the "contingent" nature of the question to be credible: if he believes that his own payment will be pivotal in determining provision of the good, he will be more inclined to state his true WTP. On the other hand, if he believes that his answer will not be pivotal, he will understate the bid.

Several critics have noted that estimates of WTP from CV studies tend to be larger than actual contributions to environmental conservation efforts,[74] suggesting that the sign if not the magnitude of the average error is predictable. That CV estimates of WTP for public goods seem upwardly biased compared to individuals' real-life contributions is not necessarily an indictment of CV; in fact, we expect that actual private payments for public goods should be systematically below maximal

---

72.   Furthermore, a respondent who would derive greater utility from the public good will have a greater incentive, on the margin, to respond in the first place. Thus, questionnaires administered via postal mail or the internet probably have a selection bias because respondents have more interest in the subject matter (and, likely, a higher WTP) than recipients who choose to not respond. *See* Richard T. Carson, Nicholas E. Flores & Norman F. Meade, *Contingent Valuation: Controversies and Evidence*, 19 ENVTL. & RESOURCE ECON. 173, 180 (2001).

73.   Green et al., *supra* note 13, at 110.

74.   *See supra* text accompanying note 4.

WTP due to free-riding.[75] This undervaluing is a feature of markets for public goods generally, rather than CV studies, but it is another reason why valuations elicited via CV may be less than a "true" value.

Complicating matters, the questions in any CV study can be permuted such that the theoretical direction of strategic bias is ambiguous. Imagine that respondents are told that in case the public good is provided, every citizen will pay an identical, predetermined amount (say, $100). Respondents with WTP above $100 then have an incentive to inflate their stated WTP, so as to increase the probability that the good is provided. On the other hand, respondents with WTP below $100 have an incentive to reduce their stated WTP (perhaps to zero), so as to reduce the probability that the good is provided.

Imagine, alternatively, that respondents are told that in case the public good is provided, every citizen will pay some unspecified proportion of his stated WTP. Many respondents will want to reduce their stated WTP, if they calculate that the resulting reduction in probability of public good provision is outweighed by their expected out-of-pocket cost saved. On the other hand, some respondents may think that the required proportion will be very small, and thus they actually will *inflate* their stated WTP in order to increase the probability of provision.[76] All this is to say that the bias inherent in any CV study—or, perhaps, the error rate, per *Daubert*—is difficult to predict *a priori*, without diagnostic studies.

The "CV studies are full of strategic liars" theory is not universal, however. Mitchell and Carson argue that the incentive for nontruthful reporting in CV surveys will be weak because of: high informational requirements for strategic behavior; each respondent's belief that many other respondents are involved, thus reducing the probability that a single person's overbidding would affect the level of provision; and the payment vehicle's evoking realistic budget constraints.[77] Looking empirically, one experiment by Bohm found a lack of evidence of strategic behavior or "cheating" among respondents, even when the incentives to cheat were explicitly described to respondents.[78]

---

    75. *See, e.g.*, K.E. McConnell, *Reflections on the* Ohio *Decision*, 34 NAT. RES. J. 93, 104-05 (1994).

    76. This example and the immediately previous one are adapted from Peter Bohm, *Estimating Demand for Public Goods: An Experiment*, 3 EURO. ECON. REV. 111, 113-14 (1972).

    77. MITCHELL & CARSON, *supra* note 3, at 155.

    78. Bohm, *supra* note 76, at 124.

B.   A Simple Nonstrategic Model

CV estimates of "willingness to pay" to obtain a public environmental good (or prevent destruction thereof) have been frequently found to be systematically lower than estimates of "willingness to accept" payment to allow destruction of the same good.[79]   This discrepancy is disturbing for proponents of CV evidence who make no attempt to explain whether WTP or WTA is a more theoretically appropriate measure of value.   How can we reconcile the WTP-WTA disparity with rational decision-making?   According to a neoclassical setup with convex preferences, maximum WTP and minimum WTA for a good should be separated only by a wealth effect: that is, if I require more compensation to relinquish the good than I was willing to pay to acquire the good, that is only because I feel richer when I possess the good.   The standard WTP question—"How much would you pay to acquire this improvement?"— is essentially measuring the compensating surplus of acquiring a commodity, whereas the standard WTA question—"How much would you accept to allow this environmental damage?"—is measuring the compensating surplus of losing the commodity.[80]   Basic consumer theory shows that WTA will exceed WTP if the transition in question is a price decrease of a tradable good.[81]   Due to the associated income effect, equivalent surplus (WTA) exceeds compensating surplus (WTP), because the former is calculated by taking an integral of the Hicksian demand at the new utility, whereas the latter is calculated by taking an integral of the Hicksian demand at the old utility.

On the other hand, for any exogenous quantity improvement of a tradable, continuously divisible commodity, WTA should equal WTP: in

---

79.   *See* Hoffman, *supra* note 12, at 113.   For logical symmetry, the final version of the willingness-to-accept question might be "How much would you accept to halt the expected creation of this public good?"   This question type is confusing and rarely used.   I will explore this issue further *infra* Part V.C.

80.   *See* David S. Brookshire, Alan Randall & John R. Stoll, *Valuing Increments and Decrements in Natural Resource Service Flows*, 62 AM. J. AG. ECON. 478, 481 (1980).   My analysis assumes that the WTA question does *not* start at the same reference point as does the WTP question. That is, my analysis requires that the WTP question might start at a "0" level of environmental quality and ask the respondent about gaining x, while the WTA question would start at an x level of environmental quality and ask the respondent about losing x.   In other words, the only difference between the WTP scenario and WTA scenario is a shift in property rights to the respondent.   Some, but not all, CV surveys take this form.   *See infra* Fig. 1.

81.   The exception would be if the income elasticity of demand is zero, in which case Hicksian demand is the same as Marshallian demand, and compensating surplus is the same as equivalent surplus.   In practice, with environmental goods, income elasticity of demand is probably positive, so this exception does not obtain.

other words, markets will always clear perfectly.[82] But with an exogenous quantity improvement to a nontradable good like clean air, we cannot apply the same analytic step, because the price is fixed at infinity and the Hicksian demand curve essentially does not exist. Unlike a price change, both substitution effects *and* income effects govern the WTP-WTA difference, so the theoretical relationship between WTA and WTP is not immediately obvious.[83] Randall and Stoll show that for "indivisible or lumpy" goods (a category of which environmental amenities are exemplary: rare redwood trees, for example, come only in discrete and small quantities) the Hicksian compensating surplus of a welfare-harming quantity reduction (i.e., the WTA of this quantity reduction) should differ from the equivalent surplus of the same welfare-harming quantity reduction (i.e., WTP), assuming nonzero income elasticity of WTP.[84]

     To illustrate this further, let us consider the graph in Figure 1, in which z is a basket of tradable private goods and x is the nontradable public environmental good. Let us assume convex preferences over z and x.

---

     82.   Randall & Stoll show this result in a two-good competitive setup in which a consumer is free to trade off one good for the other. *See* Alan Randall & John R. Stoll, *Consumer's Surplus in Commodity Space*, 70 AM. ECON. REV. 449 (1980). Similarly, Hanemann establishes the same property in a setup with a public good and private good, for the case that the two goods are perfectly substitutable. W. Michael Hanemann, *Willingness To Pay and Willingness To Accept: How Much Can They Differ?*, 81 AM. ECON. REV. 635 (1991).
     83.   This latter observation comes from Carson, Flores & Meade, *supra* note 72, at 185.
     84.   Randall & Stoll, *supra* note 82. However, the discrepancy cannot be too great. Various work including Randall and Stoll, *id.*, have attempted to estimate bounds on the potential difference between WTP and WTA. For example, Milgrom shows that if a single respondent's stated WTP for acquiring an environmental good ($CS_{0,1}$) is half as much as the respondent's stated WTA for losing the environmental good ($CS_{1,0}$) then this implies essentially that the marginal propensity of environmental spending (with respect to income) in the relevant range is 0.5. Milgrom points out rightly that this seems unreasonably large. *See* Milgrom, *supra* note 21, at 430.

Figure 1
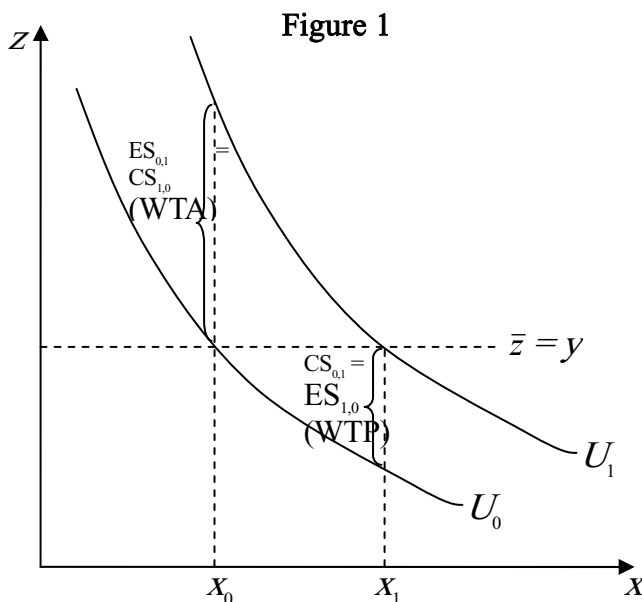
Normalize the price of z to \$1. Say the consumer has *y* of income, and the initial exogenously given endowment of the public good *x* is x_0. Then the original indifference curve is U_0, cutting through the bundle (x_0, $\bar{z}$) where $\bar{z} = y$. An exogenous improvement of *x* from x_0 to x_1 will take the consumer to the new indifference curve U_1, cutting through the bundle (x_1, $\bar{z}$). It is easy to then identify the compensating surplus on the graph, i.e., the WTP for the improvement: call this $CS_{0,1}$. We can also identify the equivalent surplus for the same improvement: call this $ES_{0,1}$. The graph also makes it easy to see that for an exogenous welfare-harming transition from x_1 to x_0, the compensating surplus, i.e., the WTA for this transition (call this $CS_{1,0}$) is the same as the equivalent surplus for the welfare-improving transition, $ES_{0,1}$. And the equivalent surplus of the welfare-harming transition, i.e., the WTP to avoid suffering the loss (call this $ES_{1,0}$), is the same as the compensating surplus of the welfare improvement ($CS_{0,1}$).

Table 1, below, together with Figure 1 make clear that no matter how the WTP question is framed—in terms of $CS_{0,1}$ or $ES_{1,0}$—the valuation should be the same in theory. And no matter how the WTA question is framed—in terms of $CS_{1,0}$ or $ES_{0,1}$—the valuation should be equal.

So the question of how to frame a CV question ultimately reduces to WTA vs. WTP—as we initially expected. As Figure 1 makes clear, it

is not obvious that one should systematically exceed the other: in the graph drawn above, WTA appears bigger because the marginal indirect utility of extra cash seems to be lower at lower levels of the public environmental good. However, given a different preference structure (if the marginal indirect utility of income is higher at lower levels of the public environmental good—which might be plausible if severe pollution is making life very unpleasant, and there is substitutability between air quality and health care, say) then the graph could easily show WTP as being bigger.[85] Another way of putting this is that if the marginal rate of substitution of z for x declines more rapidly at higher levels of utility (the environmental good satiates you more easily when you start off happier) then WTP would exceed WTA.

Table 1		
	Improvement	Loss
Compensating surplus	$CS_{0,1}$: "How much would you pay to obtain this improvement?" (WTP)	$CS_{1,0}$: "How much would you need to be fully restored? / How much would you accept in order to allow the environmental damage?" (WTA)
Equivalent surplus	$ES_{0,1}$: "How much would you accept in order to forego the expected improvement?" (WTA)	$ES_{1,0}$: "How much would you pay in order to avoid suffering the loss?" (WTP)

The description in this Part shows how WTP and WTA can systematically differ in a neoclassical model of nontradable goods. However, this model has ignored the strategic possibilities discussed in the previous Part. There may be good reason to question this assumption: why should we think that strategic lying is not a salient problem? First of all, it is important to recognize that the WTA-WTP disparity can be generated by a simple model of a rational actor, without the need to gesture to departures from rationality such as the anchoring effect. Even

---

    85.   A related explanation for the WTP-WTA discrepancy may be "income constraint consideration"—meaning that people who greatly value clean water, say, would like to bid a very big number in their WTP (i.e., their $CS_{0,1} = ES_{1,0}$ would be very high with the population average income), but they lack sufficient income (their y is low in Figure 1) to realistically bid this. On the other hand, WTA questions, phrased either of the ways shown in Table 1, can elicit unboundedly large valuations. *See, e.g.*, William D. Schulze, Ralph C. d'Arge & David S. Brookshire, *Valuing Environmental Commodities: Some Recent Experiments*, 57 LAND ECON. 151, 166-67 (1981).

if we accept that conscious lying or some sort of misrepresentation on CV surveys may be rampant, the key point is that, as discussed in Part V.A, the direction of bias is unclear and probably differs from person to person. If the distribution of valuations across the population is symmetric around some mean, and if the dichotomous-choice referendum value given to respondents is systematically varied roughly across the range of the distribution, then we should expect an equal proportion of responses that are upwardly biased as the number of downwardly biased responses. Also, even if WTP is dogged by systematic bias in one direction, then for most causes (fear of payment, the anchoring effect, etc.) WTA will be biased in the same way, preserving the WTA-WTP spread. Thus, the existence of bias or strategy should not destroy the basic explanatory power of the graphical model above for showing why WTP and WTA differ.

C.   *WTA and WTP in Policy Context*

Which measure of value should be used for legal purposes? Aside from evidentiary-quality considerations, which I will explain shortly, there is a normative component to this question. Brookshire, Randall, and Stoll in an early article on CV, show how the measure of damages to be used implies a prior judgment about property rights to environmental quality: compensating surplus of the loss (what I call $CS_{1,0}$, or WTA) measures welfare changes "as if the individual had a right to his initial level of welfare," whereas equivalent surplus of the loss (what I call $ES_{1,0}$) "treats the individual as if he had only a right to his subsequent level of welfare."[86] They conclude that compensating-surplus measures are more appropriate for valuing potential improvements, given that most CV questions involve the decision of whether or not the government should provide public goods to individuals.[87] Considering a slightly different question, Cross and Dobbins recommend using WTA as the measure of natural resource damages, because the idea of citizens selling their assets is consistent with the public, democratic character of natural resource damage assessment.[88]

If the legal question is "When you were enjoying that environmental good, how much was it worth to you?" then $CS_{1,0}$ seems to be the obvious question to measure damages: this measure is by definition the amount

---

86.   Brookshire, Randall & Stoll, *supra* note 80, at 480.

87.   In other words, in valuing a potential environmental improvement, governments should seek to estimate $CS_{0,1}$.

88.   Jeffrey C. Dobbins, *The Pain and Suffering of Environmental Loss: Using Contingent Valuation To Estimate Nonuse Damages*, 43 DUKE L.J. 879, 919 (1994).

of money that the possessor of the good would have needed in order to sell the good, and also the amount of money required to restore the victim to her original position after the loss. On the other hand, if the legal question is "Now that you no longer have that environmental good, how much would you pay to get it back?" then $CS_{0,1}$, which is a WTP measure, appears the obvious valuation. Willingness-to-pay makes sense only when the regulatory scheme is valuing a potential acquisition, not a loss.

As Table 1 shows, WTP can be stated in two very different ways, which both are consistent with *not* having a property right over the environmental good, and which should theoretically elicit the same numerical valuation. If I already have a clean river, which is scheduled to be polluted soon, I might pay $100 to keep it clean; or if I don't have a clean river, then I might pay $100 to make it clean. The mirror image is a WTA measure of environmental value: if I already have a clean river and a clear property right over it, I might accept $150 to dirty the river; or if I don't have a clean river, I might accept $150 to absolve the government of its promise to clean up the river.

Similarly, several experiments have found that the *status quo ante* fate of an environmental object has a crucial role in determining the magnitude of WTA values relative to WTP.[89] For example, if a question is "How much would you accept to sell this environmental object?," subjects are much more reluctant to sell, and the WTA-WTP gap is greater, when they know that the potential buyer intends to destroy the object. Boyce et al. explain this by saying that the subjects may not be very willing to pay to save the object from destruction by somebody else ("Why is that my responsibility?") but will be more concerned for stewardship of the object when they have a prior property right over it.[90] Returning to the framework of Figure 1, we could again say that if the marginal indirect utility of cash is higher when you hold the object in your hands than when you do not, then it is expected that WTA should exceed WTP.

To make this more concrete, I will consider the most obvious application of the WTA-WTP dichotomy. Normatively, WTA is probably appropriate in CERCLA or similar tort cases, where the purpose of the regulatory scheme is compensation or deterrence: that is the amount of money that would restore the victim's original, higher, utility level after

---

    89.   *See, e.g.*, Rebecca R. Boyce et al., *An Experimental Examination of Intrinsic Values as a Source of the WTA-WTP Disparity*, 82 AM. ECON. REV. 1366 (1992) (finding the crucial role in determining WTA-WTP valuation of the *status quo ante* fate of environmental objects).

    90.   *Id.* at 1371.

injury to clean land, or else prevent the tortfeasor (the Potentially Responsible Party) from disturbing the citizen's environmental health and enjoyment in the first place. If the statutory purpose were deterrence, expected liability (given risk-neutrality of injurers) from pollution should exactly equal the expected harm.[91]   Both of these statutory purposes assume that the citizen had an initial right to the uncontaminated land. Although either phrasing of the WTA question ($ES_{0,1}$ or $CS_{1,0}$) should elicit the correct answer, the $ES_{0,1}$ phrasing ("How much would you accept in order to forego the expected improvement?") is not apposite for a tort context; only the $CS_{1,0}$ phrasing makes sense when someone has lost a previous level of environmental quality. Presumably, the plaintiff's standing to sue establishes that some property right or perhaps some "zone of interests" has been violated, and the plaintiff wants a remedy to affirm her preexisting right.[92]

    The 1993 Arrow-Solow Report, considered the gold standard for articulating the correct techniques for a CV study, advocates a "conservative design" that "tends to underestimate willingness to pay"; and it counsels against the WTA question format, because WTP is the "conservative choice."[93]   The report does not consider the theoretical or evidentiary fit of WTP to a CERCLA or private-tort damages context, but rather favors the method that is likely to produce a lower number. It is not clear why a "conservative" or lower valuation is more desirable, other than to make CV a more politically palatable technique. If WTA is a more appropriate question format, and prevalent utility functions are such that WTA systematically exceeds WTP, it is not obvious, from a social welfare perspective, why WTA should be disfavored just because the resulting damage awards seem too large by some subjective standard. If damage awards calculated this way seem to threaten the solvency of firms, perhaps this is evidence that the industry itself is less sustainable than thought.

    The Arrow-Solow report itself admits that "the conceptually correct measure of lost passive-use value for environmental damage that has already occurred is the minimum amount of compensation that each affected individual would be willing to accept."[94]   It is thus extremely puzzling why WTP might be advanced as a requisite for state-of-the-art

---

    91.   I will ignore the less-than-100% probability of a guilty verdict.

    92.   *See, e.g.*, Lujan v. Defenders of Wildlife, 504 U.S. 555 (1992) (articulating Article III constitutional requirements for suing for environmental injury in federal court).

    93.   NOAA, Proposed Rules: National Resource Damage Assessments Under the Oil Pollution Act of 1990, 58 Fed. Reg. 4609 (proposed Jan. 1993).

    94.   *Id.* at 4603.

CV for the purpose of establishing judicial damages. Yet, when CV has been used in public and private environmental tort suits, WTP has been the survey mode of choice.[95]

D.   *Why Not To Use WTA*

There may be several good reasons not to use WTA as the measure of damages (or any form of CV studies) in environmental tort suits. WTA suffers from obvious strategic problems described above: if a CV respondent is asked to report the minimum amount she would accept to cede her claim on environmental quality willingly, and if the respondent does not think that her response will be linked to taxation of some sort, then she has every incentive to state a very big number. (This dynamic is the leading reason why mainstream economists have suggested that the WTP mode should be used for CV studies.) Also, whereas a WTP-type question can be checked for realism by looking at the respondent's income and other budget categories, a WTA-based measure could in theory be unboundedly large. In a regulatory context, if the government surveys citizens for their WTA in the process of deciding whether to implement a project, one citizen could effectively veto the project by stating a very large WTA value.[96]

A final reason to eschew WTA is that the "selling" mode of valuation is less familiar to consumers than buying. Because most citizens do not regularly sell their assets in the ordinary course of their lives, they may not be able or willing to articulate a selling price for an intangible asset like environmental quality[97]; and thus, why would we base our valuation measure on a fairly artificial construct? Carson et al. suggest that CV respondents may find a WTA question "implausible" (though, they admit, it presents the theoretically appropriate property right) because "they do not believe they possess a personal property right to sell the good."[98] This objection, though, fails to appreciate that consumers are already losing some environmental quality every day by

-----

    95.   *See, e.g.*, Brown, Champ, Bishop & McCollum, *supra* note 11, at 153 (describing thirteen CV studies from the 1980s and 1990s that all inquired as to the respondents' WTP; by contrast, studies inquiring into WTA are difficult to find).
    96.   Milgrom, *supra* note 84, at 430.
    97.   *See, e.g.*, Russell Korobkin, *Policymaking and the Offer-Asking Price Gap: Toward a Theory of Efficient Entitlement Allocation*, 46 STAN. L. REV. 663, 691, 695 (1994) (identifying a "psychic aversion to bargaining" that prevents individuals from readily selling their assets). The growth of Internet-based markets for consumer-consumer transactions, such as Craigslist and eBay, may be making consumers more cognizant of their WTA for their tangible assets, at least, if not for intangibles like environmental quality.
    98.   Carson, *supra* note 18, at 281.

living in an industrialized society, so they are vaguely aware of the tradeoff entailed in their modern lifestyles. Additionally, experiments have shown that consumers can quickly learn the ropes of "selling" their entitlement, as evidenced by relative convergence of WTA and WTP over multiple iterations of a game.[99] Thus, given a chance, consumers can provide reasonable values through a WTA-type survey.

It is for this reason that we should do "diagnostic" CV WTA surveys on a regular basis, to discover who already cares about environmental quality. Citizens with prior investments in particular environmental goods will be better experienced with making tradeoffs around their usage of the amenity. They could still have some incentive to inflate their WTA, but it will be more reflexive for such individuals to think about how much the good is worth to them, in contrast to uninvolved parties who may think that environmental quality is literally priceless.

Alternatively, putting aside concerns about strategic misrepresentation, what if a "correct" WTA exceeds the cost of restoring a resource back to its undamaged state, or providing requisite ecological amenities elsewhere to approximate $CS_{0,1}$ (in cases where this is physically possible)? If restoration cost were less than WTA,[100] it would be easy to say that restoration cost is a more efficient measure of damages, given that transfers have transaction costs. Why not just restock a river with new fish, instead of providing a more expensive money-based compensation? Putting aside the fact that the *Ohio* court rejected this line of thinking ("The fatal flaw of Interior's approach, however, is that it assumes that natural resources are fungible goods"), is it inefficient to use a "WTA" compensation measure rather than a restoration measure?[101]

John F. Daum argues that we should always use restoration value as a measure of damages, rendering contingent valuation unnecessary.[102] On this view, after the resource is restored, there would be no lasting impairment of nonuse values like bequest value, option value, or existence value, so there would be no reason to compensate anyone after the resource is restored. But, as discussed in Part III.B, we should not take too restrictive a view of utility functions. It could be the case, for

---

99.   *See, e.g.*, Don L. Coursey, John L. Hovis & William D. Schulze, *The Disparity Between Willingness To Accept and Willingness To Pay Measures of Value*, 102 Q. J. ECON. 679 (1987).

100.   In theory at least, assuming that citizens hold a property right to undamaged land, restoration cost should not be less than WTA; if it were, then there would be unexploited arbitrage opportunities for a third party. I am grateful to Richard Zeckhauser for pointing this out to me.

101.   Ohio v. U.S. Dep't of Labor, 880 F.2d 462, 456 (1989).

102.   Daum, *supra* note 21, at 398-99.

example, that the disutility comes from the process of environmental degradation itself, rather than the (temporary) absence of the resource. Relying on restoration cost saves on the transaction costs of transfers, but fails to meet the twin goals of full compensation and deterrence embodied in the statutory and common law regimes of environmental regulation.

In any case, beyond the purposivist reasons to use WTA (namely, that it tracks better with the deterrent or compensatory character of environmental regulatory regimes), there are, additionally, important evidentiary questions about whether WTA-type CV studies can be reconciled with the *Daubert* standards.[103] This question is salient whether we focus on a question framed as "How much would you accept in order to lose some environmental quality?" ($CS_{1,0}$) or "How much would you accept in order to forego the expected improvement?" ($ES_{0,1}$). In the next Part, I will outline a context for validating CV answers to a WTA question against "natural" market evidence of the same valuation decision.

VI. TESTABILITY

A.   *Finding the Right Comparison*

The *Daubert* doctrine requires that the admissibility of any scientifically based piece of evidence requires that it be "testable" and have a known error rate. There is very little commentary in the literature on whether CV passes the "testability" prong of the *Daubert* doctrine.[104]

---

103.   In Part II of this Article, I suggested that CV could be used for other administrative applications, including imputing a value to environmental quality in macroeconomic production functions and growth calculations, or calculating benefits when making regulations. WTA (rather than WTP) is still appropriate for these applications: using up natural resources in the production process involves the same "selling" of environmental quality by citizens, as does environmental injury in a tort case. Also, most regulations embracing environmental quality are based on environmental preservation rather than environmental improvement. The structure of the Clean Water Act, Clean Air Act, or Food, Drug, and Cosmetic Act, for example, all involve imposing penalties for emitting dangerous substances into the natural environment or the stream of commerce. For a regulatory regime that aspired to subsidize cleanup, on the other hand, WTA might not be so appropriate.

104.   There is, however, some discussion of the testability of hedonic pricing evidence. Kuiper argues that hedonic damages evidence, a technique that, like CV, also tries to calculate valuations of nontradables, does not pass the "testability" prong of *Daubert*. Hedonic damages studies have tended to show wide disparities in value-of-life estimates, even for multiple studies of the same data. Kuiper, following the ideas of Ireland, argues that hedonic pricing data on, say, automobile air bag purchases is flawed because consumers are probably not accurately judging risk, and thus calculated hedonic prices that use the true probabilities of harm will misestimate consumers' subjective valuation of life. *See* Joseph A. Kuiper, Note, *The Courts,* Daubert*, and Willingness-To-Pay*, 1996 U. ILL. L. REV. 1197, 1229, 1247-48; Thomas R. Ireland, Walter D.

In principle, it should be straightforward to perform this sort of check for the CV technique: we need merely to compare CV-derived valuations to "true" or revealed valuations over the environment. There are two potential complications with this approach, however: first, it is not obvious which kind of revealed valuation to use. Second, there is some controversy over whether citizens have primitive preference relations and valuations over environmental amenities in the first place. I will begin by addressing the first of these problems.

   We have several ways in theory to measure the validity of WTP-based CV results. We could track individuals and compare their actual market decisions against the responses they gave in the CV survey.[105] But following respondents' market behavior is difficult, and there are also various stochastic factors that may alter an individual's WTP in the time between the CV survey and later behavior.[106] More tractably, we could take two random samples of the same population and administer CV surveys to both of them; or a CV survey to one and an electoral referendum to the other.[107] This method of testing for validity could be problematic, however, because voters in the referendum are unlikely to be a random sample of the population.

   Paralleling these options, the Arrow-Solow report of 1993, considered to establish the state-of-the-art in CV methodology, proposes two techniques in particular for validating CV estimates:[108] comparing CV estimates for public good valuation against a subsequent opportunity for respondents to pay real money for the same public good; and comparing CV estimates for a private good against actual purchases of the private good.[109] As the Arrow-Solow report admits, the validation

---

Johnson & James D. Rodgers, *Why Hedonic Measures Are Irrelevant to Wrongful Death Litigation*, 2 J. LEGAL ECON. 49, 51-52 (1992).

   105.   In this regard, Diamond and Hausman present an excellent review of studies showing a systematic discrepancy between CV respondents' stated WTP vs. their actual donations to environmental charities and other organizations. *See* Diamond & Hausman, *supra* note 21, at 21.

   106.   *See* MITCHELL & CARSON, *supra* note 3, at 181 ("[W]hether or not the person ultimately behaves as she says she intends to depends on intervening experience, the presence or the views of relevant others, and the physical setting.").

   107.   In this light, Vossler et al. compare results of an actual political referendum on raising funds to purchase open space vs. a corresponding dichotomous-choice CV survey of the same population (finding that the WTP estimate in the CV survey matches the estimate from the referendum if "undecided" responses in the CV survey are coded as "no"). *See* Christian A. Vossler, Joe Kerkvliet, Stephen Polasky & Olesya Gainutdinova, *Externally Validating Contingent Valuation: An Open-Space Survey and Referendum in Corvallis, Oregon*, 51 J. ECON. BEHAVIOR & ORG. 261, 274 (2000).

   108.   NOAA, Proposed Rules: Natural Resource Damage Assessments Under the Oil Pollution Act of 1990, 58 Fed. Reg. 4604 (1993).

   109.   The most promising attempt to compare preferences in an "actual" decision-making scenario to preferences in the CV setting may be found in Cummings et al., wherein the authors

studies of CV for public goods seem to show that CV estimates of WTP are upwardly biased. However, note that if we discover discrepancies between CV estimates of valuation and "true" valuations, this need not disqualify CV data from admissibility in court; rather, if we can characterize the bias as systematic and regular, in terms of sign and magnitude, then we may have a "known error rate" per the *Daubert* doctrine. It would also help for these purposes to provide a theoretical account of why we have bias. As discussed above, such discrepancy can spring from a variety of causes, including lack of familiarity with the good, strategic misrepresentation, or desire to free-ride. Various tests have thus been proposed to diagnose these infirmities. For example, Bohm looks for a proper valuation mechanism to have "verifiability": the extent of strategic misstatement of preferences should be measurable.[110] Accordingly, several studies have suggested ways for testing the CV technique for the presence of strategic behavior.[111] Carson et al. discuss several studies of the reproducibility (reliability) of CV, based on two types of comparison: first, multiple samples of the same population over time, and second, the same respondent's answers at multiple points in time.[112] They conclude that CV studies seem to be reliable both between samples and within samples.[113]

    Another problem with validation studies of CV is not just that the CV estimates are flawed, but rather that "real" payments may not show

---

ask subjects to provide their hypothetical WTP valuations, using a dichotomous-choice question method, for various private goods. They then ask the same subjects whether they are willing to actually purchase the goods with real money at that moment, with the same price from the first question. They find a significant difference between the percentage of "yes" answers in the two scenarios. *See* Ronald G. Cummings, Glenn W. Harrison & E. Elisabet Rutstrom, *Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?*, 85 AM. ECON. REV. 260 (1995). Arrow and his coauthors refer to an earlier version of this research in their 1993 report. KENNETH ARROW ET AL., REPORT OF THE NOAA PANEL ON CONTINGENT VALUATION 4 (1993).

    110. Peter Bohm, *Revealing Demand for an Actual Public Good*, 24 J. PUB. ECON. 135, 137 (1984).

    111. *See* MITCHELL & CARSON, *supra* note 3, at 165 (describing tests for detecting strategic behavior); William D. Schulze, *Use of Direct Methods Valuing Natural Resource Damages*, *in* VALUING NATURAL ASSETS 225 (Raymond J. Kopp & V. Kerry Smith eds., 1993) (suggesting cutting out high bids which are beyond a "reasonable" proportion of a respondent's income); Ian J. Bateman, Ian H. Langford, R. Kerry Turner, Ken G. Willis & Guy D. Garrod, *Elicitation and Truncation Effects in Contingent Valuation Studies*, 12 ECOL. ECON. 177, 179 (1995) (determining that in the case of open-ended and iterative-bidding question formats, truncation of high bids can be used to test for "severe strategic overbidding"); Boyle & Bergstrom, *supra* note 13, at 198 (Ian J. Bateman & Kenneth G. Willis eds., 1999) (recommending "identifying groups of individuals who are likely to misstate their values either purposely or inadvertently," and testing for the robustness of survey estimates if such responses are excluded).

    112. *See* Carson, Flores & Meade, *supra* note 72, at 195.

    113. *Id.*

the individual's true valuation. Any artificially constructed (or even observed market) data on "purchases" of environmental goods are likely to be below true WTP due to free-riding, or else not an authentic market purchase situation at all: an individual cannot truly buy cleaner air in the real world, because she lacks the coercive power of the state, and Coaseian bargaining with many large polluting corporations would be difficult. Thus, efforts to validate CV by comparing data from WTP-type CV studies to "actual" purchases in an experimental setting will probably be systematically biased.

     Still, at the least, it is fair to say that we can readily test WTP-type data from CV studies for validity and error rates, and thus WTP data should be admissible under *Daubert*. But all this is moot, in a way: assuming that WTA is a theoretically more appropriate measure of damages, we must ask instead, is WTA testable? If finding real-life situations where citizens pay to improve environmental quality or to avert environmental degradation is difficult, locating real examples of citizens accepting payment for ceding their claim to environmental quality is even more challenging.

     Whereas there are few real-life cases where citizens pay to acquire public goods or to avert the destruction of public goods, there are many cases in which citizens accept *ex ante* compensation from a public or private authority in order to give up their claim (or their hope) to something which they had been promised, as in a $ES_{0,1}$ type of question. It is here that the equivalence of $ES_{0,1}$ and $CS_{1,0}$ becomes useful: instead of looking for examples of citizens literally selling environmental quality, we could find evidence of citizens electorally ratifying their acceptance of particular forms of economic stimulus in exchange for a politician's broken promise to improve the environment. Additionally, we could look at negotiated "prices" for more traditional cases in which citizens agree to give up environmental quality that they previously held. To be clear, I do not mean to invoke "takings" in which the government decrees a compensation rate at fair market value.[114] Rather, I refer to situations in which citizens agree in an arm's-length transaction to cede some environmental quality to a private developer; or competition among multiple political jurisdictions to encourage the importation of environmental disamenities in exchange for industrial growth.

---

     114. In theory, true valuation could be elicited even in an obligatory takings context via a process resembling a Groves mechanism, but this is not done in practice. *See* Abraham Bell & Gideon Parchomovsky, *Bargaining for Takings Compensation*, (U. Penn. Inst. for Law & Econ., Working Paper No. 05-22, 2005).

Take, for example, a case in which a corporation decides to build a factory which will demonstrably worsen local air quality. Local citizens complain until the corporation agrees to invest a certain amount in local nonprofits, or employ a certain number of new jobs; and then the local citizens' association agrees to the development. Presumably the specific tradeoff agreed to can be taken as an indication of the citizens' willingness to accept compensation $(CS_{1,0})$.[115] If agency costs are low between the citizens and the representative organization, then agreement is a strong signal of the average valuation placed by citizens on environmental quality. In order to calculate error rates for WTA valuation estimates, data from CV surveys can be compared against valuation data from these sorts of negotiated deals. However, it is important to focus on settlements that are freely struck between the relevant citizens and the wrongdoer, rather than those that are imposed by a judicial determination of liability, or in the shadow of liability.

One might then ask: if these real-world sources of valuation data are available, why use CV at all? Why not just use the revealed valuations? The trouble is that these quasi-natural experiments probably are infrequent; every time litigation arises over damage to a particular environmental object, we will not be able to look back to a situation in which citizens en masse indicated their valuation for that object through a deal with a politician or a developer. These natural experiments are useful for occasionally calibrating the reliability of CV, but they cannot be marshaled into action every time the legal system wonders how much a nontradable environmental object is worth. For that, only CV suffices.

B.   Sizing the Right Class

Another potential source of error in the calculation of a WTA-type CV damage assessment is the breadth of the affected class. Most CV surveys estimate an average per-person valuation for the environmental good under study, estimate the size of the relevant population or the relevant market, and then multiply the two. In particular, when estimating an existence value (if use values are unavailable or irrelevant, under CERCLA), the class of individuals who care about the environmental good, to the extent that their utility function implies

---

115. My estimation technique holds if we assume that the corporation has all the bargaining power; if the citizens hold some bargaining power in such situations, then the outcome will resemble Nash bargaining, and the final deal struck might establish a valuation above the citizens' minimal WTA. However, the pie-split via Nash bargaining can be analyzed fairly simply, unlike the complex strategic interactions involved with willingness to pay for public goods.

positive $CS_{0,1}$ (WTA), could be vast.[116] If there is a correct size of the affected population to use, then an incorrect judgment of the scale leads to estimation error.

However, the risk of error in estimation may be less with WTA than with WTP: it should be relatively easier for economists and public agencies to conduct studies at regular intervals, independent of any environmental injury, that estimate the number of people who hold some interest in various environmental goods, or categories of environmental goods. This way, in the event that a particular environmental amenity is injured, we will have some prior evidence to help calibrate the size of the interested population. Individuals surveyed in these diagnostic studies might have incentives to state their interest falsely, but a list of questions to assess the respondent's prior familiarity with the good can help to assess the population's average "psychic investment" in certain types of environmental amenities. By contrast, with willingness-to-pay studies, such diagnostic surveys are more difficult, because respondents must (as recommended by Arrow and Solow) be provided with information about the environmental good.

To be sure, plaintiffs and defendants in an environmental injury case might want to redo a CV study for the specific environmental good at issue, with possibly a new estimate of the affected population size, but the possibility of preinjury diagnostic studies helps to strengthen the case for the admissibility of WTA evidence. At the least, it can help establish a known error rate with respect to aggregate valuation, in addition to estimated error rates with respect to individual valuation.

C.   *Measuring the Right Value*

The analysis thus far has assumed, of course, that there is some true subjective valuation of environmental quality to be discerned. In addition to the challenge of preference aggregation, a theory of public goods necessarily needs to account for what sort of valuations should "count" in regulatory or judicial processes. I have previously discussed different types of nonuse value, including option value, bequest value, and existence value. Some economists believe that existence value should be dismissed as undefined and imprecise[117] or simply inappropriate for an

---

116. Murray B. Rutherford, Jack L. Knetsch & Thomas C. Brown, *Assessing Environmental Losses: Judgments of Importance and Damage Schedules*, 22 HARV. ENVTL. L. REV. 51, 68 (1998) (arguing that the act of defining the affected class may be "quite arbitrary").

117. *See* Donald H. Rosenthal & Robert H. Nelson, *Why Existence Value Should Not Be Used in Cost-Benefit Analysis*, 11 J. POL'Y ANAL. & MGMT. 117 (1992) (arguing that existence value should be excluded from economic efficiency analysis because economists should not

economic valuation,[118] while others believe that existence values are held by regular people and should be incorporated into judicial damages or cost-benefit regulatory decisions.[119] To the extent that certain nonuse values do not belong the measure of damages, inclusion constitutes a systematic deviation from the true valuation. What can we say about the significance of this "error"? Mitchell and Carson show that markets for existence value do exist (through voluntary contributions to activist groups, or to political lobbying), but are incomplete.[120] The *Ohio* court said in dictum that "option and existence values may represent 'passive' use, but they nonetheless reflect utility derived by humans from a resource, and … ought to be included in a damage assessment."[121] Departing somewhat from the teaching of *Ohio*, today's CERCLA regulations permit assessment of use and nonuse values, but nonuse values are permitted only if use values cannot be determined.

      There is a strong presumption (in a nonlegal sense) among economists that nonuse values, though perhaps a legitimate source of utility, are less legitimate or reliable than use values. If nonuse value is an argument in citizens' utility functions, though, there is no obvious reason why it should not be considered in judicial or regulatory processes. Some critics observe that, *ad absurdum*, "existence value" might be assigned to the "good" of keeping jobs in the United States or the good of maintaining French culture, say, and that assigning property rights to these goods per se would be silly. The difference here may be that with tradable goods like American manufacturing or French film,

---

concern themselves with the "basic values of society," and because "the range of possible existence values may well be limitless," thus eroding "the power of economic analysis to discriminate between more- and less-efficient states of the world"). Another problem with existence value that I have not seen raised elsewhere is that the disutility from lost existence value is endogenous to news of the accident: one million people who hold some affection for a rare species will suffer disutility from the animal's extinction only if they find out about it. How much harm will be caused by an environmental injury depends on how much the event is publicized; and thus the plaintiff's lawsuit itself may help to create the need for compensation. In these situations, it may be that optimal policy involves procedural rules to encourage a quiet settlement. Robert Sugden, *Public Goods and Contingent Valuation*, *in* VALUING ENVIRONMENTAL PREFERENCES 170, 170 (Ian J. Bateman & Kenneth G. Willis eds., 1999).

      118. *See, e.g.*, Richard B. Stewart, *Liability for Natural Resource Injury: Beyond Tort 239*, *in* ANALYZING SUPERFUND: ECONOMICS, SCIENCE, AND LAW (Richard L. Revesz & Richard B. Stewart eds., 1995) ("[T]he argument that omission of nonuse values would lead to serious undercompensation of the public for injury rests on a category mistake. If nonuse values are principally noneconomic, then their omission from an economic measure of damages does not make the resulting damages inadequate.").

      119. *See* Raymond J. Kopp, *Why Existence Value Should Be Used in Cost-Benefit Analysis*, 11 J. POL'Y ANALYSIS & MGMT. 123, 126 (1992).

      120. MITCHELL & CARSON, *supra* note 3, at 63.

      121. Ohio v. U.S. Dep't of Interior, 880 F.2d 432, 464 (D.C. Cir. 1989).

existence values could show up in market prices through extra demand, but with nontradables like public environmental goods, the valuation for uses cannot capture the valuation for option or existence—and so option and existence value should be considered separately. This debate is somewhat orthogonal to the question of WTP vs. WTA, with the following caveat: given that "existence value" or "option value" cannot be acquired or increased in any real-world sense if the environmental amenity is extinguished, WTP makes even less sense when we seek to estimate nonuse values.

Before determining what type of valuations a court ought to consider, a more basic question is where individual valuations come from. In neoclassical theory, every individual has a complete and transitive preference relation over goods, which, combined with a budget constraint, leads to "willingness to pay" (a compensating surplus) for acquiring any good of any quantity. Environmental valuation fits into this model only if a coherent preference relation can be identified over environmental goods. Sugden, for example, says that individuals might not have a primitive preference map that they "consult" or "retrieve" with each decision; rather, they make decisions on the spot. As a result, a series of separate responses may not show transitivity or other marks of consistency.[122] Relatedly, it may be that over time, valuation comes from a loop of decision-making, feedback, updating of beliefs and preferences, and new decisions. As a result, not all observed choices correspond to a person's actual, exact preferences: sometimes the person gets it too high, and sometimes too low. Because respondents in CV surveys have little opportunity for feedback, a respondent never gets a chance to adjust his reported valuation to his "true" value.[123] In this light, Blomquist and Whitehead show experimentally that providing more information about the "quality" of an environmental good can positively influence respondents' WTP.[124]

---

122. Sugden, *supra* note 117, at 170; *see also* Gregory et al., *supra* note 13, at 463 (arguing that "for many types of environmental resources, particularly those that have been unfamiliar or have not previously been the subject of trades, stable economic values may not exist. Instead, they are constructed at the time of elicitation"); Schkade & Payne, *supra* note 23, at 103-05 (arguing that decision-making is necessarily "constructive" because the problem is multidimensional and because respondents may not know in advance how they will feel upon receiving a novel and unfamiliar good; and furthermore that CV responses will be highly sensitive to the context of questioning).

123. Milgrom, *supra* note 84, at 424-26; *see also* Note, *Ask a Silly Question . . . : Contingent Valuation of Natural Resource Damages*, 105 HARV. L. REV. 1981, 1985 (1992).

124. Glenn C. Bloomquist & John C. Whitehead, *Resource Quality Information and Validity of Willingness to Pay in Contingent Valuation*, 20 RES & ENERGY ECON. 179 (1998).

If CV is not capturing a carefully reasoned valuation, what, then, is it capturing? Tautologically, we can say that CV is capturing individuals' first reaction to a difficult and unfamiliar decision problem; and the instinct is then to dismiss CV as not reliable. On the other hand, it is not clear that individuals have a prior preference relation for every decision that they confront in a regular day, such as whether to purchase a new brand of shampoo.[125] The CV question process may force respondents to spend at least as much (and perhaps more) time thinking and learning about the problem than they usually do when, say, purchasing food at the grocery store.[126] It is easy to imagine that grocery purchases are subject to many of the same infirmities as CV valuation decisions: taking clues to value from a market price; making snap decisions about a new, unknown product; and perhaps even undervaluing due to free-riding (a consumer might buy less guacamole than she really wants, believing that she will be able to enjoy guacamole at her employer's holiday party; meanwhile, the employer buys enough guacamole to fill the bowl, rather than averaging over the partygoers' carefully measured preferences).

Consider just the first of these infirmities. That value should be calculated based on primitive preferences and divorced from any strategic interaction also implies that, normatively, market prices should not matter in calculating value. However, stated values may not be independent of market price in practice, either because respondents do not have a well-defined valuation; or else because they are not familiar with how to extract a maximal value, independent of price, from their psyche. Loomis et al. wonder whether CV respondents may be playing a game of "The Price Is Right," i.e., trying to guess what a fair market price would be, rather than accurately stating their maximal WTP. In a CV-based experiment, they found that including a "reminder" for respondents to report actual WTP, rather than what they believe the good might sell for in a store, significantly lessens the correlation of stated hypothetical WTP with (perceived) market price.[127]

---

    125.   W. Michael Hanemann, *Valuing the Environment Through Contingent Valuation*, 8 J. ECON. PERSP. 19, 28 (1994) (arguing that consumers "make up a decision rule at the moment they need to use it").

    126.   Carson, Flores & Meade, *supra* note 72, at 178.

    127.   *See* John Loomis, Thomas Brown, Beatrice Lucero & George Peterson, *Improving Validity Experiments of Contingent Valuation Methods: Results of Efforts To Reduce the Disparity of Hypothetical and Actual Willingness To Pay*, 72 LAND ECON. 450, 460 (1996). If CV respondents, and citizens thinking about environmental quality more generally, behave as Bayesians, then the usual model of aggregating individual valuations to get aggregate demand goes out the window. If the market price, or the dichotomous-choice price point given in the CV survey, is a clue to the true quality of an environmental public good, and the number of other people willing to pay for the good is also a clue of quality, then it is easy to see that the population

The potential correlation of CV responses with the known market price, or alternatively the anchoring effect, seems much likelier in a WTP context than in a WTA context. Indeed, precisely because it seems closer to a market-like purchase decision, the WTP question format is likelier to invoke a mental process of linking value to a perceived price. The WTA format, because it is essentially a question about selling rather than buying (the former being something that consumers rarely do) is less likely to create anchoring problems. A critic might answer that WTP should be favored precisely because it resembles a real-life decision that consumers or citizens face every day. However, paying pivotal dollars for provision of a public good that you actually get for free in real life may be so unusual a scenario that respondents could more naturally comprehend the WTA question: if you actually owned a property right, for how much would you sell it?

Besides the independence of valuation from price, elicited valuations would ideally be independent of information provided by the questioner. (Certainly, valuations should depend on information generally available to and known by the respondent, but information learned directly due to the survey should not be determinative.) It is reasonable to assume, though, that in a real-life situation of environmental damage, the disutility a person suffers could be augmented or created by news of the event.[128] Milgrom objects to a purely utilitarian-based conception of existence value, on precisely these grounds: it would be an absurd result if "the secret destruction of an environmental resource does no damage . . . [but] real damage is wrought

---

(and perhaps the sample in a CV study) could coordinate on several possible valuations, regardless of the intrinsic features of the environmental amenity. Also, Bayesian thinking could lead to a multiple-equilibria situation whether either very few citizens decide that the good is worth providing, or many citizens decide that the good is worth providing.

    128. *See* Daum, *supra* note 102, at 393-94 ("[T]he values people hold for environmental resources obviously change over time and may be affected massively by publicity from the accident itself. . . . But the very fact that people are being educated means that their values and preferences are changing; if one tries to measure those values *after* they have been changed, as contingent valuation necessarily does when it conducts surveys in the wake of a major environmental event, one is not measuring the values that were held at the time, and one is therefore not measuring anything that *has been lost*. . . . If (as will often be the case) the respondents have no previous knowledge of the resources, the survey itself may cause them to form, or to report, a preference. If some or all of the information supplied is new to the respondents, the survey may cause existing preferences, based on other information, to change."); Diamond & Hausman, *supra* note 21, at 10 ("[T]he knowledge issue creates a clear difficulty in attempting to measure compensatory damages for the loss of nonuse value when an individual learns simultaneously about a resources existence and about an injury to it, for example, learning about the existence of Prince William Sound *and* its beauty and the oil spill in the same news report. The change in well-being when a known resource is injured is not the same as that which occurs when one learns simultaneously about the existence of a resource and an injury to it.").

by the journalist who first publicizes the destruction."[129] Milgrom instead recommends a theory of existence value that "depends as much as possible on the actual state of the environment" and does not depend on the respondent's completeness of information.[130]

It is for these reasons that WTA looks all the more attractive than WTP. Unlike WTP, WTA is based on a preexisting property right, and so can be measured without giving any extra information to respondents. It is true, as discussed above in Part V.B, that a WTP question could be framed as "How much would you pay to avoid the destruction of this environmental amenity?" and could be adjusted to discern the respondent's level of background knowledge of the already-possessed environmental amenity. However, this type of questioning is likely to provoke angry "protest" answers: who wants to give up money to save what one already has a claim on?

VII. WHAT KIND OF EVIDENTIARY RULE?

A.   *Judges and Juries*

In this Article I have advocated a new norm for CV practitioners, federal courts, and relevant governmental agencies to use in dealing with CV evidence. Agencies including DOI and NOAA currently have great discretion under their regulations about how exactly to proffer "reliable" evidence for use and nonuse values. To the extent that agencies choose to use CV evidence (which has been unusual, perhaps due to the unsettled evidentiary position of the technique), agencies should strive to use a WTA approach rather than WTP—and agencies can regularly do low-cost "diagnostic" studies to determine the extent of citizen interest in particular types of environmental goods.

It makes sense that the offerors of evidence should use the theoretically appropriate survey form, but do courts need to guard their evidentiary gates so vigilantly? One alternative approach might simply be to give juries any CV evidence that the plaintiff or defendant can bring forth, and let the jurors sort out on their own what the correct valuation should be. If WTA indeed systematically exceeds WTP (although this relationship could be reversed, as the discussion in Part V.B showed), then the plaintiff would provide the WTA estimates and the defendant would either try to undermine the CV technique or else provide WTP estimates, and the jury could arbitrate the controversy.

---

129.   Milgrom, *supra* note 84, at 419.
130.   *Id.*

    The problem with allowing WTP-type evidence is that any such evidence is inherently prejudicial, within the meaning of FRE 403. This inherent prejudice is true of most scientific evidence prone to error, but in the case of WTP, the evidence is not just biased, but categorically inappropriate. If, as is usual, the normatively correct measure of damages is WTA, but jurors fail to understand this, then WTP, while perhaps giving jurors a "ballpark" estimate of the order of magnitude of valuation, leads them away from the calculation they should be making. It would be better to let both plaintiff and defendant provide competing estimates of WTA, and let the jurors decide which CV study is more credible. This is not to suggest that WTP studies must always be excluded as a matter of law from environmental injury cases in federal courts. If there is no other available evidence, then a carefully undertaken WTP-type CV study could be acceptable. However, judges should be very careful to balance the prejudicial nature of WTP-type evidence in environmental damages cases against the necessity of using such evidence when plaintiffs have nothing else. If a party in a natural resources damages suit had the opportunity to undertake a WTA survey after the cause of action arose, but instead executed a WTP survey, that evidence should be excluded as prejudicial.

    Alternatively, we could bypass CV studies entirely and let a (presumably Bayesian) jury repeat the CV exercise on their own. In this light, Kanner and Nagy recommend allowing juries to use their intuition to quantify nonmarket or nonuse environmental values, just as juries presently do in tort cases for pain and suffering, nuisance, or damage to unique goods.[131] They suggest that rather than hearing CV evidence, juries should hear only a description of relevant physical facts, including the prepollution condition of the environmental object, the nature and extent of the release, the effect on users of that environmental object, and the public significance of the environmental object.[132] The jury would then decide on the money value of this loss.[133]

    Leaving valuation decisions entirely to the jury is suboptimal for a variety of reasons. First, it is unlikely that a jury represents a random sample of the underlying population.[134] If the case involves many thousands or millions of affected citizens who suffered a loss of

---

    131. Allan Kanner & Tibor Nagy, *Measuring Loss of Use Damages in Natural Resource Damage Actions*, 30 COLUM. J. ENVTL. LAW 417, 424 (2005).
    132. *Id.* at 440.
    133. *Id.*
    134. *See, e.g.*, David Kairys, Joseph P. Kadane & John P. Lehoczky, *Jury Representativeness: A Mandate for Multiple Source Lists*, 65 CAL. L. REV. 776 (1977).

environmental quality, it is important for correct deterrence and/or compensation that the judicial award is based on the average citizen's valuation. Unlike a personal-injury lawsuit, say, where the goal is for a jury of the victim's peers to assess the value of the victim's pain and suffering, the goal of an environmental damages lawsuit is to establish the loss felt by the average citizen. Also, juries may be no more precise than CV studies at considering the value of a hypothetical loss.[135] A University of Chicago study of jury damage awards vs. judges' "shadow" awards in the same cases found that jury awards averaged 20 percent higher.[136] A Condorcet-type argument suggests that jury awards should be more precise, but the higher mean award points in the other direction: the group dynamics in juries may lead to an upward bias. Finally, the economic significance of the distinction between WTA and WTP is subtle, and lay juries may not fully understand it. An instruction by the judge for juries to consider WTA rather than WTP could help to solve this problem, but this essentially asks jurors to run their own CV studies amongst each other, with each juror serving as both questioner and respondent. As the Arrow-Solow report suggests, CV studies are much cleaner when the questioning is uniform.

The most reasonable solution to the CV problem might be (as presently happens, though under poorly defined standards) to let each judge decide, case-by-case, when a particular CV study meets the four *Daubert* factors for admissibility of expert testimony. Dobbins recommends allowing CV studies into evidence, but giving the judge discretion to exclude it when the survey has methodological problems or the valuation estimate from the survey seems "irrationally large."[137] Although this proposal puts a fair amount of faith in the ability of noneconomist judges to decide the testability of econometric techniques, a non-discretionary rule creates a greater risk of excluding too much evidence when plaintiffs, especially private ones, have no other recourse.

B.   *Two Bodies of Law Meet*

A final interesting question is the relationship between evidentiary rules and administrative law. Under the *Chevron* standard,[138] federal

---

135.   Miriam Montesinos, *It May Be Silly, But It's an Answer: The Need To Accept Contingent Valuation Methodology in Natural Resource Damage Assessments*, 26 ECOLOGY L.Q. 48, 73 (1999).

136.   Edith Greene, *On Juries and Damage Awards: The Process of Decisionmaking*, 52 L. & CONTEMP. PROBS. 225, 228 (1989).

137.   Dobbins, *supra* note 88, at 933.

138.   Chevron v. NRDC, 467 U.S. 837 (1984).

courts should defer to federal agencies' regulatory interpretations of statutes when the statutory language is unambiguous, and otherwise should uphold an agency's reasonable decision. And under section 706 of the federal Administrative Procedure Act, coupled with the *State Farm* doctrine,[139] a reviewing court should overturn any agency action found to be arbitrary and capricious, i.e., lacking sufficient basis and explanation. In the case of contingent valuation, agencies like DOI and NOAA are interpreting statutes like CERCLA and the Oil Pollution Act in deciding to use certain types of CV studies. Presumably an aggrieved defendant could challenge the agency's use of CV under either *Chevron* or *State Farm*—but this decision must be independent of the decision about whether the evidence is admissible under *Daubert*. In other words, the CV evidence must pass both tests. The rebuttable presumptions contained in the DOI and NOAA regulations will not be operational if the CV evidence cannot get into court in the first place. In any case, agency use of CV evidence probably would pass either a *Chevron* or *State Farm* test, given the extensive research record that the agencies have developed over the years in this area, highlighted by the Arrow-Solow report.

　　Finally, should the admissibility standard in statutory-based cases such as CERCLA be any different from the evidentiary standard in private tort actions? The answer must be yes in one sense: the trustees' evidence in CERCLA cases already gets a rebuttable presumption by regulation, which does not obtain in private environmental lawsuits. However, a federal court applying *Chevron* (or even *State Farm*) deference and upholding DOI's particular interpretation of the statutory charge for "reliable" evidence must also ensure that the proffered CV evidence passes the various *Daubert* criteria, as outlined in this paper. If a particular class of CV techniques (such as a dichotomous-choice WTA survey) is deemed to pass the *Daubert* test in a CERCLA case, then it should pass the same test in a private action in federal court: CV evidence should be treated exactly the same whether the lawsuit is public or private. Outside of some judicial discretion in very close cases, there should be no need for a private plaintiff or defendant to point to a different federal court's decision to uphold DOI's CV evidence in a CERCLA case as reason for admission of his own CV evidence.

　　The interaction between *Chevron* deference and *Daubert* rules would become significant only if *Daubert* turned out to be a *looser* evidentiary standard than the federal regulations governing CV. In that

---

139.   Motor Vehicles Mfrs. Ass'n v. State Farm Mut., 463 U.S. 29 (1983).

case, CV evidence might be admitted in a private action but not in a statutory (public) action.[140] Still, this should not be too disturbing. Just as we have different standards for admissibility of evidence and burdens of persuasion in criminal trials vs. civil trials, an environmental tort lawsuit motivated by Congressional action should not be taken as the same thing as an action between private parties. It makes sense to impose a tighter evidentiary standard when the government is a plaintiff.

VIII. CONCLUSION

    For several years the consensus among economists has held that contingent valuation evidence is either inherently unreliable, or else should be restricted to "willingness to pay" evidence rather than willingness to accept. The guidelines put forward by Solow and Arrow, among others, have shown that CV can be valid, reliable, and testable (with resulting estimates of error rates) in appropriate circumstances. However, the received wisdom recommending willingness-to-pay questioning rather than willingness-to-accept is based on flawed assumptions about the normative purpose of environmental regulation. WTP could be justified because it generates lower damage awards, and it sounds like a more natural method of valuing everyday objects. However, compared to the WTA format, the WTP format is normatively less appropriate for natural resources damage assessments; less easily testable; and likelier to lead to bias. Federal regulations implementing CERCLA and the Oil Pollution Act, along with federal judges deciding admissibility of evidence in private tort suits, should begin to encourage the use of WTA; and at the same time, concomitant with an increasing use of WTA, economists should begin to carry out empirical tests of the WTA method by examining real "market" data on the "$ES_{0,1}$" type of trades outlined above. With this data, judges can begin to determine whether willingness-to-accept CV studies have any systematic error rates; if so, juries should be instructed accordingly. Thus, CV evidence can be admitted under all the factors of *Daubert.*

    Asking respondents to sell their right to environmental quality puts them in a very unfamiliar position, because citizens are not usually in the

---

    140. If *Daubert* were a *stricter* evidentiary standard than the regulatory rules for CV, then we would not expect any inconsistent results on CV admissibility in non-statutory private lawsuits (which are governed only by *Daubert*) vs. CERCLA actions (which are controlled by the regulations and *Daubert*). Under such conditions, if a proffered CV survey were to pass the *Daubert* test, then it surely would pass the regulatory requirements, and thus it would be admissible in both the private lawsuit and the public lawsuit. If the CV survey did not pass the *Daubert* test, then it would not be admissible in either type of lawsuit.

business of selling. But then, they are not in the business of buying cleaner water either. Seeking the selling price in a judicial action gets us closer to the goal of making victims whole and stopping the damage in the first place.