Journal of Economic Perspectives—Volume 26, Number 4—Fall 2012—Pages 27–42

Contingent Valuation: A Practical Alternative when Prices Aren't Available

Richard T. Carson

A person may be willing to make an economic tradeoff to assure that a wilderness area or scenic resource is protected even if neither that person nor (perhaps) anyone else will actually visit this area. This tradeoff is commonly labeled "passive use value," although it is also known by other names including "existence value" and "stewardship value." As Krutilla (1967) explained in his classic *American Economic Review* piece, "Conservation Reconsidered," passive use is not generally revealed by choices in the marketplace, unlike many local public goods which are either capitalized into property values or which require the consumption of complementary private goods to enjoy. Passive use represents the quintessential pure public good in that exclusion is not possible, nor even desirable, because enjoyment is nonrivalrous. The concept of passive use has played an increasingly important role in economic thinking concerning the value of public goods, and particularly, those involving environmental and natural resource amenities where passive and direct use values are often thought to coexist (Freeman 2003). In Carson, Flores, and Mitchell (1999), my coauthors and I provide a detailed overview of passive use value.

Without market information, other strategies must be considered to develop measures of economic tradeoffs that involve passive use value. For example, passive values can be captured through a single-issue referendum, but popular votes on ballot propositions that relate to these types of concerns are nonexistent at the national level and infrequent at the state or local level. However, more than a half-century ago, early studies of public goods like Bowen (1943) and Ciriacy-Wantrup (1947) drew the implication that, when the ballot box is not available, demand

■ *Richard T. Carson is Professor of Economics, University of California, San Diego, La Jolla, California. His email address is rcarson@ucsd.edu.*

http://dx.doi.org/10.1257/jep.26.4.27. doi=10.1257/jep.26.4.27

for public goods might be estimated through an appropriately structured survey of a representative sample of the public—in effect, what we now call a contingent valuation survey. Bowen went so far as to argue: "The polling of a "scientifically" selected sample might produce more accurate results than general voting, unless arrangements were made to insure that every person would actually vote."

Contingent valuation studies ask questions that help to reveal the monetary tradeoff each person would make concerning the value of goods or services. In Carson and Louviere (2011), my coauthor and I provide a common nomenclature for such "stated preference" questions. Such surveys are a practical alternative approach for eliciting the value of public goods, including those with passive use considerations. Thousands of contingent valuation studies have been done in over 130 countries looking at cultural, environmental, health, transportation, and other issues (Carson 2011). Almost 60 percent of the estimates in the very large Environmental Values Reference Inventory (EVRI) database maintained by Environment Canada in conjunction with the U.S. Environmental Protection Agency and the environmental agencies of several other countries come from contingent valuation (at https://www.evri.ca). The U.S. Environmental Protection Agency's (1994) estimates of the benefits of the U.S. Clean Water Act—which largely comprises recreation and passive use—is derived using contingent valuation, as are the benefits of individual regulations targeted at specific industries or water bodies (Griffiths et al. 2012). Results from contingent valuation studies are used for many purposes in benefit–cost studies: recent examples include the willingness to pay of Pennsylvania households for additional incarceration versus a rehabilitation program for serious juvenile offenders (Nagin, Piquero, Scott, and Steinberg 2006); the willingness of Lexington, Kentucky, residents to pay higher taxes to help support the construction of a new baseball stadium and basketball arena (Johnson and Whitehead 2007); the value of developing vaccine policies in Africa (Jeuland, Lucas, Clemens, and Wittington 2009); estimating the hourly value of informal care givers in the Netherlands (de Meijer, Brouwer, Koopmanschap, van den Berg, and van Exel 2010); looking at willingness to incur higher water tariffs for less river pollution in Fuzhou, China (Jiang, Jin, and Lin 2011); and the willingness of the U.S. public to pay for climate change measures (Aldy, Kotchen, and Leiserowitz 2012).

This essay begins by discussing the events set in motion by the *Exxon Valdez* oil spill of March 1989, focusing on why it is important to measure monetary tradeoffs for goods where passive use considerations loom large. Although discussions often seem to put much of their emphasis on whether contingent valuation is sufficiently reliable for use in assessing natural resource damages in lawsuits, it is important to remember that most estimates from contingent valuation studies are used in benefit–cost assessments, not natural resource damage assessments. Those working on benefit–cost analysis have long recognized that goods and impacts that cannot be quantified are valued, implicitly, by giving them a limitless value when government regulations preclude certain activities, or giving them a value of zero by leaving certain consequences out of the analysis. Contingent valuation offers a practical alternative for reducing the use of either of these extreme choices. I put

forward an affirmative case for contingent valuation and address a number of the concerns that have arisen.

Events Set in Motion for Contingent Valuation by the Exxon Valdez Spill

Soon after the *Exxon Valdez* spill in March 1989, the state of Alaska funded a contingent valuation study, contained in Carson, Mitchell, Hanemann, Kopp, Presser, and Ruud (1992), which estimated the American public's willingness to pay to avoid an oil spill similar to the *Exxon Valdez* at about $3 billion. The results of the study were shared with Exxon and a settlement for approximately $3 billion was reached, thus avoiding a long court case.[1] Our Carson et al. (1992) $3 billion estimate based on passive use dwarfed the Hausman, Leonard, and McFadden (1995) $4 million dollar estimate of the direct economic losses from lost recreation days in Prince William Sound, illustrating the importance of compensating the public for lost passive use.

In the aftermath of the *Exxon Valdez* oil spill, the U.S. Coast Guard put into place a version of the comprehensive plan for preventing oil spills put forward in the Carson et al. (1992) study. It was based to a large degree on the original risk assessment for shipping oil out of Alaska that had predicted one major spill every ten years from an accident eerily similar to that of the *Exxon Valdez* in the absence of risk-reducing measures (Moore 1994; Carson, Mitchell, Hanemann, Kopp, Presser, and Ruud 2003). In the years prior to the accident, some of the main safety requirements had been abandoned because they seemed "expensive" and unnecessary. One of these was that tankers have "escort tugs." Soon after key elements of the plan in our Carson et al. (1992) study were put into place, another supertanker lost power in the Straits of Valdez and drifted toward a reef near the one hit by the *Exxon Valdez*. One of the plan's new escort tugs pushed the supertanker away from the reef while the other tug shot it a towline. Since then, escort tugs have had to take control of a tanker in Prince William Sound three other times, with the latest being ExxonMobil's *SeaRiver Kodiak* in 2010.

Moreover, recognizing the potentially large passive use costs from oil-related activities led to other changes. The U.S. Oil Pollution Act enacted in 1990 required that tankers held by shell companies without large financial assets carry a $1 billion dollar insurance policy and required those shipping oil to develop comprehensive plans to respond to potential oil spills. These actions significantly reduced the

[1] By law, the government trustees must spend any money received for harm to its resources on restoration or acquisition of like resources. Exxon spent approximately $2 billion on response and restoration and $1 billion on natural resource damages, which the government used acquiring like resources. If the case had gone to trial, major contested issues would have included whether Admiralty law limited Exxon's liability; how much of Exxon's expenditures on response and restoration were on response, which did not count toward Exxon's liability, and on restoration, which did; and whether contingent valuation could be used to establish the loss to the public from the spill. Note that losses to commercial fishing are the subject of private, not government, legal claims.

frequency and severity of oil spills from tankers in the United States relative to the rest of the world (Chapple 2000). In contrast, Congress granted offshore wells like the BP *Deepwater Horizon* well a liability limit of $75 million dollars, although unlimited liability applies in the case of gross negligence, willful misconduct, or violating a federal regulation directly related to the spill. (In the aftermath of the March 2010 oil spill, BP waived the offshore well liability cap.)

In the aftermath of the *Exxon Valdez* settlement, Exxon put on a conference where research it had sponsored in preparation for the case was presented (see the volume edited by Hausman, 1993). Its overall conclusion was that contingent valuation was unreliable. In response, the U.S. Department of Commerce assembled a blue-ribbon panel of experts chaired by Kenneth Arrow and Robert Solow to examine contingent valuation. Their report (Arrow, Solow, Portney, Leamer, Radner, and Schuman 1993), known as the NOAA Panel Report, was cautiously supportive, finding that "well conducted CVM [contingent valuation method] studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive values." The panel also set forward an influential set of guidelines for conducting contingent valuation studies. The two conflicting views encouraged a large amount of theoretical, econometric, experimental, and empirical research on contingent valuation.

Economists, Survey Data, and Contingent Valuation

Economists are naturally skeptical of data generated from responses to survey questions—and they should be! Many surveys, including contingent valuation surveys, are inadequate. Whittington (2002), one of the pioneers of contingent valuation studies in developing countries, has lamented the tendency to implement quick and cheap studies that are likely to yield flawed results. He notes that "we are still a long way from the point where it is possible to do high-quality CV [contingent valuation] surveys with minimal effort or expense." This situation is no different from many other areas of economics that are heavily dependent on survey-based data—income, consumption, education, employment, health status, and so on—but it is sometimes less obvious because economists are often not actively involved in how their data is collected and often have no formal training in survey research.

A good contingent valuation survey is a very different process than the mental image some readers may have of a researcher walking up to people in a shopping mall and asking how much they would pay to save a sea otter. For an example of a real-world contingent valuation survey, interested readers might start at the bottom of my web-page at http://www.econ.ucsd.edu/~rcarson, where they can download and examine the survey instrument for the *Exxon Valdez* study. It is 39 pages long, plus 14 pages of maps and photos, and 10 pages of show cards and figures.

Overall, a well-designed contingent valuation survey must convey to respondents that the government is considering implementing a policy and that their responses to the questions in the survey will be used to help inform that decision. The survey describes the problem that is the focus of the survey and the plan that the government

is considering to address that problem. For complex policies, and particularly, those issues without a lot of previous survey work on closely related policies, the process of developing a survey can be lengthy: it frequently involves focus groups, cognitive interviews, pretests, and pilot studies. Details matter: the survey should be designed so that the plan is seen as an effective response to the problem. This must be done in a way that respondents without a high school degree can understand. The presentation typically involves graphics intended to help people understand the problem and the government's plan. Survey respondents need to understand that if the plan is implemented, it would be paid for using a coercive payment mechanism, typically some type of tax or utility bill; each respondent must be convinced that a mechanism exists that would ensure they would pay in that case. Because it is impossible to get all members of the public to accept all details of the scenario, it is standard practice to ask a sequence of "debriefing" questions to help gauge the likely impact of scenario rejection.

Much of the usefulness of doing a contingent valuation study has to do with pushing scientists and engineers to summarize what the project would do in terms that the public cares about. Further, the process of developing a contingent valuation survey often encourages earlier involvement by policymakers in thinking more critically about a project's benefits and costs and in considering options with lower costs or greater benefits to the public.

As long as respondents believe that there is a positive probability that the government will take the results of the contingent valuation survey into account, they should use the opportunity to influence the government's decision. In Carson and Groves (2007), my coauthor and I demonstrate that the response to a properly formulated binary discrete choice question represents "consequential" economic behavior; and, that the incentive properties of such survey questions with respect to economic behavior are identical to those of a binding ballot proposition.[2] In this sense, responses to a good contingent valuation study can reasonably be treated as *revealed economic behavior,* akin to that obtained in a vote of a representative population on a ballot proposition.

Neoclassical Economic Theory and Contingent Valuation Results

Predictions from simple versions of neoclassical economic theory can sometimes differ from outcomes found by contingent valuation surveys. Of course, predictions from simple versions of neoclassical theory can also differ quite a bit from observed real-world behavior, as the literature on "behavioral economics" has pointed out (DellaVigna 2009). Contingent valuation surveys are designed so that

[2] In Carson and Groves (2007), we show that the auxiliary conditions needed for truthful preference revelation to be a dominant strategy are that people can be compelled to comply with the payment provision of the scenario irrespective of the outcome and that the scenario offers a take-it-or-leave it choice that does not influence future offers. These conditions are the same for binding votes and advisory surveys.

the findings will reflect actual behavior, so it should be no surprise that the same behavioral influences on what people do in markets show up in surveys. Indeed, some of the best-known insights of behavioral economics were first demonstrated in contingent valuation surveys.

In their overview paper in this issue, Kling, Phaneuf, and Zhao detail the ways researchers have addressed objections that the findings of contingent valuation surveys appear incongruous with simple versions of neoclassical economic theory. My summary here can be limited to some key examples. In Carson, Flores, and Meade (2001) and Carson and Hanemann (2005), my colleagues and I offer a more detailed discussion of these issues.

One of the most persistent of the claims that contingent valuation surveys are unreliable points to a discrepancy between willingness to pay and minimum willingness to accept compensation for the same nonmarket good. This finding should actually not be a surprise, either in terms of neoclassical economic theory or in terms of behavioral economics. The predicted properties of welfare measures are often quite different for 1) cases where everyone will experience the same level or quantity of the public good, and 2) cases involving price changes where consumers can determine the amount of the good they wish to consume. Hanemann (1991) shows willingness to pay and willingness to accept for a pure public good are likely to be quite far apart, which stands in stark contrast to Willig's (1976) well-known result that willingness to pay and willingness to accept for *a price change* should typically be close together. Hicks (1943) correctly saw that welfare measurements involving rationed goods (of which pure public goods are a special case), so-called "surplus" measurements, are fundamentally different from the "variation" measures which typically involve price changes. The underlying reason is that in the price-change case, the magnitude of the difference between the two welfare measures is governed by an income elasticity, thought to be of reasonable magnitude; meanwhile for a quantity (or quality) change, the difference is governed by the ratio of this income parameter to a Hicksian composite substitution parameter between the good and marketed goods, often thought to be small in magnitude. Kling, Phaneuf, and Zhao point out that enriching the basic neoclassical framework by adding dynamic considerations also tends to drive willingness to pay and willingness to accept measures apart. From a behavioral economics view, the divergence between willingness to pay and willingness to accept is a core prediction of Kahneman and Tversky's (1979) prospect theory because whether a choice is framed in terms of gain or loss influences behavior. Again, the upshot here is that willingness to pay and willingness to accept are often not equal, whether in contingent value surveys or in market-based tests; indeed, Horowitz and McConnell's (2002) meta-analysis shows that the differences are similar in both settings.

Two other situations where it is often asserted that contingent valuation studies produce anomalous results involve estimates of income elasticities and sequence effects. The first contends that if contingent valuation studies were valid, then the estimates of the income elasticity of willingness to pay for the environment should be greater than one, because the environment is a luxury good. The main difficulty here (ignoring the plausibility of the luxury-good assumption for these goods and

the likelihood of measurement error in income) is that the income elasticity of willingness to pay is a very different statistic than the income elasticity of demand, upon which an economist's usual definition of a luxury good is based. In Flores and Carson (1997), my coauthor and I show the two elasticities are functionally related, but under most plausible assumptions, the income elasticity of willingness to pay should be considerably smaller than the corresponding income elasticity of demand.[3]

The second assertion is that large differences in the measured value of a good depending upon the sequence of other goods that were also valued in the same survey indicate that contingent valuation is unreliable. However, the basic theory of income and substitution effects suggests that sequence effects should occur. In a willingness-to-pay sequence of k goods, keeping utility constant requires that the agent give up money at each order in the sequence as a new good is acquired, whereas in a willingness-to-accept compensation sequence, keeping utility constant requires giving the agent money as goods are sequentially taken away. In Carson, Flores, and Hanemann (1998), my coauthors and I show that these sequence-related differences can easily be large: specifically, willingness-to-pay and willingness-to-accept sequences for imposed quantity changes involve a partially inverted demand system in terms of the Hicksian substitution terms, such that, if sequence order differences are small in price space, they will typically be large in terms of differences in welfare measures. Thus, a good valued first in a willingness-to-pay sequence will tend to be worth more than if it is valued "lower" in a sequence of possible projects.[4] In a way, this result should be no surprise: after all, in setting a political agenda, controlling the order in which projects are considered is thought to be extremely important.

A final criticism of the contingent valuation method is that different preference elicitation techniques often obtain different estimates of value, which has been taken by some critics as an indication that survey respondents do not have well-defined preferences for nonmarket goods. This finding has troubled contingent valuation researchers, although it is not unique to contingent valuation. Marketing researchers and experimental economists find the same phenomenon. Indeed, cognitive psychologists such as Tversky, Slovak, and Kahneman (1990) have argued that the divergence in economic values implied by framing decisions in terms of a choice rather than matching response is perhaps the fundamental problem with economic theory. A natural economic response to this issue is to study how different elicitation techniques should affect the answers given. Using neoclassical

[3] Specifically, the income elasticity of willingness to pay is equal to the income elasticity of demand times a matrix of Hicksian substitution terms scaled by the ratio of ordinary income to the sum of ordinary income and the implicit income from all public goods. By definition, this ratio is less than one and likely to be substantially less than one.

[4] There is some irony, though, that critics of the use of contingent valuation in natural resource damage assessments point to the substantial declines often seen in willingness to pay for a good as it is valued farther and farther out in a sequence as a reason not to use contingent valuation. Willingness to accept compensation is the theoretically correct welfare measure for harm from an oil spill (Arrow et al. 1993). Because willingness to accept is greater than willingness to pay for the same good valued first in a sequence and because the value of a good in a willingness-to-accept sequence is increasing in terms of sequence order, so willingness to pay for a good appearing first in a sequence is smaller than willingness to accept for the same good appearing in any sequence order.

mechanism design theory, in Carson and Groves (2007), we demonstrate that different elicitation formats have different incentive and information properties. Rational economic agents should be responsive to these properties in such a way that commonly used preference elicitation formats should produce different welfare estimates. This framework offers a comprehensive set of predictions concerning the characteristics of data collected using different preference elicitation methods and differences in welfare estimates obtained using them, and it has fundamentally changed how researchers view stated preference data (Poe and Vossler 2011).

In short, there are often divergences between predictions of simple neoclassical economic theory and actual behavior, as well as between that same theory and responses to contingent valuation surveys. In both cases, the most productive response is often to investigate both the theory and the data more carefully. Usually, a more realistic theoretical representation provides a reasonable guide to observed responses. But people are not perfect; their choices can reflect mistakes, which will be apparent if an analyst observes them under the equivalent of a microscope. At some point, a judgment has to be made as to whether to accept consumer sovereignty in the form of respecting choices involving the tradeoffs people say they are willing to make when they are observed in a context designed to facilitate careful decisions.

Sensitivity to Scope

The Arrow et al. (1993) NOAA Panel Report put forward a set of recommendations that largely followed the procedures used in the Carson et al. (1992) *Exxon Valdez* study, with one major exception. The Panel recommended that contingent valuations studies being done for litigation should pass a "split scope test." This test requires asking two separate subsamples of respondents about two different descriptions of a good, where the amount of the good along some quality or quantity dimension should make it clearly "larger."

The underlying concern here, voiced by Kahneman and Knetsch (1992) and Hausman (1993), was that respondents to contingent valuation surveys may have a certain amount that they are willing to spend on, say, environmental protection issues generally, and so they will tend to respond with this amount in mind regardless of the actual characteristics of the good being valued. An often-cited example is a contingent valuation study in which respondents to a self-administered shopping mall survey appeared willing to pay the same amount to save 2,000, 20,000, or 200,000 birds from being killed by oil (Desvousges, Johnson, Dunford, Boyle, Hudson, and Wilson 1993). However, in this study respondents were also told that the population of birds was very large, with the percent of birds being killed in the three split-sample treatments being similar: (a) "much less than 1% of the population", (b) "less than 1% of the population", and (c) "about 2% of the population. In short, the seeming insensitivity to scope shown by the respondents in this study is exactly what is likely to have been shown by many professional ecologists given this information. Hanemann (2008) replicates this study in the same shopping

mall context with two subsamples, one where 1 percent of the bird population is impacted and the other with 10 percent. He finds a sizeable statistically significant difference in willingness to pay between these two treatments.

Given the concern over the scope issue, new explicit scope tests quickly appeared, and several instances were also identified where past studies done for policy purposes had used split samples with goods differing in scope. In Carson (1997), I reviewed this literature and found 30 split sample tests which rejected the scope insensitivity hypothesis. Most of these involved goods where passive use was thought to be important. Two large state-of-the-art in-person surveys of the general public, one in the United States involving DDT deposits off the coast of Los Angeles (Carson et al. 1994) and one in Australia involving preservation of the Kakadu Conservation Zone (Carson, Wilks, and Imber 1994) included explicitly designed scope tests using goods where passive use considerations were thought to be the predominant source of value. Each of these surveys used identical descriptions of the local ecosystems involved, how the goods would be provided, and how they would be paid for, but one subsample was provided a good larger in scope than in the other subsample. In both cases, the scope insensitivity hypothesis is strongly rejected ($p < .001$) and willingness-to-pay estimates for the larger good in both cases are almost double that of the smaller good. The argument that scope insensitivity is a generic, unavoidable characteristic of contingent valuation studies has been shown to be false.

Of course, particular studies may show insensitivity to scope, and research has identified two main areas where this tends to occur. First, low-probability risks are often poorly understood in contingent valuation surveys, as they are by consumers in real-world behavior involving financial planning and insurance decisions. Various graphical representations have been shown to improve understanding in contingent valuation surveys (for example, Corso, Hammitt, and Graham 2001), and researchers are now looking at similar ways to assist consumers in making better financial planning decisions. Second, where a program is seen to provide multiple outputs, such as protecting different endangered species, it can be difficult to get distinct willingness-to-pay estimates for the individual outputs as opposed to the entire program.

While well-designed contingent valuation studies will typically pass a scope test, such tests have several conceptual problems that limit their potential usefulness. First, while contingent valuation critics sometimes contend that willingness to pay should be (almost) linearly increasing along some quantity dimension, declining marginal utility is more likely, which can influence the statistical power of scope tests (Rollins and Lyke 1998). Second, for a substantial fraction of the public, the likelihood of the government delivering on very large projects can be perceived to be much lower than that for smaller projects, in which case values placed on two goods may be entangled with beliefs about how well government functions. Finally, true willingness to pay may not even be monotonic in some instances. One can imagine a case, for example, in which a modest increase in the wolf population may be seen as a good thing, while a substantially larger increase is viewed negatively (Heberlein, Wilson, Bishop, and Schaeffer 2005). The time may have come to listen with an open mind to the message that survey respondents are seeking to convey when their answers suggest that changes in the scope of the good do not matter to them.

Contingent valuation critics sometimes also argue that the values of survey respondents must satisfy a more stringent sequential adding-up test whereby a composite good is broken into two parts and all three valued separately. Such a test is logically correct given its assumptions. But as Smith and Osborne (1996) point out, a key implicit assumption in natural resource damage assessment is that the replacement good is a perfect substitute. This requires, for example, that an agent be indifferent between saving a wild bird from being killed by oil and creating a hatchery program that produces a bird.[5] From a survey perspective, the adding-up test is problematic to implement because the survey for the second sub-components requires respondents to imagine they have received the first good and to imagine they have paid for the first good when asked about willingness to pay for the second good. Even putting moral and practical implementation objections aside, many people do not pass this adding-up test with market goods. Bateman, Munro, Rhodes, Starmer, and Sugden (1997) examined this experimentally, using students as the subjects and the highly familiar and frequently consumed goods of coffee and pizza. They fail the adding-up test. Many stores such as car dealers and cell phone providers are routinely successful selling customers additional goods and services *after* they purchase the car or phone that they were not otherwise going to purchase. Good contingent valuation studies do not engage in the survey equivalent of "upselling"; instead, they offer the complete bundle when the bundle is the relevant good for policy purposes.

Difficulties with the Hypothetical Bias Argument

Many economists instinctively think that the responses to contingent valuation questions will automatically overvalue people's true willingness to pay for public goods. In the context of contingent valuation surveys, this is called "hypothetical bias." Ironically, Samuelson (1954) saw the opposite problem in his classic article on public goods when he noted: "It is in the selfish interest of each person to give false signals, to pretend to have less interest in a given collective consumption activity than he really has," and he predicts that having the public complete "questionnaires" at different prices would fall prey to this strategic behavior. From Samuelson's view,

[5] The argument is sometimes put forward that anything that is put back physically cannot result in a loss in passive use value. This assumption is equivalent to denying the validity of a loss in utility from pain and suffering associated with a serious automobile injury, as long as the bones are eventually put back in place. In the context of an oil spill, it says agents cannot suffer a utility loss from knowing that animals suffer from being oiled, as long as the animal population and the ecosystem recovers. Because injuries to ecosystems often cannot be completely restored for any plausible cost, proponents of this assumption sometimes advocate the use of "habitat equivalency," a technique that translates loss in type of habitat into gains in another. This biological measure may be a reasonable proxy for small environmental injuries where the restoration or replacement is done in close proximity (on-site) to the original injury and involves very similar resources (in-kind). However, the approach breaks down for large-scale injuries. As an extreme example, the technique would allow the destruction of all wetlands in San Francisco Bay to be compensated for by restoring some amount of prairie grasslands in Nebraska. More important, perhaps, habitat equivalency has no direct tie to public welfare and, as such, should not be seen as a way of making the public whole.

those answering contingent valuation surveys about a public good should follow a free-rider approach of pretending to be less interested, hoping that the costs of providing the public good will fall on others. Which position does the empirical evidence support: the "hypothetical bias" prediction that surveys will overestimate true willingness to pay, or the Samuelson's prediction that strategic behavior will lead to an underestimate? The answer is "both." Survey exercises presented as purely hypothetical or having incentives encouraging overpledging can overestimate willingness to pay. However, contingent valuation surveys that are designed so that participants perceive them as consequential with a coercive payment mechanism and a reasonable set of auxiliary conditions (as discussed earlier) tend to, if anything, follow Samuelson's prediction of underestimating the true value, when they can be compared to other ways of calculating such values.

Studies of "hypothetical bias" are often done in an experimental context with students in which one group of subjects is told they will have to pay and another group of subjects are repeatedly told that responses are "purely hypothetical" in the sense of not having any effect on anything. The "hypothetical treatment" does typically lead to higher willingness to pay. In a meta-analysis of such studies, Murphy, Allen, Stevens, and Weatherhead (2005) find that the median ratio of estimated willingness to pay for purely hypothetical treatments to estimated willingness to pay in the actual payment treatments is 1.35, with a small number of very large outliers that drive up the mean ratio. Since a good contingent valuation study emphasizes the chance to influence whether the government will provide the good and the payment obligations if it is provided, it is not clear whether these purely hypothetical laboratory comparisons are of much relevance.

Another setting sometimes used to assert that contingent valuation suffers from "hypothetical bias" involves comparing actual contributions to a voluntary program to the propensity to contribute expressed in a survey context: that is, people say that they will contribute more in surveys than is actually contributed. This comparison has long been suspect. As we explain in Carson and Groves (2007), the most likely purpose for doing a survey asking about the likelihood of making a voluntary contribution is to help gauge whether to mount a fundraising effort. If the respondent wants the good, the optimal response is to appear ready to contribute in the survey to encourage the voluntary contribution campaign—and then to free-ride hoping others will contribute enough to provide the good. From this perspective, the economic puzzle then is not why the survey estimate is higher than actual contributions, but rather, why the difference between the two estimates is not larger.[6]

[6] Similarly, it has long been known that surveys of purchase intentions for new products in private markets tend to over-forecast actual purchases. In Carson and Groves (2007), we show that this result is theoretically predictable—respondents who potentially want to purchase the good should say "yes" to increase the likelihood that it is offered for sale, at which time they can then decide whether to buy. When people are surveyed about their likelihood of buying a private good that is already being offered for sale, respondents tend to have a lower propensity to buy in the survey than they do in markets. This effect is also in the predicted direction since respondents want to encourage firms to lower prices. For an interesting example involving existing toll roads where surveys underpredict usage, see Small, Winston, and Yan (2005).

With quasi-public goods, it is possible to compare estimates from contingent valuation studies with other ways of estimating values through some form of revealed preference for public goods. For example, the "travel cost method" involves people facing different travel costs for visiting a certain place—like a recreational fishing site. This price (in terms of travel costs) for going to that site can be used in conjunction with the number of trips to that site to estimate a demand curve and, in turn, willingness to pay for a trip to the site. The "hedonic pricing" method can be used with housing prices that incorporate spatially delineated amenities. Statistical methods can be used to control for other attributes of the home, like the number of bedrooms, in such a way that an estimate of the value of the environmental amenity can be obtained. The so-called "averting-behavior approach" looks at what people spend to avoid an adverse effect and allows the researcher to back out a derived demand for reducing it.

In Carson, Flores, Martin, and Wright (1996), we conducted a meta-analysis of 83 studies that included 616 comparisons of contingent valuation estimates to revealed preference estimates using these kinds of methods. We found that the mean ratio of contingent valuation to revealed preference estimates is 0.89 (with a 95 percent confidence interval of [0.81–0.96]), suggesting that contingent valuation estimates in the case of quasi-public goods are on average a bit lower than revealed preference estimates and reasonably correlated (0.78) with them.[7] Since that study, other papers have looked at valuing specific classes of goods using contingent valuation and revealed preference approaches and examined whether the details of the approach make a difference. For instance, the value of statistical life estimates from contingent valuation studies are of the same order of magnitude as those from hedonic wage studies using job risks, but smaller (Kochi, Hubbell, and Kramer 2006). Shrestha, Rosenberger, and Loomis (2007) perform a meta-analysis of studies on a large database of outdoor recreation valuation estimates. They find that contingent valuation estimates are significantly lower on average than comparable estimates based on revealed preference methods.

Yet another approach is to look at the small number of U.S. studies where a contingent valuation survey with (nearly) identical wording to an actual ballot proposition can be compared to the actual vote. Like many environmental goods, these ballot propositions tend to involve a mix of direct and passive use. As Kling, Phaneuf, and Zhao note, the comparisons between the contingent valuation estimates and actual votes are quite favorable, and they are clearly conservative when "don't knows" are treated as "no's" (the standard practice in the contingent valuation literature). This should not be surprising. Public polls taken near an actual vote, when the information set is unlikely to change, are on average quite good predictors of two-candidate races and ballot propositions. Predicting voter turnout

[7] The median ratio is somewhat lower at 0.75. There is a clear publication bias in studies comparing contingent valuation to revealed preferences estimates: published studies tend to either find a ratio of the two estimates close to one or a ratio that is very large. This two-humped distribution of published results suggests two very divergent expectations among economists, and that results can be cherry-picked to support a particular position.

is generally a harder task than predicting how people are going to vote conditional on the information they have.

Determining the Quality of a Contingent Valuation Study

A recurring theme of this essay has been that high-quality contingent valuation surveys appear to produce high-quality economic data. How does one separate the wheat from the chaff? Survey researchers point out that the most important thing to look at is the "face validity" of the entire contingent valuation survey instrument. Does the survey credibly pose a well-developed policy proposal to respondents and provide them with the necessary information to make an informed decision about it? Does the survey make respondents comfortable making a decision to either support or oppose the policy proposal and make them aware of the consequences if the policy is implemented? The best contingent valuation surveys are among the best survey instruments currently being administered while the worst are among the worst. In the hands of an expert in questionnaire design, face validity is not hard to judge. Economists are not typically trained with these skills so their judgments may need to be supplemented by those of people who do have the requisite training. Economists can judge whether a choice is consequential and when choices will reveal the desired tradeoffs.

Next, turn to the survey development effort. Ask whether adequate development and testing work was done in a deliberate, not pro forma, manner. Look at the survey administration and sampling. The Arrow et al. (1993) NOAA Panel Report recommended that surveys being done for litigation use in-person interviews with experienced professional interviewers to help motivate respondents to pay close attention to the details of the scenario, and that these surveys also have a rigorous sampling plan that is well executed. This is an enormously expensive undertaking, so it is here that one is most likely to see efforts to reduce cost. What are the implications of the survey implementation choices made?[8]

Now look at the basic results of the completed survey: Taking sampling error into account, does the percent of respondents willing to pay the randomly assigned cost amount fall as that amount increases? Is the estimate of willingness to pay derived using a statistical technique that is robust to assumptions about the far right tail of the distribution? Does the study present a construct validity equation that explains a reasonable amount (in a cross-sectional sense) of the heterogeneity in estimated willingness to pay and a comprehensive set of sensitivity analyses? Does the estimate from the study represent a sensible tradeoff that people might make to implement the policy in question?

[8] There is a lively debate in the literature over how to best deliver high-quality valuation estimates at lower costs, which is no surprise since the key question facing an agency doing a benefit–cost analysis is the value of spending a marginal dollar on a particular analysis and in allocating that dollar to one part of that analysis versus another.

Concluding Remarks

Contingent valuation is not perfect. No economic technique is. But the alternative to contingent valuation, especially in cases involving passive use considerations, is to place a zero value on goods that the public cares about—which is never likely to be the right choice.

In the two decades since the *Exxon Valdez* oil spill, the amount of research undertaken on contingent valuation has been substantial, including many thoughtful assessments starting with the Arrow et al. (1993) NOAA Panel Report by government agencies and international organizations (for example, Atkinson, Pearce, and Mourato 2006). The debate inside academic circles has often been acrimonious, but ultimately productive. The big issues concerning the reliability of contingent valuation raised by critics in the early 1990s have been resolved favorably with respect to the use of contingent valuation or have been shown to involve generic behavioral effects that also routinely characterize market data. A considerable body of evidence now supports the view that contingent valuation done appropriately can provide a reliable basis for gauging what the public is willing to trade off to obtain well-defined public goods. The time has come to move beyond endless debates that seek to discredit contingent valuation and to focus instead on making it better.

■ *I have received no compensation from any party for writing this article. However, over the last 30 years I have conducted contingent valuation studies for a number of local, state, and federal agencies as well as for foreign governments and international organizations. I have worked and continue to work for government agencies on natural resource damage assessments, including serving as principal investigator on the economic portion of the government's damage assessment for the Exxon Valdez oil spill. Helpful comments were received from the editors, David Autor, John List, and Timothy Taylor as well as from Michael Hanemann and V. Kerry Smith.*

References

Aldy, Joseph E., Mathew J. Kotchen, and Anthony A. Leiserowitz. 2012. "Willingness to Pay and Political Support for a US National Clean Energy Standard." *Nature: Climate Change* 2(8): 596–99.

Atkinson, Giles, David W. Pearce, and Susana Mourato. 2006. *Cost-Benefit Analysis and the Environment: Recent Developments.* Paris: OECD.

Arrow, Kenneth, Robert Solow, Paul Portney, Edward Leamer, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register,* January 15, 58(10): 4601–14.

Bateman, Ian, Alistar Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden. 1997. "Does Part–Whole Bias Exist? An Experimental Investigation." *Economic Journal* 107(441): 322–32.

Bowen, Howard R. 1943. "The Interpretation of Voting in the Allocation of Resources." *Quarterly Journal of Economics* 58(1): 27–48.

Carson, Richard T. 1997. "Contingent Valuation

and Tests of Insensitivity to Scope." In *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, edited by R. J. Kopp, W. Pommerhene, and N. Schwartz, 127–163. Boston: Kluwer.

Carson, Richard T. 2011. *Contingent Valuation: A Comprehensive Bibliography and History*, Northampton, MA: Edward Elgar.

Carson, Richard T., Nicholas E. Flores, and W. Michael Hanemann. 1998. "Sequencing and Valuing Public Goods." *Journal of Environmental Economics and Management* 36(3): 314–23.

Carson Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright. 1996. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics* 72(1): 80–99.

Carson Richard T., Nicholas E. Flores, and Norman F. Meade. 2001. "Contingent Valuation: Controversies and Evidence." *Environmental Resources Economics* 19(2): 173–210.

Carson, Richard T., Nicholas E. Flores, and Robert C. Mitchell. 1999. "The Theory and Measurement of Passive-Use Value." In *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries*, edited by I. J. Bateman and K. G. Willis, 97–130. Oxford: Oxford University Press.

Carson, Richard T., and Theodore Groves. 2007. "Incentive and Information Properties of Preference Questions." *Environmental and Resource Economics* 37(1): 181–210.

Carson Richard T., and W. Michael Hanemann. 2005. "Contingent Valuation." In: *Handbook of Environmental Economics*, Vol. 2, edited by K. G. Mäler and J. Vincent, 821–936. Amsterdam: North-Holland.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith. 1994. *Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight*. Report to the National Oceanic and Atmospheric Administration. http://www.econ.ucsd.edu/~rcarson/papers/SCalDDT.pdf.

Carson, Richard T., and Jordan J. Louviere. 2011. "A Common Nomenclature for Stated Preference Approaches." *Environmental and Resource Economics* 49(4): 539–59.

Carson Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 1992. "A Contingent Valuation Study of Lost Passive Use Values Resulting from the *Exxon Valdez* Oil Spill." Report to the Attorney General of the State of Alaska. http://www.econ.ucsd.edu/~rcarson/papers/AlaskaReport.pdf.

Carson Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 2003. "Contingent Valuation and Lost Passive Use: Damages from the *Exxon Valdez* Oil Spill." *Environmental and Resource Economics* 25(3): 257–86.

Carson, Richard T., Leanne Wilks, and David Imber. 1994. "Valuing the Preservation of Australia's Kakadu Conservation Zone." *Oxford Economic Papers*, October, 46 (Special Issue on Environmental Economics): 727–49.

Chapple, Clive. 2000. "The 1990 Oil Pollution Act: Consequences for the Environment." Paper presented at the Association of Environmental and Resource Economists Summer Workshop, La Jolla, CA.

Ciracy-Wantrup, S. V. 1947. "Capital Returns from Soil-Conservation Practices." *Journal of Farm Economics* 29(4): 1181–96.

Corso, Phaedra S., James K. Hammitt, and John D. Graham. 2001. "Valuing Mortality-Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation." *Journal of Risk and Uncertainty* 23(2): 165–84.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47(2): 315–72.

de Meijer, Claudine, Werner Brouwer, Marc Koopmanschap, Bernard van den Berg, and Job van Exel. 2010. "The Value of Informal Care–A Further Investigation of the Feasibility of Contingent Valuation in Informal Caregivers." *Health Economics* 19(7): 755–71.

Desvousges, William H., F. Reed Johnson, Richard W. Dunford, Kevin J. Boyle, Sara P. Hudson, and Nicole Wilson. 1993. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." In *Contingent Valuation: A Critical Assessment*, edited by J. A. Hausman, 91–159. Amsterdam: North Holland.

Flores, Nicholas E., and Richard T. Carson. 1997. "The Relationship between the Income Elasticities of Demand and Willingness to Pay." *Journal of Environmental Economics and Management* 33(3): 287–295.

Freeman, A. Myrick. 2003. *The Measurement of Environmental and Resource Values*, 2nd ed. Washington, DC: Resources for the Future.

Griffiths, Charles, Heather Klemick, Matt Massey, Chris Moore, Steve Newbold, David Simpson, Patrick Walsh, and William Wheeler. 2012. "U.S. Environmental Protection Agency Valuation of Surface Water Quality Improvements." *Review of Environmental Economics and Policy* 6(1): 130–46.

Hanemann, W. Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 81(3): 635–47.

Hanemann, W. Michael. 2008. *Embedding in Stated Preference*. Final Report to U.S. Environmental

Protection Agency, Agreement R-82966501, Department of Agricultural and Resource Economics, UC Berkeley.

Hausman, Jerry A., ed. 1993. *Contingent Valuation: A Critical Appraisal*. Amsterdam: North-Holland.

Hausman, Jerry A., Gregory K. Leonard, and Daniel McFadden. 1995. "A Utility-Consistent, Combined Discrete Choice and Count Data Model Assessing Recreational Use Losses Due to Natural Resource Damage." *Journal of Public Economics* 56(1): 1–30.

Heberlein, Thomas A., Mathew Wilson, Richard C. Bishop, and Nora C. Schaeffer. 2005. "Rethinking the Scope Test as a Criterion for Validity in Contingent Valuation." *Journal of Environmental Economics and Management* 50(1): 1–22.

Hicks, John R. 1943. "The Four Consumer Surpluses." *Review of Economic Studies* 11(1): 31–41.

Horowitz, John K., and Kenneth E. McConnell. 2002. "A Review of WTA/WTP Studies." *Journal of Environmental Economics and Management* 44(3): 426–47.

Jeuland, Marc, Marcelino Lucas, John Clemens, and Dale Whittington. 2009. "A Cost–Benefit Analysis of Cholera Vaccination Programs in Beira, Mozambique." *World Bank Economic Review* 23(2): 235–67.

Jiang, Yi, Leshan Jin, and Tun Lin. 2011. "Higher Water Tariffs for Less River Pollution—Evidence from the Min River and Fuzhou City in China." *China Economic Review* 22(2): 183–95.

Johnson, Bruce K., and John C. Whitehead. 2007. "Value of Public Goods from Sports Stadiums: The CVM Approach." *Contemporary Economic Policy* 18(1): 48–58.

Kahneman, Daniel, and Jack L. Knetsch. 1992. "Valuing Public Goods: Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22(1): 57–70.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2): 263–91.

Kochi, Ikuho, Bryan Hubbell, and Randall Kramer. 2006. "An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis." *Environmental and Resource Economics* 34(3): 385–406.

Krutilla, John V. 1967. "Conservation Reconsidered." *American Economic Review* 57(4):777–86.

Moore, William H. 1994. "The Grounding of the *Exxon Valdez*: An Examination of the Human and Organizational Factors." *Marine Technology Society Journal* 31(1): 41–51.

Murphy, James J., Geoffrey Allen, Thomas S. Stevens, and Darryl Weatherhead. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Resource and Environmental Economics* 30(3): 313–25.

Nagin, Daniel S., Alex R. Piquero, Elizabeth S. Scott, and Laurence Steinberg. 2006. "Public Preferences for Rehabilitation versus Incarceration of Juvenile Offenders: Evidence from a Contingent Valuation Study." *Criminology and Public Policy* 5(4): 627–51.

Poe, Gregory L., and Christian A. Vossler. 2011. "Consequentiality and Contingent Values: An Emerging Paradigm." In *The International Handbook on Non-Market Environmental Valuation*, edited by J. Bennett, 122–41. Cheltenham, UK: Edward Elgar Publishers.

Rollins, Kimberly, and Audrey Lyke. 1998. "The Case for Diminishing Marginal Existence Values." *Journal of Environmental Economics and Management* 36(3): 324–44.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *Review of Economics and Statistics* 36(4): 387–89.

Shrestha, Ram, Randall S. Rosenberger, and John B. Loomis. 2007. "Benefit-Transfer Using Meta-Analysis in Recreation Economic Valuation." In *Environmental Value Transfer: Issues and Methods*, edited by S. Navrud and R. Ready, 161–77. Dordrecht, Netherlands: Springer.

Small, Kenneth A., Clifford Winston, and Jia Yan. 2005. "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability." *Econometrica* 73(4): 1367–82.

Smith, V. Kerry, and Laura L. Osborne. 1996. "Do Contingent Valuation Estimates Pass a 'Scope' Test? A Meta-analysis." *Journal of Environmental Economics and Management* 31(3): 287–301.

Tversky, Amos, Paul Slovic, and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review* 80(1): 204–217.

U.S. Environmental Protection Agency. 1994. *President Clinton's Clean Water Initiative: Analysis of Benefits and Costs*. EPA 800-R-94-002 (NTIS Document No. PB94-154101). Washington DC: Office of Water, United States Environmental Protection Agency.

Whittington, Dale. 2002. "Improving the Performance of Contingent Valuation Studies in Developing Countries." *Environmental and Resource Economics* 22(1): 323–67.

Willig, Robert D. 1976. "Consumer's Surplus without Apology." *American Economic Review* 66(4): 589–97.