

Scope (in)sensitivity in elicited valuations

SHANE FREDERICK and BARUCH FISCHHOFF*

Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

The contingent valuation method (CV) is often used to value goods that are not traded directly in markets. Measures of value elicited by this method sometimes appear inadequately sensitive to the quantity, or 'scope', of the good being offered for hypothetical sale. Critics of CV have argued that such insensitivity casts doubts on the method. Advocates of CV have either contended that such results are consistent with economic theory, or have rejected the studies finding such results for failing to familiarize respondents with the goods being evaluated or for deviating from current methodological conventions. This article offers a critical interpretation of these accounts of scope insensitivity, and discusses the unresolved theoretical issues highlighted by scope tests and other tests of economic predictions.

1. Introduction

Policy decisions often affect the quantities of goods that people value. For example, damming a river may increase the quantity of electricity and water, but decrease the quantity of arable land; increase wind-surfing opportunities, but decrease rafting opportunities; increase the populations of some fish species, but decrease the populations of others. Making appropriate policy choices requires accurate assessments of the values of those changes. Price is routinely used to measure the value of goods that are exchanged in markets. However, some other procedure is needed to assess the values of goods like health, safety, and environmental quality, that are not exchanged in markets and have no associated market price.

Some resource economists have championed the Contingent Valuation (CV) method for estimating such 'non-market' values (e.g., Cummings *et al.*, 1986; Mitchell and Carson, 1989; Hahneman, 1994). CV attempts to measure the value of a specified quantity of goods by asking respondents to imagine how much they would be willing to pay for that amount *if* it were offered in a market (making their valuations 'contingent'). CV has received legal support in the United States from the Comprehensive Environmental Response Compensation and Liability Act (CERCLA), which granted it equal standing with revealed-preference techniques, which infer values from actual behaviour (Gregory *et al.*, 1993).¹

However, some critics view willingness-to-pay (WTP) responses to CV surveys as indicators of general attitudes, rather than expressions of economic value (Fischhoff

*We thank Dean Behrens, Colin Camerer, Robyn Dawes, Karen Jenni, Steven Klepper, Jack Knetsch, Greg Pogarsky, Thyra Uth and several anonymous reviewers for comments received on earlier drafts. This research was supported by the National Science Foundation, through the Center for the Integrated Assessment of Human Dimensions of Global Change at Carnegie Mellon University, and by a grant to Baruch Fischhoff from the Environmental Protection Agency.

¹See Daum (1993) for a dissenting view of the import of the CERCLA ruling.

et al., 1980; Kahneman, 1986; Fischhoff, 1991; Kahneman and Knetsch, 1992; Diamond and Hausman, 1993; Gregory *et al.*, 1993; Kahneman *et al.*, 1993; Kahneman and Ritov, 1994; Diamond and Hausman, 1994). For evidence, these critics often point to studies in which different groups of respondents express similar WTP despite wide variation in the number, amount, or extent of the good(s) they are valuing. For example, Tolley and Randall (1986) found about the same WTP for 10 days or 180 days of improved atmospheric visibility. Desvousges *et al.* (1993) and Schkade and Payne (1994) each found that respondents' WTP to protect migratory waterfowl was similar, whether the intervention was described as saving 2000 birds or 200 000 birds. This phenomenon is usually called the 'embedding effect' (Kahneman and Knetsch, 1992)² or 'insensitivity to scope' (Carson and Mitchell, 1993, 1995). It has been observed in elicitation formats ranging from brief telephone surveys (Kahneman and Knetsch, 1992) to intensive one-on-one interviews with detailed descriptions of the good (Schkade and Payne, 1994).

Many critics believe that scope insensitivity undermines the credibility of CV, because it shows that CV fails a fundamental test of measurement validity: yielding different values for differences in the measured object.³ Even the blue ribbon panel convened by the National Oceanic and Atmospheric Administration (NOAA), which tentatively endorsed CV, considered scope insensitivity to be 'perhaps the most important internal argument against the reliability⁴ of the CV approach.' (Arrow *et al.*, 1993, p. 4607). Despite the NOAA panel's lingering concerns, many CV practitioners have defended the method against demonstrations of scope insensitivity by alternately arguing: (1) that it is consistent with economic theory, (2) that the studies demonstrating it failed to familiarize respondents with the contingent good, or (3) that those studies failed to follow currently 'accepted' CV practice, such as in-person interviews, using a referendum format. This paper offers a critical analysis of these three accounts, followed by a general discussion of issues that complicate the design of scope tests, the creation of contingent markets, and the interpretation of contingent responses.

2. Could scope insensitivity be consistent with economic theory?

Economic theory typically assumes diminishing marginal rates of substitution between goods. With tradeoffs involving money, this assumption implies that the marginal value of an additional unit of good should decrease as the *ratio* of money to good decreases. Because every unit of good purchased both reduces one's wealth

²Hahneman (1994) notes that the term 'embedding' has also been used to refer to two related, but distinguishable phenomena: (a) when WTP for a composite good is less than the sum of the WTPs for its separate components and (b) when WTP for a good altered by its serial position, when multiple goods are valued sequentially.

³These critics also point to other cases where CV fails the complementary measurement test – producing different values across fundamentally similar situations. For example, when multiple goods are valued sequentially, the order of presentation may influence the total value assigned to the collection of goods (e.g., Samples and Hollyer, 1990).

⁴Here, as elsewhere in the CV literature, the term *reliability* was used to connote what psychologists would call *validity*. That is, it was used to imply that the measurement (a respondent's reported WTP) corresponded with the measured construct (the 'true' value as defined by some other criterion) and not merely that the measurements were similar across time, or interviewers, or interview locations.

and increases one's amount of the good, the monetary value of successive units of a good should decline. If this effect is sufficiently large, WTP for a small number of units of a good might equal (or approximate) WTP for a much larger number of units. In that case, scope insensitivity would be consistent with economic theory, as several CV practitioners have pointed out:

'... (embedding) is exactly what would be expected from simple consumer theory. The *absence* of this embedding phenomenon, not its presence, would be an indictment of CV. As one adds more substitute goods to the choice set of a budget constrained individual, or allows the individual the ability to adjust more goods in the choice set, the value of the individual components will decrease. This result is not unique to CV or the valuation of public goods, but would be expected in a market context with private goods as well.'

(Kopp, 1992, p. 128)

'(Kahneman and Knetsch's demonstration of embedding) fails to provide the test (of economic theory) that they imply because their primary results *would be* consistent with standard value theory.'

(Smith, 1992, p. 96)

'... embedding effects are standard economic phenomena induced by substitution relationships... and evidence of its effects in CV data sets does not, *per se*, invalidate the particular studies or the method in general.'

(Randall and Hoehn, 1996, p. 378)

Although it is theoretically *possible* that (rapidly) diminishing marginal rates of substitution could account for the scope insensitivity observed in CV studies, such an interpretation is inconsistent with the evidence we review next.

3. Evidence against economic accounts of scope insensitivity

3.1. *Insensitivity for willingness to accept (WTA) judgements*

The assumption of diminishing marginal rates of substitution means that successive units of good *obtained* should be worth less and less. It also means that successive units of good *given up* should be worth more and more. Thus, WTA should increase *more than* linearly with quantity, so that respondents should require *more than* ten times as much money to give up ten units of good than to give up one unit. However, when WTA judgements have been elicited for different amounts of a good (Baron and Greene, 1996; Dubourg *et al.*, 1994), WTA not only increased less than linearly with quantity, but was highly insensitive to it.⁵ That is, respondents stated the same or very similar WTA for widely differing amounts of good being given up.⁶

⁵The magnitudes varied in the study by Dubourg *et al.* (1994) were different *probabilities* of injury or death, or different *lengths of time* that some risk must be incurred, and were not different physical *quantities* of good in the sense used elsewhere in this paper. However, these valuations should still rise linearly with 'quantity' rather than less than linearly, as was observed.

⁶See Gregory (1986) and Gregory and McDaniels (1987) for discussions of WTP and WTA as measures of value, and why CV practitioners predominantly favour WTP (even in situations where WTA is theoretically appropriate). See Hicks (1943) and Willig (1976) for discussions of differences between the two, including the theoretical rationale of why they differ, but to only a small degree for transactions involving small changes in wealth.

3.2. Sub-proportionality for quantity-demanded judgements

Baron and Greene (1996) turned the typical CV task around, by asking respondents how many units of a good they would demand for a specified payment of money. Economic theory predicts that the quantity of good demanded should increase *more than* proportionally to the amount of money specified. (A person who is willing to pay \$5 for one unit should be willing to pay somewhat less than \$10 for 2 units. Thus, to give up \$10, this person should demand *more* than 2 units). Baron and Green found, instead, that quantity demands were not only subproportional, but substantially insensitive to the amount of money specified, which cannot be explained by diminishing marginal rates of substitution.

3.3. Violations of 'adding-up' tests

The monetary value of a given quantity of a good should not depend on whether it is bought all at once or in successive increments (assuming that the purchases occur at the same time and the transaction costs are equal). For example, it should not matter whether one first buys one unit, and then buys another, or whether one buys two units to begin with. This is formalized in Equation 1, where the two sides of the equality are alternative expressions for the value of acquiring $A + B$ additional units, for someone who currently has i units. On the left, the evaluation is done in two steps; on the right, it is done in one. Thus, if individuals were randomly assigned to groups evaluating each of these three changes, the sum of the WTPs for the two groups on the left should equal the WTP of the group on the right (neglecting wealth effects).⁷

$$\text{WTP}(X_{i \rightarrow i+A}) + \text{WTP}(X_{i+A \rightarrow i+A+B}) = \text{WTP}(X_{i \rightarrow i+A+B}) \quad (1)$$

Diamond *et al.* (1993) performed such an 'adding-up' test. In that study, one group of respondents was asked to value the preservation of the Selway-Bitterroot wilderness area, assuming that the Washakie wilderness area would be developed ($\text{WTP}(X_{i \rightarrow i+A})$).⁸ A second group was asked to value the preservation of the Washakie area, assuming that Selway-Bitterroot would be preserved (WTP

⁷As a reviewer pointed out, Equality (1) does not strictly hold in separate valuation by independent groups, because that is not exactly analogous to sequential purchase by an individual. In a true sequential valuation, WTP to acquire A [$\text{WTP}(X_{i \rightarrow i+A})$] would actually be subtracted from wealth. Thus, the valuation of B [$\text{WTP}(X_{i+A \rightarrow i+A+B})$] would be made from a slightly reduced wealth level. However, when the three purchases are made by three separate groups, the group which values the change $X_{i+A \rightarrow i+A+B}$ has not actually purchased A , but has been given A , resulting in a slightly higher WTP for B than if they had actually expended money to purchase A . Thus, for "simultaneous" valuations among different groups (as in the Diamond *et al.* study), economic theory would predict $\text{WTP}(X_{i \rightarrow i+A}) + \text{WTP}(X_{i+A \rightarrow i+A+B})$ to slightly exceed $\text{WTP}(X_{i \rightarrow i+A+B})$, because "endowing" the one group with A is equivalent to raising their wealth by the value of A , which should raise their WTP slightly. However, with the WTPs in the Diamond *et al.* study (between \$47 and \$66), an income of \$20000, and an income elasticity of 1, wealth effects would predict that the left side of equality one to exceed the right by about one cent – nowhere near the \$31.00 difference that was actually observed.

⁸In theory, WTA judgements should have been elicited in this study, because subjects were asked to express values for *preventing reductions* in the current level of wilderness areas.

($X_{i+A \rightarrow i+A+B}$). A third group was asked to value the preservation of both areas, assuming that neither would otherwise be preserved ($WTP(X_{i \rightarrow i+A+B})$).⁹ The responses violated economic predictions. The WTP for all three groups was similar, such that the total value of the two areas was roughly twice as high when valued 'sequentially' (the left side of Equality 1), than when valued simultaneously (the right side of Equality 1).

The 'adding-up' test should hold for *any* monotonic value function. The further assumption of diminishing marginal rates of substitution means that a given quantity of good should be worth less the more of the good one already has. Thus, $WTP(X_{i \rightarrow i+A}) \geq WTP(X_{i+B \rightarrow i+B+A})$ and $WTP(X_{i \rightarrow i+B}) \geq WTP(X_{i+A \rightarrow i+B+A})$, which, with the assumption of negligible wealth effects,¹⁰ implies Inequality 2:¹¹

$$WTP(X_{i \rightarrow i+A+B}) \geq WTP(X_{i+A \rightarrow i+A+B}) + WTP(X_{i+B \rightarrow i+B+A}). \quad (2)$$

Samples and Hollyer (1990) tested Inequality 2 by asking three groups of subjects for their independent or joint valuation of preserving 600 humpback whales and 1200 Hawaiian monk seals – two endangered marine mammals which inhabit the waters around Hawaii.¹² Again, the responses violated economic predictions. The sum of the mean WTPs to preserve seals separately and whales separately (for subjects who were not told that the other species was threatened) substantially *exceeded* the mean WTP to preserve both mammals (for subjects told that both species were threatened). We replicated their results, using scenarios that restored timber wolves to Maine, to Wisconsin, or to both states (Frederick and Fischhoff, 1997a). The median WTP was \$20 in all three conditions, violating Inequality 2.

⁹An 'adding-up test' is more complicated than a 'scope test,' because it requires the manipulation of *both* the *initial* and *final* levels of the good. In this study, the initial level was manipulated between Groups 1 and 2 (Group 1 was asked to assume that areas *A* and *B* were both threatened, whereas Group 2 was told that only area *B* was threatened), whereas the final level was manipulated between Groups 1 and 3 (Group 1 valued only area *A*, whereas Group 3 valued areas *A* and *B*).

¹⁰Inequality 2 would not strictly hold in the presence of wealth effects. To use an extreme but illustrative example from Kahneman and Knetsch (1992), suppose that state *i* is the condition of having only one leg and one arm. In that case, a person's WTP to obtain *either* an arm or a leg would likely exhaust their wealth, and therefore approximate their WTP to obtain *both* an arm and a leg. Thus, the sum of the two WTP's [$WTP(X_{i+LEG \rightarrow i+LEG+ARM}) + WTP(X_{i+ARM \rightarrow i+ARM+LEG})$] would be *greater than* the WTP of an individual purchasing both an arm and a leg [$WTP(X_{i \rightarrow i+ARM+LEG})$], which would violate Equality 2. However, for the levels of WTP elicited in CV surveys (typically less than \$100), such large wealth effects would not arise, and Inequality 2 should be expected to hold.

¹¹To see that Inequality 2 follows directly from the assumption of diminishing marginal rates of substitution, note that $WTP(X_{i \rightarrow i+A+B})$ can be decomposed as either $WTP(X_{i \rightarrow i+A}) + WTP(X_{i+A \rightarrow i+A+B})$, or $WTP(X_{i \rightarrow i+B}) + WTP(X_{i+B \rightarrow i+B+A})$. Substituting the first of these decompositions into the left side of the Inequality 2 produces the following expression:

$$WTP(X_{i \rightarrow i+A}) + WTP(X_{i+A \rightarrow i+A+B}) \geq WTP(X_{i+A \rightarrow i+A+B}) + WTP(X_{i+B \rightarrow i+B+A})$$

When the common terms are cancelled, the remaining expression merely states the assumption of diminishing marginal rates of substitution; that the marginal monetary value of a given number of units of a good declines as the levels of the good already possessed increases: $WTP(X_{i \rightarrow i+A}) \geq WTP(X_{i+B \rightarrow i+B+A})$.

3.4. Scope insensitivity for unrelated goods

In interviews with Vancouver residents, Knetsch (1993) found that the median WTP to protect fish populations (\$52), or to maintain historic buildings (\$56), was similar to the median WTP for *both* projects (\$51). In another study, he found that the proportion of people willing to pay \$50 for both grocery coupons *and* a weekend at an Ontario resort (59%) was no higher than the proportion willing to pay for either the coupons (60%) or the weekend (59%) separately. It seems unlikely that the apparent scope insensitivity here could be accounted for by rapidly diminishing rates of substitution, because the component goods of the scope test (arguably) satisfy no similar preferences, serve no common purposes, and share no common functions.¹³

3.5 Summary

Contingent responses that show scope insensitivity for WTA judgements, sub-proportionality for quantity-demanded judgements, and violations of 'adding-up' tests cannot be considered expressions of economic value, because they violate properties that the theory assumes. CV proponents have typically dismissed these findings by claiming that the contingent goods in those studies were not sufficiently 'familiar' or that 'accepted' methodological conventions were not followed, such as in-person interviews and a referendum-style elicitation procedure. We consider these explanations next.

4. Is scope insensitivity due to a lack of familiarity with the contingent good?

CV practitioners have often emphasized the importance of establishing 'familiarity' with the contingent good. For example, the concept of familiarity appears in two of the four 'reference operating conditions' that Cummings *et al.* (1986, p. 104) identified as essential for CV's reliability:¹⁴

1. Subjects must understand, be familiar with, the commodity to be valued.
2. Subjects must have had (or be allowed to obtain) prior valuation and choice experience with respect to consumption levels of the commodity.

¹²In Inequality 2, *A* and *B* represent the whales and seals, *X* denotes a superordinate good that includes humpback whales and monk seals as elements, and *i* is the initial level of this good, excluding the humpback whales and Hawaiian monk seals that would be preserved with the contingent payment. The notation is inappropriate if one considers whales and seals as two separate goods rather than as elements of some larger good, such as *endangered marine mammals*. For example, if *A* were an apple and *B* were a battery, it would not be meaningful to denote them as elements of some encompassing good *X*. Nevertheless, Inequality 2 should hold, because one would still expect that $WTP(A \& B) = WTP(A) + WTP(B)$, neglecting wealth effects.

¹³Some CV enthusiasts have interpreted such results as revealing the substitutability of the component goods (i.e., historic buildings and fish) rather than as indictment of the method used to assess their value (e.g., Cummings *et al.*, 1994, p. 212).

¹⁴Some have criticized these conditions for restricting the applicability of CV to goods that are already marketed.

Familiarity has been stressed elsewhere as well:

'The most striking implication from our study is the extremely difficult task of valuing marginal changes in a natural resource ... Part of this difficulty may arise from ... the respondents' general lack of familiarity with the commodity being evaluated.'

(Boyle *et al.*, 1994, p. 80)

'The bane of CVM researchers is the fact that CVM subjects are often uninformed about environmental goods to be valued in CVM studies.'

(Carson and Mitchell, 1995, p. 213)

'There appears to be general agreement among CVM practitioners that (one of the) major problems encountered in the design of 'good' CVM questionnaires (is) assuring that subjects understand the substance of the good that they are to value (the 'familiarity issue').

(Neill *et al.*, 1994, p. 146)

Mitchell and Carson (1989, p. 296) explicitly state that CV can measure the values of familiar goods. Unfortunately, there are no formal definitions of 'familiarity' that permit an empirical test of this claim. However, in a recent series of experiments, we elicited valuations for a set of goods that should satisfy any definition of familiarity: supermarket products, such as canned tuna, applesauce, and toilet paper (Frederick and Fischhoff, 1997b). In one experiment, respondents in a *between-subjects* condition valued *either* the small or the large quantity of each good, whereas respondents in a *within-subjects* condition valued *both* the small and the large quantities of each good. The larger quantities typically received higher valuations under both conditions, indicating some scope sensitivity for these goods. However, the particular values for each quantity were often dramatically affected by the elicitation procedure. The within-subject design generated much more quantity sensitivity and, therefore, implied much greater marginal value for the additional quantity. (On average, $WTP(\text{large}) - WTP(\text{small})$ was about 2.5 times greater using the within-subject design).¹⁵

These results show (1) that the choice of elicitation procedure can influence contingent responses even for 'familiar' goods, and (2) that at least one of these two procedures is invalid. Unfortunately, the data do not indicate which procedure to reject. The within-subject design is more valid if respondents' 'true' value functions are *less* concave than their collective responses to between-subject quantity manipulations imply; it is less valid if it induces *artificial* sensitivity (if respondents' 'true' value functions were actually *more* concave than their responses implied).

We could not find any other studies in the peer-reviewed literature that compared

¹⁵Of course, a sceptic of this study might argue even these respondents were insufficiently familiarized with the described goods to understand exactly what they were buying. For example, perhaps '24 ounces' and '48 ounces' evoke similar mental images of an applesauce jar. When both numbers are presented together, the bigger one would obviously indicate a larger quantity, but in isolation the numbers may have little absolute meaning, causing them to be treated similarly. However, the 'between-within discrepancy' occurred even for toilet paper, which was described in terms of the *number* of rolls – a quantity description which should be easily appreciated in an absolute sense. Thus, even respondents who understand how much a quantity *is* may not know how much that amount is *worth*, and just respond with a number that seems like the right amount to spend on that *type* of good.

the values elicited from between-subject and within-subject quantity manipulations, or that discussed the relative validity of the two procedures. The NOAA panel broached the issue, but concluded that within-subject designs would not constitute a fair test of scope sensitivity:

‘We must reject one possible approach (for testing scope sensitivity), that of asking each respondent to express willingness to pay to avert incidents of varying sizes; the danger is that embedding will be forcibly avoided, *still without realism*’.

(Arrow *et al.*, 1993, p. 6048, emphasis added)

We share NOAA’s scepticism of within-subject scope tests, because we suspect that valuations of any particular quantity would be sensitive to its relative position within the range selected for valuation but insensitive to *which* range is chosen, resulting in insensitive (or incoherent) values across studies using different quantity ranges. For example, we predict that the value of cleaning up ten miles of a polluted river would be higher in a study where respondents valued one mile and ten miles of cleanup than in a study where they valued ten miles and 100 miles. However, we do not believe that the demands of within-subject designs necessarily decrease the validity of the contingent responses – indeed, respondents in a within-subject quantity manipulation who report that a lot more of a good is worth a lot more to them may be revealing more about their true values than respondents in a between-subject design, who (collectively) indicate that a lot more of a good is only worth only a little more.

5. Is scope insensitivity caused by a failure to use ‘state of the art’ CV methodology?

As noted earlier, the studies cited here as unfavourable to CV typically did not follow the NOAA panel’s methodological standards. Consequently, some CV advocates have dismissed those studies as unrepresentative of ‘good’ CV surveys (e.g., Carson, 1997). Furthermore, advocates have often claimed evidence of ‘scope sensitivity’¹⁶ by citing CV studies showing statistically significant differences between groups valuing different quantities of a good (e.g., McDaniels, 1988; Loomis *et al.*, 1993).¹⁷ The next section describes one such demonstration of ‘scope sensitivity’ in the Long Beach Harbor Study commissioned by NOAA – one of the most expensive and intensively prepared CV surveys to date.

5.1. A demonstration of ‘scope sensitivity’: The Long Beach Harbor Study

From May 1991 to March 1994, NOAA retained Natural Resource Damage Assessment, Inc., to estimate the lost use value resulting from some of the effects of DDT and PCB residues in ocean sediments in Long Beach Harbor (Carson *et al.*, 1994). The study implemented 20 of the 23 guidelines recommended by the NOAA panel, including extensive pretesting, in-person interviews, and the use of a

¹⁶At a recent conference, Richard Carson claimed that a meta analysis of split sample scope tests shows ‘large, consistent’ effects. We assess the size and consistency of these effects in a separate paper (Fischhoff and Frederick, 1998).

¹⁷Some of these studies followed (many of) NOAA’s methodological guidelines, but others did not (Fischhoff and Frederick, 1997).

'referendum' format, in which respondents were asked whether they would endorse a restoration programme at a specified cost (Would you pay \$Y for [SPECIFIED PROGRAMME]? YES or NO).¹⁸

The 'scope test' varied the extent of injury that would be mitigated by the proposed clean-up programme, using a between-subject design. The 'small injury' scenario promised to accelerate the recovery of two fish species (kelp bass and white croaker) by ten years (from 15 years to 5 years). The 'large injury' scenario promised to accelerate the recovery of these two fish species *and* two bird species (bald eagles and peregrine falcons) by 45 years (from 50 years to 5 years).

The mean WTPs elicited for the 'large-injury' and the 'small-injury' programmes were \$63 and \$34, respectively – a statistically significant difference in the direction predicted by economic theory. However, because economic theory predicts only the *sign*, and not the *size* of the differences expected, readers must judge for themselves whether a WTP ratio of 1.9 ($\$63 \div \34) adequately reflects the relative value of the two programmes. We doubt that anyone who evaluated the programmes side by side would think so. The *small injury* scenario involved two obscure, and unthreatened fish species 'living in the millions up and down the California coast,' whose 'reproduction problemmes' were restricted to the area directly above the polluted sediments near Los Angeles Harbor, and which would recover by themselves in 15 years. The *large injury* scenario involved those fish, *and* two locally extinct, and nationally endangered bird species whose 'reproduction problems' extended over the whole South Coast (a vastly greater area than the area directly above the polluted sediments), and whose recovery would otherwise take 50 years.¹⁹ (Appendix 1 presents verbatim excerpts from the CV protocol).

6. What are the boundaries of the contingent good?

The difficulty of quantifying the magnitude of the scope manipulation in the Long Beach Harbor study reflects an unresolved theoretical issue in CV studies. A meaningful evaluation task must specify the *initial* and *target* levels of the good associated with the proposed transaction (Fischhoff and Furby, 1988). Yet a change can often be defined or interpreted in many different contexts, each having a different initial and target level. For example, subjects in the Diamond *et al.* (1993) study were asked to value the preservation of one or two western wilderness areas, assuming that there were 57 'other such wilderness areas' in the western US (of which eight or nine were described as threatened with 'development'). However, it is unclear why respondents should consider only designated wilderness areas in the

¹⁸ Y was varied across respondents, with initial values of \$10, \$25, \$80, \$140, and \$215. Depending on the respondents' vote, a second vote was proposed at a different price (higher, if the respondent voted YES initially; lower, if they voted NO). This was done to obtain a better estimate of respondents' reservation value (because an open ended question eliciting this information directly is forbidden by NOAA's guidelines).

¹⁹In another 'demonstration' of scope sensitivity, McDaniels (1988) found that the average WTP to prevent one auto accident death was \$14, and that the average of WTP (of a second group) to prevent 10000 deaths was \$182. Thus, a 10000 fold difference in the magnitude of the good elicited a 13 fold difference in valuations.

West, rather than all other designated wilderness areas in the US, or in the world, or all the wilderness that is not officially designated. Indeed, because the contingent 'development' in that study was a 1% annual harvest of mature timber, respondents might want to think in terms of changes in the total acreage of large trees left standing after different levels of 'development,' rather than changes in the number of areas labelled 'preserved' or 'developed' in some arbitrarily specified region.²⁰

Similar issues arise in other studies. In Samples and Hollyer's (1990) 'Contingent Valuation of Wildlife Resources in the Presence of Substitutes and Complements,' it is unclear what substitutes (or complements) respondents are meant to consider when determining the value of 600 humpback whales or 1200 monk seals – other *endangered animals?*, *endangered marine mammals near Hawaii?*, *marine mammals near Hawaii?*, *endangered mammals?*, *mammals?*, *Hawaiian animals?*, *aquatic organisms?*... and so on.²¹ Desvougues *et al.* (1993) attempted to impart a meaningful sense of the magnitude of the effect of the programme described in their study by describing the death of 200 000 waterfowl as '*less than 2% of the total population of migratory waterfowl in the Mississippi Flyway.*' However, that description assumes that respondents' values are defined over that category, rather than a narrower one (*the number of waterfowl in a particular state*), or a broader one (*the number of birds in the US*), or a different one (*the number of animals that die from human hazards*). Moreover, to decide what is 'meaningful' one must ask '*relative to what?*' – 200 000 waterfowl may represent less than 2% of the migratory waterfowl in the Mississippi Flyway, or an infinitesimal fraction of all the animals in the world or almost all of the waterfowl in the continental US who drown from having their feathers coated by oil.

These examples illustrate the (as yet, unmet) challenge of determining the appropriate specification for contingent scenarios. Economic theory can say nothing about whether two things are substitutes without observing demand (which, of course, cannot be done for the goods where CV is needed).²² Thus, it provides

²⁰Milgrom (1993) makes a similar point: 'in order to assess the rate of substitution between an environmental amenity and some private goods, a consumer would need to know the quantities of the various environmental amenities that already exist. If environmental amenities are like ordinary consumer goods, then a consumer's willingness to pay for additional units must depend on the quantities that are already available... It is doubtful that many consumers even have a rough idea of how many acres of national forest or wildlife refuge exist or how many more acres are likely to be set aside or restored in the future. Yet, in principle, this information is essential (for determining the value of preserving a specific wilderness area).' (Milgram, 1993, p. 425)

²¹Our replication of their study (Frederick and Fischhoff, 1997a) is subject to the same problem. Reestablishing 100 timber wolves to Maine could be described or interpreted as 100 more timber wolves in *Maine*, 100 more timber wolves in the *United States*, 100 more timber wolves in the *world* ... as 100 more *timber wolves*, 100 more *wolves* (a category which includes coyotes, red wolves, etc.), 100 more *large mammals*, 100 more *animals*, one more *environmental project*, one more *good cause* ... and so on. Each of these possible categories would have different initial and target levels, even though the effect of the project is the same in each case (100 more timber wolves in Maine).

²²This problem is not limited to the exotic goods valued in CV studies, but is true of goods generally. Stigler (1966, p. 24) remarks on the difficulty of measuring the goodness of substitution between non-identical goods: 'There is no simple 'technological' measure of substitution: not only is it difficult to compare heterogeneous things (is radio a better substitute for television than for a theater or a newspaper?) but substitutability varies with circumstances (a tractor is a substitute for a horse to a farmer, less so for a riding academy).'

practitioners' no guidance for constructing the survey, and no justification for their decisions to include some information in a scenario (e.g., telling respondents how many wolves currently live in the US), while excluding other potentially relevant information (e.g., how many coyotes currently live in Maine).²³

7. What are the boundaries of the contingent market?

If a contingent good is conceived broadly (e.g., if timber wolves in Maine are considered part of the world stock of *large predatory mammals*), any proposed project (e.g., adding 100 timber wolves to Maine) would represent such a negligible proportional change in the level of that good that it would have little value. On the other hand, conceiving such goods narrowly would create such an enormous universe of different goods that an individual asked to consider them all would have to either (a) assign a negligible value to each or (b) be forced to adopt a deliberately narrow focus, where they decide which problems to neglect entirely (i.e., which desirable environmental and social projects to withhold payment from; which threatened animal species to deny protection or support).

In practice, most respondents express a substantial WTP for nearly any good (or cause) that researchers set before them. Studies have found, for example, that people are willing to pay \$95 for improved visibility in the Grand Canyon (Schulze *et al.*, 1981), \$68 to clean up the nation's rivers (Mitchell and Carson, 1986), \$50 to prevent logging in the Selway-Bitterroot wilderness area (Diamond *et al.*, 1993), \$80 to protect migratory waterfowl from waste water holding ponds (Desvougues *et al.*, 1993), \$85 to reduce the chance of oil spills off the coast of Alaska (Kemp and Maxwell, 1993), \$86 to protect spotted owls (Hagen *et al.*, 1992), \$62 to preserve whooping cranes (Bowker and Stoll, 1988), and \$33 to manage the nesting habitat of loggerhead turtles (Whitehead, 1992).

The collective infeasibility of such payments has long been recognized (e.g., Hoehn and Randall, 1989). However, current CV theory offers no solution except for recommendations to survey designers to place a greater emphasis on budget constraints, substitute goods, or alternative uses of the contingent payment. For example, the NOAA panel (1, p. 4609) states: 'Respondents must be reminded that their WTP for the environmental program in question would reduce their expenditures for private goods and other public goods. This reminder should be *more than perfunctory, but less than overwhelming.*' (emphasis added).²⁴

²³For any nonlinear value function defined over some good X , determining the value of additional k units, $WTP(X_{i \rightarrow i+k})$, depends on the value of i , which depends on how broadly one defines the good. For example, if one considers Maine timber wolves as *wild animals*, then mice, salmon, kangeroos, and cockroaches could all be seen as substitutes. A somewhat narrower conception of timber wolves (say, as *large, predatory mammals*) would still not rule out Siberian timber wolves, Alaskan grizzly bears, Kenyan chimpanzees, and Pacific sperm whales. Moreover, the population of those substitute goods may be theoretically relevant for respondents contemplating the economic value of a project that would increase the number of timber wolves in Maine by 100.

²⁴Presumably, telling respondents 'Remember, this is real money.' would be considered *perfunctory*, whereas forcing them to prorate their income among an exhaustive list of market and non-market goods would be considered *overwhelming*.

8. Conclusions

Debates over the validity of CV have often centred around the issue of scope sensitivity. Advocates of the method have focused on between-subject or split-sample studies that satisfy the (weak) economic prediction that the relation between quantity and WTP is nondecreasing. For example, Carson *et al.* (1994) showed that, on average, respondents placed a higher value on a program helping two common fish *and* two endangered birds than on a program helping just the two common fish. Critics point to other studies where responses violate economic predictions. The studies demonstrating such violations have typically deviated from the methodological conventions advocated by some CV proponents (as do many of the studies proponents cite as demonstrations of monotonicity; see Fischhoff and Frederick, 1997). We know of no studies that satisfy both the stricter methodological standards advocates endorse and the stricter validity standards critics demand.

Both advocates and critics agree that if respondents are asked to take the market metaphor seriously, they need complete specifications of the substantive features of the hypothetical market. However, few CV studies specify these features in the detail that we argue is necessary. Thus, respondents may have no idea what type of contingent market to imagine; whether the focal good is the only additional good in the hypothetically expanded market, or whether other such goods will also be included (and, if so, what the costs of those alternative goods would be).²⁵

We suspect that this is not for lack of effort (the Long Beach Harbor study involved extensive pretesting by a professional survey firm, and several hundred pages are devoted to describe the stimuli development). More likely, it is for the want of a substantive theory defining the relevant features for a specific transaction. Creating that theory cannot occur in the current fragmented institutional environment which promotes narrowly focused studies of a few arbitrarily chosen goods and which discourages scholars with different expertise from working collaboratively on these practically important, and intellectually challenging questions.

Appendix: Summarized description of contingent scenarios used in scope test

Small injury mitigation

The white croaker and kelp bass are having problems producing young in this one place off the south California coast (a small area between Santa Monica Bay and

²⁵Although the economic value of a good is not necessarily related to its *own* market price, it is intimately tied to the market price of close substitutes. Indeed, in a two-good economy of perfect substitutes (say, blue or black ballpoint pens), the relation is a strict equivalence – the economic value of good *A* is the market price of good *B*, because the consumer would always buy *A* at a price up to P_B but would never pay more. Thus, if the price of substitute goods is not provided in a contingent market, the meaning of the contingent response becomes ambiguous. Suppose, for example, that a respondent: (1) would prefer to spend \$20 to preserve *whooping* cranes to *every* alternate bundle of market goods that could be purchased for \$20; (2) considers *sandhill* cranes to be a perfect substitute for *whooping* cranes and would, therefore, prefer preserving *sandhill* cranes at any price lower than \$20; and (3) could preserve an equal number of *sandhill* cranes for \$10, though this substitute good is not explicitly mentioned in the survey (and therefore not included in the contingent market that the survey defines). Is the correct value of *whooping* cranes \$10 or \$20?

San Pedro Bay). Millions of these two fish live in other places along the California coast, neither species is in any danger of becoming extinct. These two species will not fully recover for another 15 years, unless the clean up program is implemented, in which case they will fully recover in 5 years.

Large injury mitigation

The white croaker and kelp bass are having problems producing young in this one place off the south California coast (a small area between Santa Monica Bay and San Pedro Bay). Millions of these two fish live in other places along the California coast, neither species is in any danger of becoming extinct. In addition to these two fish species, two of the many species of birds living along the South Coast also have reproduction problems – the peregrine falcon and the bald eagle. Unlike the white croaker and kelp bass, which only have problems in this place (a small area between Santa Monica Bay and San Pedro Bay) these birds are having reproduction problems everywhere they live along the South Coast, including the Channel Islands. No eagles have hatched young on their own in this area and only rarely have some peregrine falcons been able to do so. As a result, there are no local populations of bald eagles and peregrine falcons along the South Coast. The eagles and falcons are listed as endangered by the State of California. These birds are also listed as endangered in most of the other states where they live. These two fish species and two bird species will not fully recover for another 50 years, unless the clean up programme is implemented, in which case they will fully recover in 5 years.

References

- Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R. and Schuman, H. (1993) Report of the NOAA Panel on Contingent Valuation, 58 *Federal Register* 46.
- Baron, J. and Greene, J. (1996) Determinants of insensitivity to quantity in valuation of public goods: Contribution, warm glow, budget constraints, availability, and prominence, *Journal of Experimental Psychology: Applied* **2**, 107–25.
- Bowker, J. and Stoll, J. (1988) Use of dichotomous choice nonmarket methods to value the whooping crane resource, *American Journal of Agricultural Economics* **70**, 372–81.
- Boyle, K., Desvousges, W., Johnson, F., Dunford, R. and Hudson, S. (1994) An investigation of part-whole biases in contingent-valuation studies, *Journal of Environmental Economics and Management* **27**, 64–83.
- Carson, R. (1997) Contingent Valuation Surveys and Tests of Scope Sensitivity, in: W. Pommerehne, N. Schwarz and R. Kopp (Eds) *Determining the Value of Nonmarketed Goods*, New York: Plenum.
- Carson, R. and Mitchell, R. (1993) The issue of scope in contingent valuation studies, *American Journal of Agricultural Economics* **75**, 1263–67.
- Carson, R. and Mitchell, R. (1995) Sequencing and nesting in contingent valuation surveys, *Journal of Environmental Economics and Management* **28**, 155–73.
- Carson, R., Hanemann, W., Kopp, R., Krosnick, J., Mitchell, R., Presser, S., Ruud, P. and Smith, K. (1994) Prospective interim lost use value due to DDT and PCB contamination in the Southern California Bight. NOAA Contract #50-DGNC-1-00007.
- Cummings, R., Brookshire, D. and Schulze, W. (Eds) (1986) *Valuing Environmental Goods: An assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman and Allenheld.
- Cummings, R., Ganderton, P. and McGuckin, T. (1994) Substitution effects in CVM values, *American Journal of Agricultural Economics* **76**, 205–14.

- Daum, J. (1993) Some Legal and Regulatory Aspects of Contingent Valuation, in: J.A. Hausman (Ed.) *Contingent Valuation: A Critical Assessment*, Amsterdam: North Holland.
- Desvousges, W.H., Johnson, F., Dunford, R., Hudson, S., Wilson, K. and Boyle, K. (1993) Measuring Resource Damages with Contingent Valuation: Tests of Validity and Reliability, in: J.A. Hausman (Ed.) *Contingent Valuation: A Critical Assessment*, Amsterdam: North Holland.
- Diamond, P.A. and Hausman, J. (1993) On Contingent Valuation Measurement and Nonuse Values, In: J.A. Hausman (Ed.) *Contingent Valuation: A Critical Assessment*, Amsterdam: North Holland.
- Diamond, P.A. and Hausman, J. (1994) Contingent valuation: is some number better than no number? *Journal of Economic Perspectives* **8**, 45–64.
- Diamond, P.A., Hausman, J., Leonard, G. and Denning, M. (1993) Does Contingent Valuation Measure Preferences? Experimental Evidence, In: J.A. Hausman (Ed.) *Contingent Valuation: A Critical Assessment*, Amsterdam: North Holland.
- Dubourg, W., Jones-Lee, M. and Loomes, G. (1994) Imprecise preferences and the WTP-WTA disparity, *Journal of Risk and Uncertainty* **9**, 115–33.
- Fischhoff, B. (1988) Specifying Value Measurements, in: B. Driver, G. Peterson, and R. Gregory (Eds) *Evaluating Amenity Resources*, New York: Venture.
- Fischhoff, B. (1991) Value elicitation: is there anything in there? *American Psychologist* **46**, 835–47.
- Fischhoff, B. and Frederick, S. (1997) Large, consistent effects: a study in linguistic norms. Working Paper, Carnegie Mellon University.
- Fischhoff, B. and Furby, L. (1988) Measuring values: a conceptual framework for interpreting transactions with special reference to contingent valuation of visibility, *Journal of Risk and Uncertainty* **1**, 147–84.
- Fischhoff, B., Slovic, P. and Lichtenstein, S. (1980) Knowing What You Want. Measuring Labile Values, in: T. Wallsten (Eds) *Cognitive Processes in Choice and Decision Behavior*, Hillsdale, NJ: Erlbaum.
- Frederick, S. and Fischhoff, B. (1997a) An empirical test of ‘adding-up.’ Restoring the American Timberwolf. Working paper, Carnegie Mellon University.
- Frederick, S. and Fischhoff, B. (1997b) Scope insensitivity at a contingent supermarket: examining the role of familiarity in scope sensitivity. Working paper, Carnegie Mellon University.
- Gregory, R. (1986) Interpreting measures of economic loss: Evidence from contingent valuation and experimental studies, *Journal of Environmental Economics and Management* **13**, 325–37.
- Gregory, R., Lichtenstein, S. and Slovic, P. (1993) Valuing environmental resources: a constructive approach, *Journal of Risk and Uncertainty* **7**(2), 177–97.
- Gregory, R. and McDaniels, T. (1987) Valuing environmental losses: what promise does the right measure hold? *Policy Sciences* **20**, 11–26.
- Hagen, D., Vincent, J. and Welle, G. (1992) Benefits of preserving old-growth forests and the spotted owl, *Contemporary Policy Issues* **10**, 13–26.
- Hahneman, M. (1994) Valuing the environment through contingent valuation, *Journal of Economic Perspectives* **8**, 19–43.
- Hicks, J. (1943) The four consumer’s surpluses, *Review of Economic Studies* **11**, 31–41.
- Hoehn, J. and Randall, A. (1989) Too many proposals pass the benefit cost test, *American Economic Review* **79**, 544–51.
- Kahneman, D. (1986) Comments, in: R. Cummings, D. Brookshire, and W. Schulze (Eds) *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, Totowa, NJ: Rowman and Allenheld.
- Kahneman, D. and Knetsch, J. (1992) Valuing public goods: the purchase of moral

- satisfaction, *Journal of Environmental Economics and Management* **22**, 57–70.
- Kahneman, D., Ritov, I., Jacowitz, K. and Grant, P. (1993) Stated willingness to pay for public goods: a psychological perspective, *Psychological Science* **4**, 310–15.
- Kahneman, D. and Ritov, I. (1994) Determinants of stated willingness to pay for public goods: a study in the headline method, *Journal of Risk and Uncertainty* **9**, 5–38.
- Kemp, M. and Maxwell, C. (1993) Exploring a budget context for contingent valuation estimates, in: J.A. Hausman (Ed.) *Contingent Valuation: A Critical Assessment*, Amsterdam: North Holland, 503pp.
- Knetsch, J. (1993) *Environmental Valuation: Some Practical Problems of Wrong Questions and Misleading Answers*, published for the Resource Assessment Commission, Canberra: Australian Government Publishing Service.
- Kopp, R. (1992) Why existence value *should* be used in cost-benefit analysis, *Journal of Policy Analysis and Management* **11**, 123–30.
- Loomis, J., Lockwood, M. and DeLacy, T. (1993) Some empirical evidence on embedding effects in contingent valuation of forest protection. *Journal of Environmental Economics and Management* **24**, 45–55.
- McDaniels, T.L. (1988) Comparing expressed and revealed preferences for risk reduction: different hazards and question frames, *Risk Analysis* **8**(4), 593–604.
- Milgrom, P. (1993) Is Sympathy an Economic Value? Philosophy, Economics and the Contingent Valuation Method, in: J.A. Hausman (Ed.) *Contingent Valuation: A Critical Assessment*, Amsterdam: North Holland.
- Mitchell, R. and Carson, R. (1986) *The Use of Contingent Valuation Data for Benefit-Cost Analysis in Water Pollution Control*, Report to USEPA, Washington, DC.
- Mitchell, R. and Carson, R. (1989) *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Washington, DC: Resources for the Future.
- Neill, H., Cummings, R., Ganderton, P., Harrison, G. and McGuckin, T. (1994) Hypothetical surveys and real economic commitments, *Land Economics* **70**, 145–54.
- Randall, A. and Hoehn, J. (1996) Embedding in market demand systems, *Journal of Environmental Economics and Management* **30**, 369–80.
- Samples, K. and Hollyer, J. (1990) Contingent Valuation of Wildlife Resources in the Presence of Substitutes and Complements, in: R. Johnson and G. Johnson (Eds) *Economic Valuation of Natural Resources: Issues, Theory, and Applications*, Boulder: Westview Press.
- Schkade, D. and Payne, J. (1994) How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation, *Journal of Environmental Economics and Management* **26**, 88–109.
- Schulze, W., Brookshire, D., Walther, E. and Kelley, K. (1981) *The Benefits of Preserving Visibility in the National Parklands of the Southwest*. Vol. VIII. Methods Development for Environmental Control Benefits Assessment. Report prepared for the USEPA, Washington, DC.
- Smith, V.K. (1992) Arbitrary values, good causes, and premature verdicts, *Journal of Environmental Economics and Management* **22**, 71–89.
- Stigler, G. (1966) *The Theory of Price*, New York, NY: Macmillan Publishing Company.
- Tolley, G. and Randall, A. (1986) *Establishing and Valuing the Effects of Improved Visibility in the Eastern United States*. Report to US Environmental Protection Agency, Cooperative Agreement 807768-01-0. Chicago: University of Chicago.
- Willig, R. (1976) Consumer's surplus without apology, *American Economic Review* **66**, 587–97.
- Whitehead, J. (1992) Ex ante willingness to pay with supply and demand uncertainty: implications for valuing a sea turtle protection programme, *Applied Economics* **24**, 981–988.