7.   Some findings from further exploration of the "composite good" approach to contingent valuation[1]

Michael Kemp, Edward Leamer, James Burrows, and Powell Dixon[2]

THE CONTEXT AND RATIONALE FOR THIS INVESTIGATION

In this chapter, we report on some components of a larger study addressing "budget awareness" considerations in contingent valuation ("CVM") studies. The study explored general population willingness to pay ("WTP") for essentially the same environmental resource under several different budget awareness approaches, using a split-sample design. In the fall of 2014 (after substantial development work), we undertook a survey with a total achieved sample size of almost 4,000 households. This chapter reports two of those approaches, based on subsamples totaling approximately 2,400 households: the use of within-questionnaire "wording additions" and the disaggregation of a value obtained for a larger composite good.[3]

   As well as the budget awareness focus, our study had a second major emphasis: on respondent cognition and on other survey design considerations in the use of CVM techniques to estimate non-use (or "existence")

[2]   Respectively, Senior Consultant, Charles River Associates; Chauncey J. Medberry Professor in Management and Professor in Economics and Statistics at UCLA; Vice Chairman, Charles River Associates; Associate Principal, Charles River Associates.

[3]   The survey additionally included a subsample presented with three comparable goods, and our analysis has also employed parametric methods (alongside the non-parametric analysis techniques presented in this chapter). It is anticipated that these aspects of the overall study will be reported in other papers.

values[4] among representative samples of the general public. Empirically, it has long been recognized that reminding respondents about their household budget constraints can sometimes significantly affect responses to valuation questions.[5] Specifically, there has been an effort to design surveys in ways that induce respondents to view their responses as *consequential* in terms both of contributing significantly to the framing of public policy, and of the financial impacts on their own households. The NOAA Blue Ribbon Panel explicitly endorsed such concerns too.[6]

To this end, a variety of approaches has been used within valuation questionnaires to intensify the perceived consequentiality of responses to the posited hypothetical. We refer to these within-questionnaire tactics as "wording additions." These approaches have often appeared to lower *some* of the hypothetical bias,[7] typically by reducing valuation estimates by levels in the 5% to 40% range, but rarely to an extent approaching the levels of "over-statement" observed in the meta-analyses of marketplace valuations.[8]

The Single-focus and "Composite Good" Approaches

The budget constraint that should underlie the answer to a single-focus willingness-to-pay question holds total spending constant so that the purchase of the item being studied comes at the cost of one or more goods currently consumed. WTP survey methods – whether involving direct (CVM) or indirect (choice experiment, or "CE") questioning techniques – usually embody this budget constraint by having respondents promise to give up cash to acquire the good or service being studied. That cash is probably interpreted by respondents in terms of the private goods it represents, since trading cash for public goods is an unfamiliar experience.

However, the cheapest way to pay for a public good might be by giving up an equivalent or a close substitute public good. Furthermore, the amount of the private goods one would surrender to fund a public good

---

[4]   While it is sometimes helpful to think of a non-use value (or "existence value") as just another public good, there are very fundamental differences. A non-use value cannot be observed from *any* marketplace behaviors, and (unlike other public initiatives) it has no substantive manifestation except in the *minds* of the affected population. The non-use valuation task is essentially one of finding credible, robust, reliable methods of eliciting those thoughts, when survey respondents have no prior experience (or marketplace frames of reference) for articulating any values for them.

[5]   See, for example, Kotchen and Reiling (1999) and Whitehead and Cherry (2007).

[6]   See Arrow et al. (1993).

[7]   See Morrison and Brown (2009) and Loomis (2014).

[8]   See Foster and Burrows (Chapter 10, this volume).

depends on the existence of already available close substitutes for the new public good, which may not be present (or prominent) in the respondent's thinking. For both of these reasons, it seems wise to encourage respondents to think about a *variety* of public goods as an alternative to the traditional single-focus survey.

Our study pursued this idea with a survey exploring the WTP *in aggregate* for a large basket of alternative public goods, including one specific program for which (with a separate sample) we conducted a single-focus valuation survey. We put the respondents through a top-down budget-allocation exercise in which they first prioritized the components of an extensive composite of environmental goods, before eliciting their WTP for the complete basket.[9] They were then given, as *one example* of the kinds of things the portmanteau program would do, the same details of the specific program of interest as were provided to the single-focus split sample. Following this extra information, respondents were given opportunities to adjust their binary-choice WTP responses and/ or their priority allocations among the components of the composite good.

A single-focus valuation method might be intended to estimate the amount that would be needed to compensate for some accidental damage (using WTP as a conservative approximation of willingness to *accept*). This single-focus question involves a different budget constraint that does not include public goods currently unavailable, but it does include public goods currently provided, and one might reasonably expect that WTP should depend on the existence of these close substitutes.[10]

As Hoehn and Randall (1989) and others have pointed out, the entities being valued by a single-focus survey and a composite good survey are not identical, and it should not be expected that the WTP value for the bundle equal the sum of the single-focus WTPs for each of the goods in

---

[9]   This exercise is close in spirit to traditional public budgeting, in which elected officials typically allocate a given tax revenue over a broad set of expenditure items, and consider tax increases only when the revenue seems inadequate to fund all the appealing programs. This contrasts with the (relatively unfamiliar) single-focus referendum that proposes a new tax specifically for funding one new public good. If the single-focus question were intended for public policy it would be essential that respondents recognize that this use of the tax revenue precludes other uses, something that is an explicit feature of the multi-good budget exercise.

[10]   Our study explored the possibility that the budgeting exercise makes respondents more aware of substitute public goods, and that a single-focus question asked after the budgeting exercise might produce a different (lower) valuation than the single-focus question without the budgeting exercise. In a sense, this is an alternative to the various "wording addition" treatments. Another potential benefit of a multi-good survey is that it may reduce the "importance" bias by including a very broad set of alternatives instead of one special one.

the bundle. Valuing a single good implicitly assumes that all other goods are being held constant, while valuing a bundle does not share the same constraint.

Both CVM and CE valuation methods are based on the assumption of rational and well-behaved neoclassical utility maximization by respondents, which is a questionable assumption in the case of unfamiliar, non-traded goods such as existence values for environmental amenities. There are several empirical reasons to be skeptical about the validity of the basic assumptions underlying this model. For instance, qualitative research associated with similar respondent tasks[11] frequently shows that, no matter how carefully a hypothetical scenario is crafted to present a credible *mandatory* payment mechanism, many respondents have a pervasive difficulty interpreting the valuation questions in ways other than being asked to consider a *voluntary* contribution.[12] And detailed post-valuation debriefing questions often show that relatively few respondents pass *all* of the criteria thought necessary for a valid interview, and specifically, few may pass a consequentiality filter.[13]

A different cognitive approach for exploring public valuations posits that, faced with a task to respond financially to a completely novel, hitherto unconsidered category of commodities, some respondents might prefer to tackle the task from a different perspective. Consistent with the "mental accounting" approach to the analysis of consumer choices,[14] several environmental valuation researchers[15] have explored constructs in which respondents determine their WTP by budgeting their wants in categories of similar goods or services. Not surprisingly, the quantitative values resulting from this approach can be considerably smaller than those from the single-focus methods.

[11]   See particularly the "cognitive interviews" (or "verbal protocols") reported by Schkade and Payne (1993). Similar evidence is frequently seen in focus groups or one-on-one intensive interviews.

[12]   Respondents often use terms like "give" or "donate" in this context, probably in part because being asked about how much one is prepared to pay for some governmental action is such an uncommon, perhaps unrealistic, phenomenon.

[13]   See, for example, Schläpfer and Brauer (2007) and Myers et al. (Chapter 9, this volume).

[14]   See, for example, Thaler (1985).

[15]   See Kahneman and Knetsch (1992), Magnussen (1992), Kemp and Maxwell (1993), Brown et al. (1995), Li et al. (2005), and Smith et al. (2005).

The Design of Our Study

We used Carson et al. (2004) as a "test bed" for our work, exploring how that project's findings might change were different approaches to budget awareness adopted in the study's survey. This "California oil spill" study ("COS") addressed the prevention and remediation of oil spills off the central California coast.[16] Based on a spring 1995 home-interview survey of a probability sample of about 1,000 English-speaking California households, the survey followed a dichotomous choice quasi-referendum structure.

The questionnaire described a scenario in which, as a direct result of oil spills anticipated off the central coast over the next ten years without further action by the State of California, 12,000 sea birds would be killed, 1,000 more birds would be injured but survive, and many small animals and saltwater plants would die over a total of about ten miles of shoreline. None of the involved species were considered "threatened," and these numbers represent minuscule proportions of the species populations. All projected harms would recover, naturally and fully, within ten years at the most, and the state initiative would no longer be necessary after ten years because of the full implementation by then of federal double-hulling mandates for oil tankers.[17]

Survey respondents were asked how they would vote on a state referendum about a hypothetical program that would provide for escort ships for *every* oil tanker sailing there.[18] The operating costs of this initiative would be borne by "the oil companies." However, "(b)ecause individual oil companies cannot legally be required to pay the cost of setting up the program, all California households would pay a special *one* time tax for this purpose." The referendum essentially offered complete and certain avoidance of the specified ten-year harms in return for approving a one-year state tax surcharge on the respondent's household.

The values that the 1995 survey sought to measure comprise *both* use

---

[16]   The methodologically detailed report allowed us to follow the design closely. Moreover, as an example of CV methodology, the COS study claimed to be "arguably the first and only valuation study to meet in full the reference study standards set by NOAA's Blue Ribbon Panel on Contingent Valuation. . .effectively a 'how-to' guide for undertaking state-of-the-art contingent valuation studies."

[17]   This scenario distorted the actual legal situation. The Oil Pollution Act of 1990 ("OPA") required single-hull oil tankers operating in US waters to be phased out, *starting* in 2005, with achievement deadlines set for 2010 and 2015. Thus, in 1995 the OPA requirement set a 20-year time horizon, not ten years.

[18]   Additionally, the program would establish three oil spill prevention and response centers along the central coast. Sea fences and skimmers would be used quickly by the escort ships to contain and remove any spilled oil, augmented by other response ships.

and non-use values for residents of California. However, since onshore effects are limited to the oiling of "about ten miles of shoreline," the level of harm that can be associated with "use" would be relatively small for the overriding proportion of the state's residents.

Reflecting interim changes in survey research methods, our own 2014 survey was carried out *online*. *Version 1* ("V1") of the study translated the 1995 COS questionnaire into an online, computer-assisted form, staying as close to the original as possible. It was completed by an achieved total sample of about 1,200 respondents. *Within* V1 we used split samples to investigate three different budget awareness-related enhancements to the original survey: a "cheap talk" entreaty consistent with modern practice, an "environmental contributions" script reminding respondents of their disposable income and allocation to environmental concerns, and a "dissonance minimization" variant of the referendum question (that gives a respondent the opportunity to indicate agreement with the identified cause while giving justification for voting *no*). We also tested a variant using a much shortened scenario description, necessary for use in other versions of the questionnaire.

The *Version 2*[19] ("V2") questionnaire also had an achieved sample size of almost 1,200 respondents. It presented a composite environmental good – a prospective "California Environmental Improvement Program" – and asked the respondent to prioritize spending under that program among several tiers of competing elements (including, in the last tier, "Reduce the risk of oil spills at sea off the California Coast").[20]

After the prioritization exercise, a referendum valuation question was asked in a manner closely analogous to that of the test bed study.[21] The 1995 oil spill prevention and remediation scenario was next introduced (in its shortened form), as an example of just one of the many programs that the proposed California Environmental Improvement Program might include, and (after that added detail) respondents were offered the opportunity to amend either or both of their earlier referendum response or their earlier priority allocations. Finally, the questionnaire asked respondents how they would vote in a referendum about the oil spill program alone,

[19]   In the questionnaire specification and all other survey documents, available on request to the lead author (MKemp@crai.com), this version is labeled Version 3. We use Version 2 in this chapter solely to avoid confusion here.

[20]   The questionnaire design was based largely on Kemp and Maxwell (1993), which reports a 1991 *pilot* study, carried out (after qualitative development work) using only a convenience sample. The matter settled before an anticipated subsequent production survey was undertaken.

[21]   Version 2 did not employ any of the budget awareness wording variants tested in Version 1.

if it were to appear on the ballot *in place of* the much larger portmanteau program.

THE SINGLE-GOOD VERSION

Similarities with and Differences from the 1995 Test-bed Study

To take advantage of the substantial development work carried out for the 1995 COS survey, we designed our own V1 baseline survey to align as closely as possible to the original, particularly with respect to question wordings and the choice of bid levels ("design points"). However, with a 19-year gap since the original fieldwork, some changes and differences in survey methods were inevitable.

First, we changed the mode of administration from an in-home face-to-face interview to an online, self-administered survey, using a high-quality online survey platform designed to provide a probability sample representative of the general population of California households, regardless of their telephone, computer, or online access status.[22]

Second, presenting the detailed COS scenario had occupied a large proportion of the 1995 interview time, with several scripted points at which the interviewer would engage the respondent by asking a general interest question or checking on comprehension. Such personal interactions could not be emulated fully online. And the 1995 survey used procedures designed to minimize non-committal "don't know" ("DK") responses to the referendum valuation question. After the lengthy COS scenario exposition, respondents were asked just "[. . .]would you vote *for* the program or would you vote *against* it?" It seems impolite (at a minimum) – or unengaged, unintelligent, or lazy – to respond after that effort that one had no opinion about voting for or against the proposed program.[23]

---

[22]   We used the pre-recruited KnowledgePanel[SM], maintained by GfK Custom Research (formerly Knowledge Networks), which recruits participants using an address-based random sampling technique and provides Internet access free of charge to any recruit who doesn't already have such access. In recent years, the same panel has been used for several large-sample environmental valuation studies on behalf of NOAA and other governmental agencies (see, for example, Bishop et al., 2011 and Wallmo and Lew, 2011). Our sample used *all* available California respondents in the KnowledgePanel[SM], and screened to ensure that the respondent was "one of the people who make most of the decisions about major purchases and other financial matters for your {household}."

[23]   Moreover, to such responses as "I don't vote," "I'm not registered," "I'm not a citizen," the interviewer was instructed to probe with "If you did vote. . .would you vote for the program or against it?"

The NOAA Blue Ribbon Panel had recommended a referendum format valuation question, with the inclusion of an *explicit* "no-answer" option.[24] There has been considerable debate in the survey research literature about the pros and cons of including explicit "no-opinion filters,"[25] and the balance of much current commercial market research practice appears to favor excluding them. We judge this practice troubling on several grounds. Most importantly, we worry about the (conscious or subconscious) alienation of those respondents who do not find an acceptable, truthful choice among the range of options proffered them, and how such alienation might affect their overall perception of the researchers' interest in accuracy and truthfulness throughout other parts of the questionnaire.[26]

Accordingly, we opted to include an explicit "I'm not sure" response option for our valuation questions, which led to a significantly higher level of non-committal responses. In the 1995 survey, about 4% of respondents gave a DK response; for our V1 respondents, the proportion choosing "I'm not sure" for the same question was over 19%.[27]

Other wording changes reflected developments in the interim between the two surveys. The COS study scenario had envisaged a ten-year program that would then become redundant by the federally mandated migration to double-hulled tankers. While information about the extent of double-hulling along the Central California coast by 2014 was sparse, we changed the proposed program slightly (lest some respondents should doubt the scenario's credibility) to include some additional dredging work and changes in navigational rules.[28]

The 1995 survey had employed five "design points" (or "bid levels") at which different subsamples were offered the subject environmental good.[29] We generally inflated the 1995 survey by about 55% to reflect

---

[24]   See Arrow et al. (1993).
[25]   See Krosnick and Presser (2010), at §9.6, for a review of this literature. In connection with CVM studies specifically, see Krosnick et al. (2002).
[26]   Also, the inclusion of an explicit DK option more realistically represents the real-world situation of a statewide referendum, where a voter may opt not to vote at all in a particular election, or if voting, not to provide a response to any specific initiative question on the ballot. We note also that the circumstances of an *online survey* much more closely resemble those of an actual referendum than does a lengthy, in-home, face-to-face *oral interview*.
[27]   As with the 1995 survey, the proportion of DK responses increased with the bid level.
[28]   Additional changes from the 1995 survey design expanded the sample scope to include Spanish-speaking (as well as English-speaking) households, and removed any reference to the participation of state and federal agencies in the sponsorship of the study.
[29]   While mentioning that the five points were selected on the basis of the questionnaire development work, the COS report does not identify the specific criteria used in making those decisions, nor justify the critical choice of $220 as the uppermost design point, an amount

interim inflation,[30] resulting in Version 1 design points of $7.50, $40, $95, $185, and $340.

Some Comparisons with the 1995 Study Findings

In analyzing their referendum responses, Carson et al. (2004) use a non-parametric, maximum likelihood estimator of the lower bound on the mean WTP, following Turnbull (1976).[31] This statistic is derived from the response curve of the subsamples voting "*not-for*" the hypothetical referendum at each of the proffered design points. In both the 1995 and 2014 surveys, the *not-for* respondents are those who did not vote an explicit *yes* to the referendum, and they comprise both explicit *no* votes and those who did not express an opinion.

Theoretically, one would expect the proportion of *not-for*s to increase monotonically as the design point increases. But natural sampling variations and number cognition–related effects may mean that this is not always the case, and we refer to such a situation as the "non-monotonicity issue."[32] In the analysis of the 2014 survey data, we frequently encountered subsamples of interest for which the *not-for* responses were non-monotonic. Following a now common practice in the non-market valuation literature to cope with this situation, we used the ABERS approach.[33]

---

that does not appear to have been tested in any of the development work. See Carson et al. (2004), at pp. 20–21 and 225–8.

[30]   As measured by the consumer price index for all urban consumers (CPI-U) as the 2014 questionnaire was being finalized. Inflated values were mostly rounded to the closest $5, except that the lowest value was kept at $7.50 and the value that would otherwise have been $100 was set at $95 as a result of non-monotonicity issues detected in our pilot survey.

[31]   See Carson et al. (2004), at Appendix F.

[32]   The original COS study did not have this problem, because *all* of the *not-for* distributions reported were monotonic. Given the magnitude of the confidence intervals for the individual design point *not-for* percentages, this outcome is somewhat surprising.

[33]   Ayer et al. (1955) – often referenced in the literature as "ABERS" – and Turnbull (1974, 1976) describe non-parametric estimators for the lower bound on the mean WTP that have frequently appeared in the environmental literature. ABERS provide a method for smoothing non-monotonic empirical distributions. Turnbull extends ABERS and Kaplan and Meier (1958) to accommodate doubly bounded data, but for single-bounded dichotomous choice data, the ABERS and Turnbull techniques yield identical results. The ABERS algorithm compares proportions for adjacent design points, and when the *not-for* proportion does not increase monotonically as the design point increases, the affected subsamples are pooled. That pooled proportion is then compared to the proportion for the next higher design point. This "pooled-adjacent-violators" algorithm continues until the estimated empirical cumulative distribution is monotonic. Not surprisingly, the resulting estimates are sensitive to the actual categories pooled by the algorithm, and it is not clear how best to compare estimates derived from different patterns of pooling of adjacent design point subsamples.

Three different "choice measures" were discussed in the report of the 1995 survey. The first (labeled **B1**) was based on the answers originally given by respondents to the referendum question, while measure **B1CH** (the one most relied upon by the authors) reflected two possible subsequent changes to the initial vote. The third measure (**B1CHNT**) also used the amended votes but additionally edited any *for* votes from households who were not currently paying California income taxes to *not-for* votes.

For the **B1** measure, we estimate[34] that the lower bound on the sample mean WTP per household was $90.32, which translates to $141.1 in 2014 prices.[35] This estimate fell to $85.39 ($133.4 in 2014 prices) for the **B1CH** measure, and fell further to $76.45 ($119.4 in 2014) when non-taxpayer *for* votes were also changed to *not-for*.

One-eighth of our 2014 V1 sample (153 respondents) had a questionnaire most closely matching the 1995 survey. These were the respondents for whom the COS scenario was described using the "long" form, and who did not experience any "wording additions" intended to enhance budget awareness.[36] For this group, using initial referendum responses in accordance with the 1995 study's **B1** measure, we estimated a lower bound on the sample mean WTP per household of $100.5,[37] about 70% of the 1995 study value at 2014 prices.

Our closest feasible replication[38] of **B1CH** from those 2014 respondents answering the most analogous questionnaire was $92.6. And the closest replication of **B1CHNT** for those same respondents was $68.1.[39] These estimates (both of which use *unweighted* data, to align with the 1995 survey treatment, and derive from monotonic distributions) are approximately 69% and 57% of the 1995 survey values expressed in 2014 prices.

---

[34] The COS report does not explicitly present this figure, but we have estimated it from the underlying frequency of *not-for* votes by design point.

[35] The 2014 estimates for values derived from the 1994 survey reflect an approximately 56.2% increase in the consumer price index (CPI-U) between the two surveys.

[36] However, this subsample did include some interviews completed in the Spanish language, excluded by design from the 1995 survey.

[37] This estimate derived from a non-monotonic distribution, and involved pooling. To match the 1995 treatment, it is derived from *unweighted* data.

[38] Because our questionnaire did not include the original survey's opportunity to revise the referendum vote at the very end of the interview, we were unable to match the 1995 **B1CH** and **B1CHNT** choice measures exactly.

[39] 15.9% of the sample were non-taxpayers, and the fraction of these making *for* votes (after any vote revisions) was 13.3%.

It appears, therefore, that the net effect of all of the factors that differentiate the 2014 survey methodology from the 1995 original has been to *lower* these key WTP statistics by roughly 30% to 45%.[40]

Moving from just the one-eighth of the V1 sample that most closely matched the 1995 treatment to the full Version 1 sample of a little over 1,200 respondents, the estimated lower bound on the sample mean WTP for the full sample was $105.5.[41] This full sample comprised subsamples given both long and short forms of the COS scenario, and (orthogonally) four different "wording addition" treatments (including one with *no* wording additions).

We investigated the effect of using the long and short COS scenario descriptions.[42] The weighted WTP statistic (based on original, unadjusted responses) was $93.0 for the long scenario and $118.8 for the short scenario, both derived from monotonic, unpooled response data. And the proportions of *for* votes were higher for the short scenario at all five design points. If the only response effect at play here were an importance bias, then one might expect that the longer, more detailed exposition would increase the expressed value. But other factors also appear to be at work. Perhaps some specific language omitted from the longer version reduced the respondents' perception of possible reasons to reject the project (concerning, for example, the abundance of the damaged species, the rate of natural restoration, or the substitution possibilities), or perhaps some aspect of the greater detail impaired credibility, or increased respondent frustration with the exercise, for example.[43]

---

[40]   These factors include, most importantly (we would judge), the translation from a face-to-face to a (more "anonymous") online administration mode, the use of computer-assist to greater personalize the questionnaire, the adoption of an explicit "I'm not sure" voting option, and the 19-year gap between the two survey periods.

[41]   The *issued* sample was designed to provide an accurate representation of the California non-institutionalized population, but differential non-response can lead an *achieved* sample to differ from reference proportions derived from Census data. Individually for each of the V1 and V2 samples, we weighted the achieved sample composition to more closely align with geographic, demographic, and socioeconomic reference proportions for the state. Here and subsequently throughout this chapter, unless otherwise indicated, the presented statistics are those for the *weighted* sample. Weighting effects were small in most instances, and typically did not materially influence our conclusions.

[42]   In these comparisons, the wording addition variations were distributed very close to equally for both the long and short scenario descriptions.

[43]   While the shorter version included all salient facts regarding expected harms (in the absence of the COS program) and the specific initiatives proposed to prevent or remediate any spills, it lacked (1) a number of diagrams and maps from the longer version, (2) some repetition and added emphasis around the types of shoreline and species affected, and (3) intermediate questions designed to maintain respondent interest and attention.

THE COMPOSITE GOOD VERSION

Valuing the Version 2 Composite Good

Since single-focus, direct questioning CVM methods are known to be very sensitive to context and survey design considerations,[44] it is sensible to anticipate that the composite good approach would be similarly sensitive to design artifacts. It seems highly likely that the results from a disaggregation survey are significantly sensitive to the somewhat arbitrary assumptions used in defining both the parent good and the disaggregation structure.

The V2 questionnaire incorporated several changes from the Kemp and Maxwell pilot study for a variety of reasons, not least that the switch to a *computer-assisted* questionnaire provides the possibility of framing the questions in a much more personalized way than was feasible with in-person, pencil-and-paper methods. More recent empirical explorations of applying CV methods to composite and component goods[45] have focused on much more tightly defined situations, with relatively constrained sets of choices among which to allocate the funds.

The 1991 experiment had asked its valuation question early in the questionnaire, *before* setting out the list of specific issues or programs that the aggregate good might include. It is possible that the specific details provided later in the disaggregation tree might incline respondents to indicate a greater WTP if the valuation question (changed to a referendum, dichotomous choice format) were delayed until *after* the disaggregation information, and this is the approach we took in the 2014 survey.

Version 2 of the questionnaire posited a very expansive new environmental program on the part of the State of California.[46,47] Then, after

---

[44]   Not least, because a non-trivial proportion of respondents appear ready to assent to the proffered good no matter how high the price is set, the selection of the highest design point is a crucial determinant of the measured WTP. See, for example, Parsons and Myers (2016, and Chapter 2, this volume).

[45]   See, for example, Magnussen (1992), Bateman et al. (2003), and Boxall et al. (2012).

[46]   "Consideration is being given to introducing a new program – the "*California Environmental Improvement Program*" – that will help *protect California's environment and natural areas* over the next ten years. We will ask some questions to find out which environmental issues are most important to {your household}, to find out what types of environmental improvements should be included in the program if it is to appeal to the majority of California's residents.

To pay for starting up the program, and for costs that cannot legally be charged to private companies, there would be a special *one-time* tax for this purpose."

[47]   The computer-assisted questionnaire tailored second person references to the most appropriate form for the circumstances of the particular respondent: "you," "your household," or "your family unit." In footnote 46, and subsequently when quoting from the

respondent prioritization of the wide spectrum of environmental ini-
tiatives that such a program might comprise (listed in Table 1), the V2
respondents were asked a valuation referendum question similar in form
to the V1 valuation.[48]

The payment scenario was directly analogous to that posited in the
original COS study: the responsible government entity would impose a
one-year income tax surcharge to fund a *ten-year* government program.
The V2 design points were set at four times those used for V1: $30, $160,
$380, $740, and $1,360. We added a sixth design point, set at $2,500, to
accommodate the greater scope of the proposed portmanteau good.

For the 1,172 completed V2 interviews, the *not-for* vote percentage
ranged from 52% at the lowest ($30) amount to 90% at the highest ($2,500)
level. However, there was non-monotonicity for the $740 and $1,360 bid
levels, which necessitated pooling of those two subsamples to estimate the
proportion of *not-for*s. The ABERS estimate for the lower bound on the
sample mean WTP is $413.2[p].[49] Hence this statistic for the large portman-
teau program is about 3.8 times that for the COS program alone from the
V1 sample[50] (a multiplier not much different from the factor of four used
to inflate the design points).

questionnaire, for ease of reading we use the "your household" form, and indicate the exist-
ence of the tailoring {thus}.

[48]   "Now that you and other survey participants have told us what environmental improve-
ments would be most important to {your household}, we want to find out how California
residents would be likely to vote on the *California Environmental Improvement Program* if it
were put on the ballot in a California election.

Even though (to shorten the interview) we have only asked you about just *some* of the
environmental improvements that you think are important, we want you to assume that the
ballot proposition on which you would be asked to vote will make significant improvements,
over the next ten years, on the environmental issues that are important to you and to other
California residents. Please assume also that it would do so efficiently, without wasting money.

Remember that to pay for setting up and administering the program, and for costs that
cannot legally be charged to private companies, there would be a special *one time* tax *added to*
your California income tax for one year. You would pay this tax increase in the same way as
you usually pay any California income tax now, so for many people the extra amount would
increase the amount that is withheld from their pay checks. . .but the extra tax withholding
would end after one year.

If the *California Environmental Improvement Program* were put into place, it would cost
{your household} $X in a special *one time* tax *added to* your next year's California income tax.

If a California-wide ballot proposition about this proposal were on an election being held
today, and the total cost in extra tax to {your household} for the program would be $X, would
you vote *for* the program or would you vote *against* it? [*Select one only.*]"

[49]   Here and subsequently, we use the superscript [p] to identify any WTP estimate based on
pooling of design point subsamples.

[50]   To make best use of the available 1,213-member V1 sample, here (and in the discussion
that immediately follows), we used the *whole* of the sample, including both long and short
COS scenarios and all of the tested wording additions, rather than just those 153 V1 respond-
ents whose questionnaires most closely followed the 1995 survey.

Priority Allocations for Components of the Composite Good

Similar to the disaggregation structure developed (on the basis of quali-
tative survey research) for the 1991 experiment, respondents were asked
to allocate the total funding for the proposed California Environmental
Improvement Program in three stages. At each stage, the allocation
question involved assigning 100 cents across a range of specified choic-
es.[51] Figure 1 summarizes the logical structure that was followed to
disaggregate from the complete California Environmental Improvement
Program down to the share for marine oil spills off the California
coast.

If at any stage a respondent chose *not* to allocate any money to the
category that was of particular interest – that is, in the direct path to the
next stage of the disaggregation – the questionnaire conservatively tried to
ensure that this decision was not due to a lack of understanding about that
category.[52] Other language was used to suggest that the particular disag-
gregation path chosen for questioning was randomly determined,[53] or to
justify following a path for which the respondent had previously allocated
a zero amount.[54] All of these features had been developed for the 1991
pilot.

With the translation to a *computer-assisted* online survey we were able
to *offer* respondents the opportunity to learn further details about the
alternative choices that they were being asked to prioritize at each stage of
the disaggregation. A respondent could find out more about the allocation
options (definitions, examples, and representative "prices") in two ways.
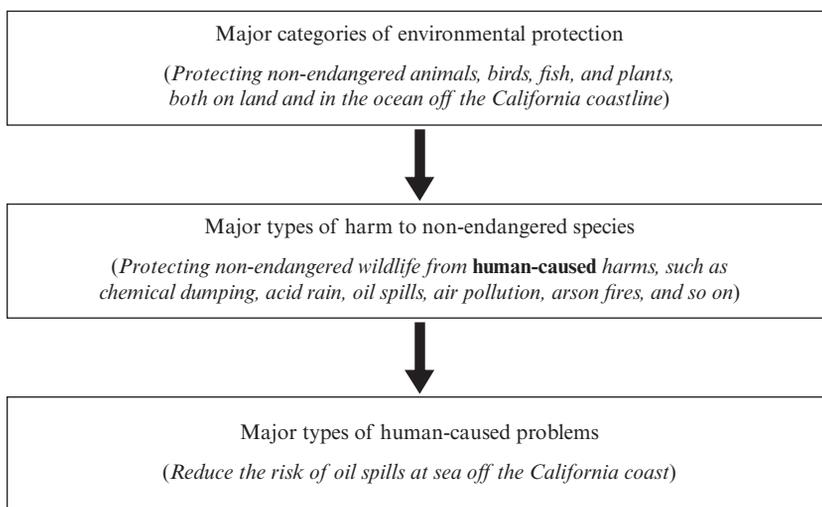Hovering the cursor over applicable text on the screen brought up a small

---

[51]   "Out of *every additional dollar* – that is, 100 cents – that {your household} would have
to pay in a one-time state tax to fund the program, how many of the extra cents should go to
each of the issues on the list below?"
[52]   "Some people think that California should do more to. . .[expanded explanation of the
category of interest]. We see that you have allocated none of {your household's} additional
tax to that category. Is that what you intended, or would like to make some changes to how
you allocated your payments?"
[53]   "To keep this survey from being too long, we will focus on just <u>one</u> of the environmen-
tal issues for each of the people taking the survey.

In your case, we are going to ask for your opinions about: [short delay before adding
remaining text]{Protecting <u>non-endangered</u> animals, birds, fish, and plants, on the land and
in the ocean off the California coastline}."
[54]   "Even though you've told us that you do not want any of your money to go to this
particular issue, we're going to suppose that some of your one-year tax surcharge would in
fact be spent on that category." In this way, unless the respondent replies with a "not sure"
answer at any of the disaggregation stages, the questionnaire still obtains each respondent's
allocation priorities throughout the full structure.

Major categories of environmental protection
(*Protecting non-endangered animals, birds, fish, and plants, both on land and in the ocean off the California coastline*)

Major types of harm to non-endangered species
(*Protecting non-endangered wildlife from **human-caused** harms, such as chemical dumping, acid rain, oil spills, air pollution, arson fires, and so on*)

Major types of human-caused problems
(*Reduce the risk of oil spills at sea off the California coast*)

Note: The particular category identified in italics is the one of primary interest at each stage of the disaggregation.

Figure 1   Logical structure used for the disaggregation

box with a brief amplification, or still more detail could be obtained (on a separate screen) by clicking on the text as a hot link.[55]

Table 1 shows the alternative environmental improvement categories offered at each stage of the disaggregation, and summarizes the mean allocations of resources for each of the three tiers of the disaggregation process. The data in the table imply that the proportion of the total California Environmental Improvement Program that respondents would, on average, dedicate to the category of interest – "[r]educe the risks of oil spills at sea off the California coast"[56] – was ($\sim 0.071 \times \sim 0.329 \times \sim 0.258 =$) 0.60% (with a standard error of 0.03%). For just those people who (later)

---

[55] A similar approach to providing details that respondents might or might not wish to learn was used in a recent, large, online choice experiment study sponsored by NOAA. See Wallmo and Lew (2011).

[56] Note that, for tractability reasons, this category of interest is a little different from the focus of the original COS study (and our own V1 replication) in two principal ways. First, "at sea off the California coast" is a more expansive geographical region than that defined for the COS program. Second, "Reduce the risk of oil spills at sea" makes no reference to the *remediation* activities described as part of the program, just as other details (e.g., the cause and magnitude of the harms, projected recovery times, and the program scale) also exceed the detail feasible in this approach.

Table 1   Initial mean allocations across environmental categories

Subsample	Allocation (%)	St. Err. (%)
First tier allocation:		
Protecting the quality of California's drinking water	19.1	0.7
Developing renewable sources of energy, like solar and wind power	14.2	0.5
Encouraging greater energy conservation	10.3	0.3
Improving disposal of household/industrial wastes, encouraging recycling	9.3	0.3
Protecting types of animals, birds, fish, and plants that are endangered	8.8	0.3
Enhancing the safety of nuclear power plants in California	9.3	0.4
Reducing air pollution (other than greenhouse gases)	8.6	0.3
Reducing the production of "greenhouse gases"	7.8	0.3
Protecting non-endangered animals, birds, fish, and plants, both on land and in the ocean off the California coastline	7.1	0.2
Other (write-in) topics	5.5	
Total	100.0	
Allocation of "protecting non-endangered species":		
Protecting non-endangered wildlife from human-caused harms (such as chemical dumping, acid rain, oil spills, air pollution, arson fires, and so on)	32.9	0.7
Preserving habitats in a wild state rather than land development	28.5	0.6
Protecting undeveloped areas or wildlife from harm from natural causes	21.3	0.5
Wildlife reduced through hunting or fishing for food or sport	17.2	0.5
Total	100.0	
Allocation of "protecting non-endangered wildlife from human-caused harms":		
Wildlife harmed from chemical or oil spills on the land or in rivers	31.4	0.6
Reduce the risks of oil spills at sea off the California coast	25.8	0.5
Human-caused fires in California wilderness areas	25.3	0.6
Other human-caused harms to California wilderness areas and wildlife	17.5	
Total	100.0	

Note: Descriptors of categories have sometimes been shortened, except for those (italicized) descriptors on the disaggregation path.

voted *for* the referendum, the overall allocation to the COS good was similarly 0.60%.

So the lower bound on the sample mean WTP for the portmanteau good was $413.2^p (based on pooled design points), and the mean allocation of

composite program dollars to marine oil spills was 0.60%. Under this particular disaggregation structure, therefore, the lower bound mean WTP to reduce the risks of oil spills off the California coast is approximated by ($\sim$\$413.2^p $\times$ $\sim$0.006 =) \$2.47^p.[57]

Reactions to the Standalone COS Scenario

At a later stage of the V2 interview, respondents were given an opportunity to revise their referendum votes and/or their allocations of portmanteau program revenues. In both cases, these revisions followed the presentation of the short version of the COS scenario, described as "an example of just *one* of the many ideas that have been proposed for funding under the program."

In total, just over a quarter of unweighted V2 respondents revised their referendum votes following the COS scenario.[58] The net effect of presenting the COS scenario was to increase the *for* proportion at the expense of both the *against* and *not sure* votes. In general, the propensity to shift votes was slightly higher at the higher bid levels, resulting in a relatively large impact on the WTP statistic. Using the revised referendum responses, the ABERS estimate for the lower bound on the sample mean WTP for the portmanteau program adjusted to \$639.5^p.

After any revision made to the referendum vote, respondents were shown the "bottom line" monetary implications (that is, the percentage allocation multiplied by the design point) of their previous allocation decisions, focusing in particular on the last tier of the allocation but with amounts also representing the totals from higher tiers of the disaggregation. We displayed these amounts, and provided the opportunity for the respondent to adjust the allocation.

About 40% of the unweighted respondents (45% of those voting *for* the program, after reconsideration) decided to make some adjustments to their implicit monetary allocations. Of those choosing to adjust their

---

[57]   Statistical dangers lurk in the combination of two independently determined distribution means. Depending on how the two variables are distributed interactively across the sample, the product of the means for each variable may not equal the mean of the products calculated individually for every sample member. Other factors equal, such a discrepancy is likely to be greater as the correlation between the two variables increases. This was not a significant problem in the current case.

[58]   The 24% of respondents who had originally voted *not sure* were responsible for the highest number of changes, with about 72% of their changed votes moving to *for*. The second largest number of vote shifts was among the respondents initially voting *against* the composite good referendum – almost 21% of those votes were changed, and almost three-quarters of the changed votes went to *not sure*. The initial *for* votes were least affected. Only 11% of those votes were switched, with 71% of the changes ending up as *not sure*.

allocations, 67% *increased* the amount apportioned to "oil spills at sea," while 9% *decreased* the amount and 24% left that particular item unchanged. The proportion of respondents not allocating anything to marine oil spills fell from 22% to 15%.

The general pattern of allocation shifts was to move money (in net) to *all* of the bottom tier options. "Oil spills at sea" benefited the most, but the amounts assigned to chemical/oil spills on land, human-caused fires, and other human-caused problems increased as well. At the next tier up, even the average amount allocated to protecting wilderness areas and wildlife from *natural harms* increased. All of these increases came at the expense of the catch-all category "All other environmental problems" – the respondents "went to the bank" to benefit the issue that had just been described to them, along with others sharing some similar characteristics.

How far the incidence and magnitude of the adjustments are ascribable to specific features of the COS scenario, or would be replicated for *any* single-focus environmental initiative that might have been presented at that point, is impossible to say, since we did not vary the scenario that we chose to present as "an example of just *one* of the many ideas that have been proposed for funding under the program."

Accepting *all* of the reallocations of budget share to marine oil spills, the mean share of the portmanteau program revenues increases to 3.9%, roughly 6½ times the original allocation. But this figure is highly influenced by a relatively small number of people making very large shifts to increase the marine oil spills share (as we later amplify). For example, if one edits to exclude the 6.4% of revised responses that allocated 20% or more of their total bid level amounts to marine oil spills, this adjusts the mean revised allocation to 2.4%, and if revised allocations of 10% or over are excluded (15% of the sample), the mean allocation for the rest of the sample is 1.0%.

Values for the Standalone COS Program

At the end of the Version 2 questionnaire, immediately before the debriefing and classification questions, respondents were asked how they would vote if, instead of the portmanteau California Environmental Improvement Program, the referendum were limited to just the COS scenario alone.[59]

---

[59] Among other things, we were interested to explore whether and how the "training aspects" of the top-down disaggregation logic – the fact that respondents had been exposed to a much broader litany of environmental concerns (many with potential harms of a much greater magnitude than those described in the COS scenario) – might affect their valuations of COS as a "standalone" good. For the standalone COS question, we used the five COS design points from V1.

Table 2   V2 ABERS estimates of the lower bound on the sample mean WTP for "standalone COS"

Subsample	Unweighted n	WTP ($)	St. Err. ($)
Version 1 respondents with short scenario:			
All (including wording additions)	603	118.8	10.0
No wording additions	153	146.9[p]	na
Version 2 "standalone COS" responses:			
All respondents	1,172	107.9	7.0
All voting *for* the portmanteau program	853	86.2	8.0
All voting *not for* the portmanteau program	319	159.8	14.7
All except those with $\Phi \geq 0.75$	900	126.7	10.0
All except those with $\Phi \geq 0.25$	786	132.6[p]	na
All except those with $\Phi \geq 0.10$	429	91.7[p]	na
All except for "yea-sayers"	654	79.4	8.1
All with marine oil spill share < median value	586	93.0[p]	na
All with marine oil spill share ≥ median value	586	121.7[p]	na

Note: The variable Φ is defined as the ratio between the (randomly assigned) design point for the standalone COS exercise and the previously seen, randomly assigned design point for the portmanteau program. Estimates with the superscript [p] are derived from pooled models, and no standard errors were computed in those cases.

Table 2 summarizes the ABERS estimates derived from responses to the standalone COS referendum. Across the full V2 sample the *not-for* proportions rose monotonically with the bid level, and consequently no pooling across adjacent design points was necessary. The WTP estimate (again, the lower bound for the sample mean) for the most comparable unweighted subsamples from Version 1 were around the $120 to $150 level. By comparison, for the full, unweighted V2 sample the equivalent WTP measure falls to about $108.

KEY FINDINGS OF THIS STUDY

1. **The study evidenced a very marked lack of sensitivity to a huge scope difference.**

We reported earlier that the ABERS estimate for the lower bound on the sample mean WTP for the V2 California Environmental Improvement Program was $413.2[p], about 3.8 times that for the COS program alone

from the V1 sample.[60] Recall that the bid levels for the composite good were set at four times those for V1 COS, with the addition of an extra bid level at the top of the scale. The similarity between the ratio of design points and the ratio of estimated WTP is remarkable. Such a result echoes the finding of Burrows et al. (Chapter 1, this volume) that a fourfold increase in the cost scale for endangered and threatened species resulted in an approximate tripling in estimated WTP. Results like these strongly suggest that there is a scope failure, with respondents providing WTPs for the lower-scope good as if they were valuing the larger-scope good.

Closer examination revealed that this fourfold difference between the WTP statistics for the composite good and the much smaller COS good is driven by the three highest bid levels for the portmanteau good, and *not* by any apparent difference in the WTP responses at similar price levels. Figure 2, which shows the proportion of *not-for* votes by design point for both cases, reveals an effect that is masked when just the WTP summary statistics alone are compared.[61] Where the two bid level ranges overlap in the lower part of the response curve, the two graphs are barely distinguishable after considering sampling errors.

The similarity of these two response curves reveals a colossal scope failure. If the survey responses are to be believed, at least for the overlapping portions of the cost scales the expressed WTP for a good that effectively encompasses the entire environment is virtually identical to the expressed WTP for a trifling fraction of the larger good: avoiding harms to minuscule proportions of the populations of a small number of non-threatened species, that will all recover naturally by no later than ten years.[62] In other words, within the $30–380 bid level range (which accounts for 100% of the V1 sample and 50% of the V2 sample) respondents apparently see very little difference to them in the value of this single-good COS program and the substantially more ambitious California Environmental Improvement Program.

Clearly, it cannot be the case that the "true" utility of the COS good alone is the same as the larger good incorporating the whole gamut of

[60]   To make best use of the available 1,213-member V1 sample, here (and in the discussion that immediately follows), we used the *whole* of the sample, including both long and short COS scenarios and all of the tested wording additions, rather than just those 153 V1 respondents whose questionnaires most closely followed the 1995 survey.
[61]   Figure 2 displays the one standard error confidence bands of *not-for* votes at each design point for both the V1 and the V2 composite good referenda. This figure uses a logarithmic transformation of the design point axis to improve clarity in the region of greatest interest.
[62]   As Table 1 shows, V2 respondents initially allocated only 0.6% of their composite program budget to marine oil spills off the California coast.
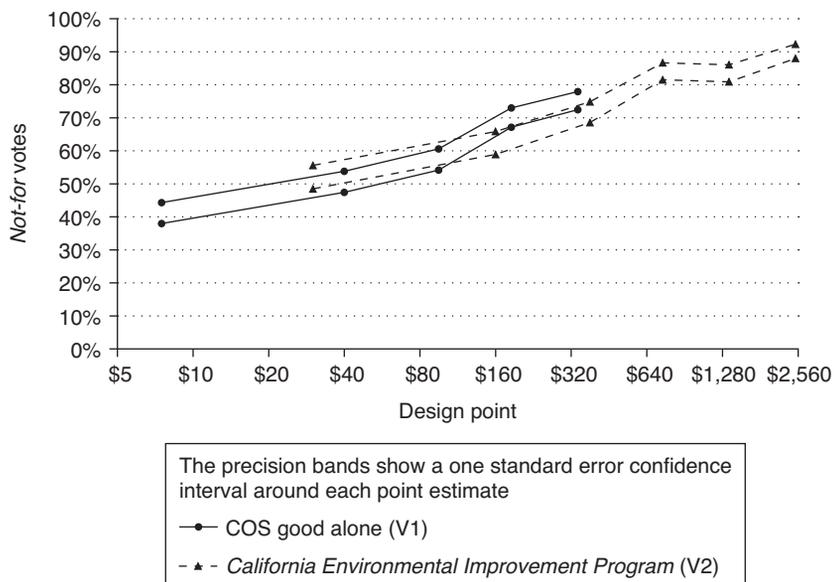
*Figure 2   Confidence bands for referendum responses to COS and the
composite good, by design point (log scale)*

state environmental policies. The only interpretation that makes sense is
that in the V1 COS referendum a sizeable proportion of the sample is
providing a WTP response for environmental goods in general, or perhaps
for "good deeds" in general, and not the WTP for COS specifically. It is
not credible that the anomalous findings are the result of a monumental
embedding problem in which respondents *mistake* the COS good for the
portmanteau good. As the respondents valuing the COS good are provided
information only about that single good, it is not likely that their responses
are really providing a value for the whole environment. An alternative
and much more credible explanation is that respondents are providing
a symbolic measure of the warm glow benefit of making a contribution
(*any* contribution) to "the environment."

2.   The composite good estimate of WTP allocated to marine oil spills is markedly smaller than the single-focus estimate.

The composite good approach used in Version 2 resulted in estimates
of the WTP statistics that were considerably smaller than those obtained
from the single-focus CVM questionnaire of Version 1 (from around

$120[63] for the single-focus estimate to, from disaggregation, $2 to $3 before reallocation and $15 to $20 after reallocation).

This conclusion per se is neither particularly novel nor surprising. Consistent with the analysis of Hoehn and Randall (1989), the two different approaches are measuring two different *entities*. For one of them, respondents are shown just a single public good in isolation, and are expected to articulate whether its non-use value to them is greater or less than a given amount. For the second, the subject public good is but one possibility among a plethora of similar goods (including possible substitutes), part of a larger category for which the respondents might have an aggregate non-use value.

A potentially valuable side benefit of the disaggregation approach is that it allows us to identify other environmental goods that have utility levels *less* than COS (consider giving these up to fund COS) and other environmental goods that have utility levels *greater* than COS (consider these as ways to compensate for COS-like environmental damage). Our results showed that when respondents had the opportunity to allocate dollars over a wide range of different potential state environmental initiatives, they did not find "Reduce the risks of oil spills at sea off the California coast" to be very appealing in comparison to the other public goods being considered, allocating only 0.6% of the total budget to that category. This means that there is not much scope for giving up other environmental goods to fund a COS-like initiative, but there are lots of ways of compensating for COS-like damage with more attractive alternative environmental goods.

3.   Sizeable proportions of respondents reported cognition difficulties in their responses, and the resulting WTP estimates are sensitive to those difficulties.

As had the original COS Study,[64] we explored the sensitivity of the WTP estimates to excluding from the analysis various categories of respondents who might, for various reasons, be judged of particular interest. For the 2014 survey, we added a markedly more extensive battery of post-valuation debriefing questions to those used in 1995.[65]

---

[63]   This estimate is specifically for those V1 respondents who saw the "short scenario" also used for V2.

[64]   See Carson et al. (2004), at Table 6.4.

[65]   Most of the additional questions used this rubric:

"We have a few more questions about what things were important to you as you answered our earlier questions about voting for or against the proposed {insert program name}.

Different people go about answering questions like these in different ways. On the screens that follow, we will show a number of statements that other people have made about *how* and *why* they answered the questions in the way that they did.

For each of the V1 and V2 samples, Table 3 summarizes the effect on
the lower bound of the sample mean WTP value (using the original, unad-
justed referendum responses) as various types of respondents are removed
from consideration. Several features of this table merit discussion. First,
one is struck by the high variability of the WTP lower-bound estimates as
the various groups are excluded. For V1, for example, the resulting esti-
mates range from 40%[66] to 134% of the "all respondents" value. Second,
the pattern of variations for the V2 values apportioned to marine oil spills
appears to be broadly similar to that observed with the V1 values.[67]

In reviewing other aspects of the table, we will focus attention on those
categories that either (1) remove the highest proportions of responses, or
(2) result in the largest adjustments to the WTP statistic, mostly using the
V1 sample data for illustration. Of particular note in both regards are
the effects of eliminating those respondents who might be judged to have
misconstrued the COS good in some way (row a).[68] In total, about 85% of
the full Version 1 sample indicated "misconstruction" in at least one way
or another. Some of the people eliminated might (for example) be just
expressing a general skepticism about what they are told about government
programs, or be exhibiting acquiescence response bias when presented with
rationales that in reality may not have occurred to them. But the large per-
centage per se will come as little or no surprise to those who have viewed

When each statement appears, please select a number between 1 and 5 to indicate how well
that statement describes the way in which *you personally* thought about your own answers to
those questions. A '5' answer means that the statement 'describes me perfectly,' and a '1' answer
means the statement 'doesn't describe me at all'."

The battery of statements relating to Versions 1 and 2 comprised 29 in total, not all
of which applied to any one respondent. Presented in randomized order, each statement
would appear alone across the top of the screen, with a five-point numerical scale arrayed
horizontally beneath it. Consistent with our strong belief in allowing respondents to indi-
cate uncertainty or discomfort about a question, an "I'm not sure" response option was
also provided.

When using this form of debriefing question in the sensitivity analyses, we based the clas-
sification on the two points at the appropriate end of the scale; that is, scale positions 4 and
5 (or 1 and 2).

[66]   Excluding the atypical row h entry.

[67]   The sizes of the variations shown in Table 3 generally increase when unweighted
data are used instead of the weighted statistics shown in the table. We also note that the
1995 survey's exclusion of Spanish-only households does not appear, on this evidence, to
be material: the V1 Spanish language interviews showed only a slightly lower WTP statistic
than the English language interviews.

[68]   These people reported either (1) that they thought the harm from oil spills would be "a
lot *more*" than presented in the questionnaire; or (2) that they thought that they would have to
pay the special tax "for more than one year;" or (3) that they "were *consciously* thinking" that
the hypothetical program was in some way more beneficial than had been described (either
"other types of animals or plants. . .would also benefit," or that the program infrastructure
"might also be useful if a spill occurred along *other parts* of the California coast," or "if a spill
occurred at an oil drilling well or because of a broken underwater pipeline").

Table 3   Sensitivity of the lower-bound mean WTP estimates to respondent exclusion criteria

Subsample	V1 Single-focus COS		V2 Marine Oil Spills	
	Un-weighted *n*	WTP ($)	Un-weighted *n*	WTP ($)
Original referendum responses for all respondents:	1,213	105.5	1,172	2.47[p]
*Comparable estimates, **excluding**:*				
a. Respondent misconstrues the COS good	179	49.9	410	1.16[p]
b. COS scenario judged not fully credible	441	138.5	348	3.74[p]
c. Survey judged inconsequential	993	113.3	954	2.93[p]
d. Surveys, ballots shouldn't be used in this case	612	99.6	414	2.37[p]
e. "Protest no" votes	948	141.6	778	4.21[p]
f. Costs shouldn't (or didn't) matter	643	42.1	654	1.09[p]
g. "Not sure" referendum responses:	986	130.5	888	3.49[p]
Respondent is not a taxpayer	924	112.0	879	2.86[p]
Changed from *for* to *not-for* on re-ask	1,165	98.7	1,121	2.59[p]
WTP more than 5% of income	1,204	100.2	1,151	2.51[p]
Respondent compares bid level to donations	692	99.7	641	2.02[p]
Survey too long or too complicated	835	106.5	752	2.37[p]
Survey gave too much or too little information for the allocation task			637	2.52[p]
Survey pushed one way or another	1,023	106.8	921	3.07[p]
Survey pushed to vote for	1,085	106.3	993	2.93[p]
Respondent not currently registered to vote	992	109.2	973	2.76[p]
Respondent unlikely to vote	900	89.4	826	2.85[p]
Spanish language interviews	1,033	105.7	1,023	2.69[p]
h. Any one or more of the above criteria	4	12.5[p]	12	3.21[p]

Note: WTP amounts are estimates of the lower bound on the sample mean. Estimates with the superscript [p] are derived from pooled models.

video recordings of "cognitive interviews" or "verbal protocols" used as respondents complete CVM tasks,[69] and it is also notable that for the people *not* admitting to any of the misconstruction indicators, the WTP statistic falls markedly, by over a half of the reference value.

[69]   Such as those filmed in connection with Schkade and Payne (1993).

Eliminating the respondents who in some way do not find the scenario fully credible (row b)[70] also reduces the available sample markedly, by about 64%. Having credibility concerns apparently lowers WTP, because removing such respondents from consideration *increased* the WTP statistic by a little over 31%. Relatedly, removing the roughly 17% of respondents appearing to view the survey as inconsequential (those *dis*agreeing that "my answers to this survey will help state government decision-makers to make better decisions about new environmental programs like the one described in the survey", row c) increased the WTP statistic by almost 7½%.

However, there were other respondents agreeing with statements that, for issues like these, the state should rely more on scientists and other trained experts than on either "surveys (like this one) of the general public" or "how California voters vote on ballot propositions." Row d shows that just under half of the sample, in total, agreed with one or other of those statements, and when those people were excluded the WTP statistic fell by roughly 5½%.

CVM practitioners have long sought to identify "protest *no*" votes, and remove them from consideration. In our case, we classified about 22% of the respondents as giving protest *no* votes (row e),[71] and eliminating those respondents increased the WTP statistic by roughly a third. At the opposite end of the spectrum are respondents – sometimes characterized as "*yea*-sayers" – who will agree to proposed environmental improvements no matter how high the price attached to the proposal (row f).[72] We classified 47% of the V1 sample in this category, and when those respondents were removed, the WTP statistic fell by 60%.

Removing the people giving "not sure" referendum responses (row g, almost 19% of the sample) increases the WTP statistic by about 24%, to $130.5.[73]

---

[70]   These were people who agreed with statements indicating that some aspect of the COS scenario was either "not factually correct" or "just didn't make sense to me," or who were not sure the proposed program would be able to prevent oil spills.
[71]   These respondents voted *no* in the referendum and expressed agreement with statements either that "I would not vote in favor of any tax increase, no matter how worthy the program" or that the oil companies should pay.
[72]   These were respondents agreeing with either (1) "In decisions about preserving and protecting our environment, the costs shouldn't really matter;" or (2) "I voted for the *Central California Coast oil spill prevention program* mostly because we need to do more to care for our environment. I was less interested in the actual details of the program, or its cost to {my household}, than the fact that it would help the environment."
[73]   The similarity of this estimate to the comparable one from the 1995 survey, expressed in 2014 prices ($141.1), is most likely a chance outcome, but it does suggest that the different treatment of DK responses (strongly discouraged in 1995, despite the Blue Ribbon Panel's advocacy, while offered as an acceptable response in 2014) may be a significant element in the differences between the two surveys' numerical results.

In the V1 survey, the proportion of respondents who opined that the survey had "pushed me to vote for" the proposal was about 10.5%. For the California Environmental Improvement Program in V2, the proportion expressing the same sentiment rose to over 15% (p value for the difference = 0.001), not surprising given the additional time spent in describing the COS program and voting on it as a standalone proposition.

Other researchers[74] have observed that the cumulative effect of eliminating all those CVM respondents whose answers, in one respect or another, do not conform with the theoretical notions of what constitutes a "good respondent" can result in very few remaining observations. In our case, of the 1,213 California residents interviewed in Version 1, only four of them were not removed in *any* of our sensitivity tests (row h).

For practically every one of the rows in Table 3, there could be justification for using *that* particular entry as *the* non-use value estimate best representative of the general California population. . .for framing public policy, or for damages estimation, or for whatever purpose. Yet the variation in the implicit aggregate monetary values, when multiplied by the applicable number of households (currently approaching 13 million in California), can be very large.

In coming up with an appropriate estimate, the researcher needs not only to fix on design details for the survey itself – which, as we see in this study, have great potential to influence the results[75] – but also on which responses should be edited out.[76] Since the resultant value estimate can vary greatly with such decisions, those choices merit greater attention and justification than appears common in many CVM study reports.

4. Respondents who were presented the single-focus COS referendum after having done the budget exercise were slightly less favorable to COS than those who did not have the budget exercise.

After completing the composite good allocation and valuation tasks, V2 respondents were asked to vote on an alternative referendum for the COS good alone. This resulted in an estimate for the lower bound on the sample mean WTP of $107.9. One hypothesis shaping our study design was that the budgeting exercise might serve effectively as "training," helping respondents to better understand the part-whole relationship

---

[74] See, for example, Myers et al. (Chapter 9, this volume).

[75] Consider, for instance, the effects of using a long or short version of the scenario, including or excluding an explicit "not sure" option, or opting for one specific type of wording addition.

[76] Because, say, the respondent is a non-taxpayer, a non-voter, a tax-protester, a yea-sayer, or does not view the survey as consequential.

between COS and the portmanteau program, and the relative scale of the two initiatives. We plotted the precision bands for the V1 sample response curve and for the V2 sample's standalone COS question, showing that the *not-for* vote proportion had increased in V2 at all design points. While the confidence intervals were not greatly separated, and the implied reduction in the WTP was not large, the budget exercise does appear to have had a modest effect.

The limited impact of the budget exercise on the WTP for COS may come from the fact that there are two different substitutions that under-lie the two types of questions. The response to the standalone version should depend on the existence of close substitutes for COS that are already available, but the budget exercise deals with substitution among public goods that are not presently available. Recent experience with the latter seems not to have had much effect on the awareness of the former.

5.   A sizeable proportion of respondents experiences cognition difficulties with part-whole relationships.

As well as the evidence provided by the immediately preceding con-clusion, this finding reflects two additional empirical observations revealed by our study when (after the composite good allocation and valuation tasks) Version 2 respondents were given the COS scenario as an example of just one component of the portmanteau good. First, when allowed to reallocate budget shares, some respondents exhibited behaviors suggesting little conception of the part-whole relationship between the example component and the portmanteau good. Second, the valuation estimate for the COS standalone good also points to a similar problem.

Some V2 respondents changed their allocations markedly after being presented the COS scenario. For the original allocation to marine oil spills, 98.6% of the implied amounts were $50 or less, and 99.9% were $100 or less. After reallocation, these proportions changed to 88.7% and 95.4% respectively, but 2.9% of the sample now had implied amounts of over $200. At the extreme, three respondents increased their marine oil spills allocation from an original zero to *100%* of their bid levels, and a fourth adjusted from zero to 95%.

Arraying respondents by the percentage and dollar values of their adjustments shows, in the upper tail of the distribution, a spectrum of respondents who clearly are responding to the stimulus of the most recently presented COS scenario without placing it in any perspective of the much larger scope of the portmanteau program on which they had previously voted. For these outliers, the experiment has effectively ceased

to be a disaggregation survey.[77] Retaining the relative advantage of the disaggregation survey in reducing importance bias obviously requires some editing of allocation adjustments allowed after introducing a single-focus scenario, but where to draw the line for that editing is an arbitrary decision.

Viewed in the context of the various "wording additions" explored in subsamples of Version 1 (to be discussed subsequently), the systematic delineation of possible other environmental issues that might compete for a share of the respondent's total WTP appears to be comparably effective in reducing valuations to the "cheap talk," "environmental contributions," or "dissonance minimization" approaches.

But from a different viewpoint, one might ask why the value obtained from the standalone COS exercise is as *large* as this.[78] Part of the standalone COS valuation might reasonably be ascribed to "anchoring effects." There is considerable evidence in the non-market valuation literature that responses to later valuation exercises can often be strongly influenced by the prices that respondents see in their *first* (or *any* prior) exercise.[79] For that reason, many studies examine the findings from an initial valuation exercise separately from those derived from any subsequent exercises, and prefer to use split samples rather than multiple valuations from the same respondents. In our case, 80% of standalone COS respondents had been shown a higher bid level for their earlier valuation of the composite good, and for those people the ratio between the composite good price and the standalone COS price, while highly variable, averaged about 38. For a respondent paying more attention to the price than to the specifics of the environmental good under offer, standalone COS would frequently appear to be a comparative bargain.[80]

When respondents express a WTP value for the *first* environmental good with which they are presented, they may be using a large portion of their

[77]　In one sense, this portion of the sample can be regarded as operating in a zone between a single-focus survey – to which they are being drawn perhaps because of the recency, specificity, substance, or some other feature of the COS good – and a composite good survey.

[78]　After all, the "COS alone" referendum question came *immediately* after respondents were shown (in dollar terms) the size of their allocation to marine oil spills under the California Environmental Improvement Program. Across the full range of design points for the portmanteau program, the monetary allocation to marine oil spills (before any respondent adjustments of the allocations) ranged from zero (for over 23% of the respondents) to about $56 at the 99th percentile, with a mean of $5.3 and a median of $1.3.

[79]　Burrows et al. (Chapter 5, this volume) provide a list of studies drawing this conclusion.

[80]　Relatedly, the independent randomization of design points allocated for the two very different program scopes in Version 2 potentially creates a credibility issue for the minority of respondents offered a bid level for standalone COS that is relatively high compared with the bid level offered earlier for the portmanteau program. For each respondent we computed a variable ϕ as the ratio between the standalone COS design point and that for the composite good. But as Table 2 shows, the effects of removing those respondents with the highest values for ϕ do not indicate that those respondents were much affected by that design artifact.

"budget for good deeds," so that this value is essentially a warm glow value. This explanation is also consistent with our finding that large proportions of the respondents do not find the survey consequential or fully credible,[81] so they are spending essentially "free" virtual dollars to achieve their warm glow.

6.   Within-questionnaire "wording additions" intended to enhance budget awareness had a relatively small effect on WTP estimates.

Employing various within-questionnaire "wording additions" touching on budget awareness, intended to reduce some of the budget-related hypothetical bias, produced results that are broadly in line with the literature reports for such techniques. Version 1 of our survey tested three different types of wording additions within the interview. Each treatment (including the "no wording additions" baseline) was allocated randomly to a quarter of the total V1 sample.

   References to household budget matters in the 1995 questionnaire were quite sparse, limited to some text on one visual aid that was also read out to the respondent,[82] and all three of the 2014 subsamples receiving wording additions also retained this 1995 language. For two of the three subsamples, the additional text came immediately after the 1995 summarization of reasons for voting for and against the proposal. The first variant of additional text was a "cheap talk" entreaty in the spirit of Cummings and Taylor (1999), but shortened somewhat in a manner used for a recent, large, online choice experiment study sponsored by NOAA.[83] This variant also emphasized that respondents might have non-use values for *other* environmental goods as well as for the subject program.[84]

---

[81]   Across the full V1/V2 sample, about 29% agreed with "In decisions about preserving and protecting our environment, the costs shouldn't really matter," and 39% agreed with "When I thought about whether to vote for or against the proposed program. . .I was *consciously* thinking about the amounts I typically give when asked to *donate* to a good cause that I believe in." Roughly 57% of the respondents voting in favor of the program agreed with "I voted for the {program} mostly because we need to do more to care for our environment. I was less interested in the actual details of the program, or its cost to {my household}, than the fact that it would help the environment."

[82]   The visual aid card summarized two reasons why the respondent might want to vote for the proposal and three reasons to vote against it. The second and third reasons for "voting against" were expressed in the (slightly longer) oral script as "Your household might prefer to spend the money to solve *other* social or environmental problems *instead*" and "The program might *cost more* than your household wants to spend for this."

[83]   See Wallmo and Lew (2011).

[84]   "As you consider how you would vote on the Central Coast oil spill prevention program, please take into account the following considerations. The extra one-time tax funds that {your own household} and all other California households would have to pay would help to prevent oil spills off the Central Coast over the next ten years, and to prevent the harm if spills do occur.

The second tested wording addition focused on the respondent's perceived disposable income and on the current budget allocation to environmental concerns.[85] The fourth subsample of Version 1 changed the response options for the valuation question itself. In the other three variants, the response options were just "FOR," "AGAINST," and "I'm not sure." For the "dissonance minimization" variant, following Morrison and Brown (2009), three additional "AGAINST + excuse" response options were added.[86]

Table 4 summarizes the estimated WTP statistics for each of the wording

For survey questions like these, studies have shown that many people *say* they are willing to pay more for new environmental protection programs than they actually *would pay* out of their pockets. We believe this happens because, when answering these types of questions, people do not really consider how big an impact an extra cost actually has to their {household's} budget. It is easy to be generous when you do not really need to open your wallet.

To avoid this, as you consider each question please suppose that {your household} will actually have to pay the cost indicated in the question, out of your {household's} budget.

There are *other* environmental improvements that you might like to see made, both here in California and elsewhere in the United States. There are *other* areas of public policy that you might also like to influence, both here in California and nationally. Those other programs might also imply increases in the amount of State or federal income tax that {your household} would have to pay, either as a one-time tax or annually. You may or may not prefer that the State or the federal government spend your money on other things, if those other things have a higher priority for you than the proposed Central Coast oil spill prevention program.

The proposed one-time tax increase might also require {your household} to spend less on other items that you need or want, or to put less money into your savings."

[85] Following Li et al. (2005), this "environmental contributions" variant read: "But before we ask you how you would vote on this ballot proposition, please think about your {household's} average monthly income and expenses. After you have paid all the necessary bills for such things as housing, transportation, groceries, insurance, debt, and taxes, what percent of your income is left over for optional uses on things like recreation, savings, and giving for charity and other causes?

Now please think about the portion of your {household's} total monthly income available for "optional uses" (like recreation, savings, and giving for charity and other causes). On average, what percent *of that amount* do you use for contributions to environmental causes, such as donations for specific programs or contributions and memberships to environmental advocacy groups?"

[86] "If a California-wide ballot proposition about this proposal were on an election being held today, and the total cost to {your household} for the program would be $X, which one of these statements best describes your own reaction to the ballot proposition? [*Select one only.*]

I would vote FOR the proposition.

I support the goal of the Central Coast oil spill prevention program, but I'm not prepared to pay that much and so I would vote AGAINST.

I support the goal of the Central Coast oil spill prevention program, but I cannot afford to pay that much and so I would vote AGAINST.

I support the goal of the Central Coast oil spill prevention program, but I prefer that my money be spent on other priorities and so I would vote AGAINST.

I support the goal of the Central Coast oil spill prevention program, but I would vote AGAINST for the following reason [*Please enter:*_____].

I would vote AGAINST the proposition.

I'm not sure."

*Table 4   Version 1 ABERS estimates of the lower bound on the sample
mean WTP*

Subsample	Unweighted n	WTP ($)	St. Err. ($)
All Version 1 respondents:	1,213	105.5	7.2
All Version 1, by wording addition variant:			
No wording addition	306	109.8[p]	na
"Cheap talk" wording added	307	118.8	14.0
"Environmental contributions" wording added	300	103.4[p]	na
"Dissonance minimization" wording added	300	85.1[p]	na
All Version 1, by scenario length:			
Long scenario version (as 1995)	610	93.0	9.2
Short scenario version (as used for V2)	603	118.8	10.0

Note: The estimates in this table are based on initial responses, equivalent to the measure
B1 in the 1995 study. Estimates with the superscript ^{p} are derived from pooled models, and
no standard errors were computed in those cases.

addition treatments.[87] The situation was confounded somewhat by uncer-
tainty around our baseline ("no treatment") measure, caused by non-
monotonicity that required the pooling of the highest two design points
for that subsample.

For respondents who saw the "cheap talk" variant before the referendum
question, the WTP statistic is surprisingly *higher*. While an occasional few
other studies have observed a similar effect,[88] the preponderant result in the
literature is that "cheap talk" scripts *do* tend to lower the WTP estimate.[89]

In addressing non-monotonicity when comparing WTP estimates, we
found it valuable to plot the one-standard error confidence bands for
not-for vote percentages by design point, because with the added assump-
tion that the true curve should increase monotonically, neighboring points
add statistical support (over and above the precision provided by just each
point's sample observations considered independently). This approach
would appear to be more insightful than the "level step" diagrams typically
generated for singly bounded dichotomous choice responses.

For example, Figure 3 plots the precision bands for our cheap talk and
baseline cases. It also shows for each design point the estimate when the

---

[87]   Each of the table rows comprises approximately equal numbers of respondents given
the long and short COS scenario descriptions.
[88]   See, for example, Aadland and Caplan (2006).
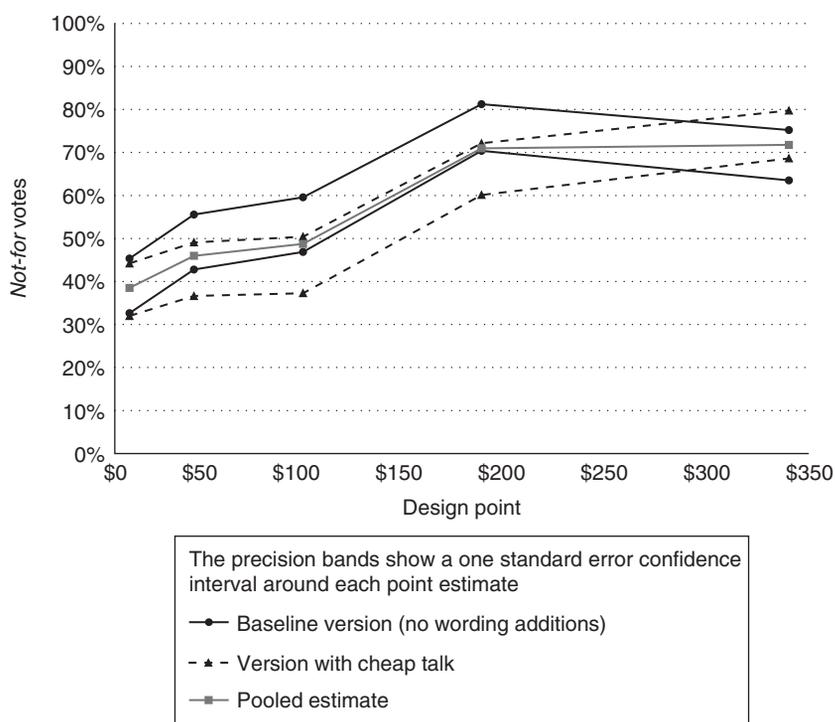[89]   See Loomis (2014) at Table 1, and Morrison and Brown (2009).

Figure 3   *Confidence bands for referendum responses with and without cheap talk, by design point*

baseline and cheap talk subsamples are pooled, and these always lie within the one standard error precision bands for both of the two subsamples. On this evidence, the baseline and cheap talk responses appear statistically indistinguishable.

The response curves for both the "environmental contributions" and "dissonance minimization" methods were (barely) non-monotonic for the weighted sample, but not for the unweighted sample (which yielded WTP statistic estimates of $102.8 and $92.4 respectively). Perhaps because of the issue with the non-monotonic baseline, the impact of our tested wording additions appeared somewhat lower than those reported by the original developers of the two techniques.[90] Our general ranking of the tested

[90]   For "environmental contributions," Li et al. (2005) observed a drop in their estimate of *median* WTP of between 50% and 60% from their base, untreated case. For "dissonance minimization" Morrison and Brown (2009) measured a 32% reduction in their Turnbull

techniques did appear to support the literature experience – "dissonance minimization" produced the most marked reduction in the WTP statistic, followed by "environmental contributions" and "cheap talk" – but examination of the confidence bands for our response curves makes it hard to conclude that these differences are statistically significant.

A PLAUSIBLE RATIONALIZATION FOR SEVERAL OF THESE FINDINGS

The evidence from our study that many respondents are more focused on the type of message that their answers convey than on balancing a household budget call into question whether cognitive disjunctions threaten the validity of trying to improve the financial consequentiality of CVM responses.

As a thought experiment, consider the following simple hypothetical. The survey sample comprises two segments, one of which has respondents who do their very best to consider carefully all of the information they are given, and try their hardest to answer the questions as accurately and realistically as possible, just as the CVM theoreticians would want them to do. The second group – the polar opposite – has people with a much simpler set of decision rules. Is the proffered public good (without getting into all of that superfluous detail) one of which I broadly *approve* or *disapprove*? Is the price at which I'm being offered that good acceptable to me, as a once-only payment with no longer-term commitment (just like a one-off charitable donation)? If I approve, and the price is OK, then I'll vote *for*. Task completed; move on to the next question when I will apply the same approach again a priori, regardless of any part-whole or other association between this public good and any previous one.

Such a hypothetical would explain not only our V2 "standalone COS" result but also the disconnect that a proportion of V2 respondents showed when allowed to adjust their portmanteau program allocations after being shown the COS program scenario. It would help explain the serious scope failure when the V1 responses are compared with those for the much more expansive V2 cornucopia of environmental initiatives. Indeed, the hypothetical also would rationalize many of the other hypothetical bias phenomena observed with CVM surveys, such as "warm glow altruism" and the pervasive and persistent similarities between WTP findings and average charitable donations.

WTP estimate from including the same number of *no* categories. The resulting measure was slightly below a revealed preference estimate of WTP, but not by a significant amount. See also Loomis (2014), at pp. 38–9.

If something like that hypothetical (albeit in reality, doubtless a more subtle and complicated version of it) is, in fact, at work here, one key problem for the survey designer and the analyst is how to identify the different sub-samples, and how best to interpret their various responses once having identified them.[91] Some clues can come from a comprehensive debriefing battery (as was used with this survey), and thorough sensitivity testing.

REFERENCES

Aadland, D. and A.J. Caplan (2006), "Cheap talk reconsidered: New evidence from CVM," *Journal of Economic Behavior and Organization*, **60**(4), 562–78.

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman (1993), "Report of the NOAA Panel on contingent valuation," *Federal Register*, **58**, 4,601–14.

Ayer, M., H.D. Brunk, G.M. Ewing, W.T. Reid, and E. Silverman (1955), "An empirical distribution function for sampling with incomplete information," *Annals of Mathematical Statistics*, **26**(4), 641–7.

Bateman, I.J., M. Cole, P. Cooper, S. Georgiou, D. Hadley, and G.L. Poe (2003), "On visible choice sets and scope sensitivity," *Journal of Environmental Economics and Management*, **47**(1), 71–93.

Bishop, R.C., D.J. Chapman, B.J. Kanninen, J.A. Krosnick, B. Leeworthy, and N.F. Meade (2011), *Total Economic Value for Protecting and Restoring Hawaiian Coral Reef Ecosystems: Final Report*, Silver Spring, MD: NOAA Office of National Marine Sanctuaries.

Boxall, P.C., W.L. Adamowicz, M. Olar, G.E. West, and G. Cantin (2012), "Analysis of the economic benefits associated with the recovery of threatened marine mammal species in the Canadian St. Lawrence Estuary," *Marine Policy*, **36**(1), 189–97.

Brown, T.C., S.C. Barro, M.J. Manfredo, and G.L. Peterson (1995), "Does better information about the good avoid the embedding effect?," *Journal of Environmental Management*, **44**(1), 1–10.

Carson, R.T., M.B. Conaway, W.M. Hanemann, J.A. Krosnick, R.C. Mitchell, and S. Presser (2004), *Valuing Oil Spill Prevention: A Case Study of California's Central Coast*, Dordrecht: Kluwer Academic Publishers.

Cummings, R.G. and L.O. Taylor (1999), "Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method," *American Economic Review*, **89**(3), 649–65.

Hoehn, J.P. and A. Randall (1989), "Too many proposals pass the benefit cost test," *American Economic Review*, **79**(3), 544–51.

Kahneman, D. and J.L. Knetsch (1992), "Valuing public goods – the purchase of moral satisfaction," *Journal of Environmental Economics and Management*, **22**(1), 57–70.

---

[91] Leamer and Lustig (Chapter 8, this volume), who use latent class models to explore the heuristic decision rules that respondents may adopt in addressing non-market valuation questions, address a similar hypothesis.

Kaplan, E.L. and P. Meier (1958), "Nonparametric observation from incomplete observations," *Journal of the American Statistical Association*, **53**(282), 457–81.

Kemp, M.A. and C. Maxwell (1993), "Exploring a budget context for contingent valuation estimates," in J.A. Hausman (ed.), *Contingent Valuation: A Critical Assessment*, Amsterdam: North-Holland, pp. 217–65.

Kotchen, M. and S. Reiling (1999), "Do reminders of substitutes and budget constraints influence contingent valuation estimates? Another comment," *Land Economics*, **75**(3), 478–82.

Krosnick, J.A. and S. Presser (2010), "Question and questionnaire design," in P.V. Marsden and J.D. Wright (eds) (2010), *Handbook of Survey Research*, 2nd edition, Bingley, UK: Emerald Group Publishing, pp. 263–313.

Krosnick, J.A., A.L. Holbrook, M.K. Berent, R.T. Carson, W.M. Hanemann, and R.J. Kopp et al. (2002), "The impact of 'no opinion' response options on data quality: Non-attitude reduction or an invitation to satisfice?," *Public Opinion Quarterly*, **66**(3), 371–403.

Li, H., R.P. Berrens, A.K. Bohara, H.C. Jenkins-Smith, C.L. Silva, and D.L Weimer (2005), "Testing for budget constraint effects in a national advisory referendum survey on the Kyoto Protocol," *Journal of Agricultural and Resource Economics*, **30**(2), 350–66.

Loomis, J.B. (2014), "Strategies for overcoming hypothetical bias in stated preference surveys," *Journal of Agricultural and Resource Economics*, **39**(1), 34–46.

Magnussen, K. (1992), "Valuation of reduced water pollution using the contingent valuation method – testing for mental accounts and amenity misspecification," in S. Navrud (ed.), *Pricing the European Environment*, Oxford: Oxford University Press, pp. 195–220.

Morrison, M. and T.C. Brown (2009), "Testing the effectiveness of certainty scales, cheap-talk, and dissonance minimization in reducing hypothetical bias in contingent valuation studies," *Environmental and Resource Economics*, **44**(3), 307–26.

Parsons, G. and K. Myers (2016), "Fat tails and truncated bids in contingent valuation: An application to an endangered shorebird species," *Ecological Economics*, **129**(C), 210–19.

Schkade, D.A. and J.W. Payne (1993), "Where do the numbers come from? How people respond to contingent valuation questions," in J.A. Hausman (ed.) (1993), *Contingent Valuation: A Critical Assessment*, Amsterdam: North-Holland, pp. 271–93.

Schläpfer, F. and I. Brauer (2007), "Theoretical incentive properties of contingent valuation questions: Do they matter in the field?," *Ecological Economics*, **62**, 451–60.

Smith, A.E., M.A. Kemp, T.H. Savage, and C.L. Taylor (2005), "Methods and results from a new survey of values for eastern regional haze improvements," *Journal of the Air and Waste Management Association*, **55**(11), 1767–79.

Thaler, R.H. (1985), "Mental accounting and consumer choice," *Marketing Science*, **4**(3), 199–214.

Turnbull, B.W. (1974), "Nonparametric estimation of a survivorship function with doubly censored data," *Journal of the American Statistical Association*, **69**(345), 169–73.

Turnbull, B.W. (1976), "The empirical distribution function with arbitrarily

grouped, censored and truncated data," *Journal of the Royal Statistical Society B*, **38**, 290–95.

Wallmo, K. and D.K. Lew (2011), "Valuing improvements to threatened and endangered marine species: An application of stated preference choice experiments," *Journal of Environmental Management*, **92**(7), 1793–801.

Whitehead, J.C. and T.L Cherry (2007), "Willingness to pay for a green energy program: A comparison of ex-ante and ex-post hypothetical bias mitigation approaches," *Resource and Energy Economics*, **29**(4), 247–61.