



THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

Alternative Value Elicitation Formats in Contingent Valuation: A New Hope

Christian A. Vossler and J. Scott Holladay

August, 2016

WORKING PAPER #2016-02

WORKING PAPER SERIES
DEPARTMENT OF ECONOMICS
HASLAM COLLEGE OF BUSINESS
<http://econ.bus.utk.edu>

Alternative Value Elicitation Formats in Contingent Valuation: A New Hope

Christian A. Vossler and J. Scott Holladay

Department of Economics and Howard H. Baker Jr. Center for Public Policy,

University of Tennessee, Knoxville, TN

*Correspondence should be directed to Christian Vossler, Department of Economics, 523 Stokely Management Center, The University of Tennessee, Knoxville, TN 37996-0550. E-mail: cvossler@utk.edu. Telephone: 865-974-1699. Fax: 865-974-4601. This work was funded in part through a grant from the Alfred P. Sloan Foundation.

Alternative Value Elicitation Formats in Contingent Valuation: A New Hope

Abstract: The single binary choice (SBC), referendum format has long been the recommended approach for eliciting values in stated preference surveys, based on respondent familiarity and incentive compatibility arguments. Nevertheless, researchers and practitioners commonly use alternative elicitation formats, and defend their design choices on the basis of efficiency or other criterion. While we are agnostic as to what format is best, in this paper we seek to advance the idea that incentive compatible elicitation using alternative formats is possible, and that designing surveys through the lens of theory can be beneficial. We highlight this paradigm by identifying a set of conditions under which two continuous response formats – purely open-ended (OE) questions and payment cards (PCs) – are incentive compatible. We then implement theory-informed value elicitations in the context of a flood control policy for New York City. We fail to reject convergent validity when comparing the theory-driven OE format with SBC, but reject convergent validity between the theory-driven PC and SBC formats. As an informative counterfactual, we find that a “standard” OE elicitation congruent with prior work leads to significantly lower values and a lower proportion of respondents who view the elicitation as consequential.

Keywords: contingent valuation, mechanism design, field experiment, flood protection

JEL classification: H41, Q51, C93

1. Introduction

Though stated preference surveys remain a standard approach for estimating values for public goods in the context of government cost-benefit analysis and in litigation over damages to natural resources, no consensus has been reached over a large number of important design issues.¹ Perhaps the most central issue is the choice of question formats for eliciting Hicksian welfare measures. Dating back to at least the report of the US National Oceanic and Atmospheric Administration Blue Ribbon Panel (Arrow et al., 1993), a single binary choice (SBC) question framed as an advisory referendum has been viewed as the industry standard. Nevertheless, many alternatives to SBC are used in practice.² Researchers routinely advocate for alternative formats on the grounds that they conceptually gather more preference information from each respondent – reducing the sample size needed – by asking about multiple goods in the same survey and/or by eliciting more precise information for a given good. At the same time, alternative value questions tend to be more complex, unfamiliar, and are theorized to give rise to strategic, untruthful responses. To some extent the choice among value elicitation formats can be characterized as a tradeoff between bias and efficiency.

In this paper we adopt the philosophy that a set of assumptions *can* be identified under which alternate elicitation formats are incentive compatible. Such theory development has several advantages. First, theory allows one to make informed comparisons across formats, leading to arguments based on the strength of assumptions and thus dampening the standard defense of an alternative format on the basis of convenience or budget. Second, theory can lead to important survey refinements, such as framing and information provision, to engender respondent beliefs that are consistent with theory assumptions. As evidenced by work on SBC

¹ See Kling, Phaneuf and Zhao (2012) for a thoughtful discussion of accumulated evidence on contingent valuation.

² We refer to any format other than a SBC elicitation as an “alternative” format.

questions, enhancing beliefs over policy consequences has been shown to increase construct validity (Herriges et al., 2010). Third, the information content of surveys as relevant for mechanism design varies widely across studies. Economists are predisposed to use theory to guide empirics, and therefore theory development is likely to cultivate consistency in standards of practice. This has many potential benefits, such as eliminating confounds in the context of benefit transfers or meta-analyses. Last, as evidenced by work on SBC questions, conditioning analyses on theoretical assumptions has been pivotal when testing external validity (Vossler and Watson, 2013).

We illustrate this paradigm using as examples two continuous response formats: purely open-ended (OE) questions and payment cards (PC). Relative to SBC, such questions, respectively, elicit a point value or define a narrow interval containing the agent's value. In this study we first identify theoretical assumptions for incentive compatibility for OE and PC questions. In the context of a flood control policy for New York City (NYC), we develop survey elicitation mechanisms informed by the theory, and test for convergent validity through comparisons with parallel SBC elicitations. As an important counterfactual, we include in the experimental design a "standard" OE elicitation that better resembles current practice.

Continuous response formats face two incentive challenges related to SBC. First, as in market settings, e.g. when purchasing a car, a participant may believe that her response can influence the price paid for the good thus incentivizing her to under-reveal demand. The second challenge stems from the lack of an implementation rule. Although an explicit rule is largely absent in SBC applications, it is natural for respondents in this familiar setting to believe that (if anything) a response in favor will increase the chance of implementation. In contrast, implementation rules that would seem natural in the continuous response case give rise to

untruthful responses. For example, as discussed by Carson and Groves (2007), respondents may believe that the chance a policy is implemented increases with the sum (or average) of stated valuations. Thus, if the respondent believes that the cost to her is fixed, but her true valuation is less than the expected cost, the optimal response to an OE question is to state zero. To overcome these issues in theory, we assume that the respondent believes her stated valuation will be interpreted as a yes or no vote to the proposed policy at the actual cost, which is unknown to the respondent when taking the survey and determined exogenously from responses at a later time. The basic logic of using an uncertain and exogenous price stems from the Becker-DeGroot-Marschak (BDM) mechanism (Becker, DeGroot and Marschak, 1964) and the random price voting mechanism (RPVM) of Messer et al. (2010), which elicit a continuous response (bid) for private and public goods, respectively.

The theory, in turn, provides new insight for survey design. Indeed, standard OE and PC implementations are very unlikely to adhere to theory stipulations for incentive compatibility, as highlighted above. In the context of a valuing a proposed flood control policy for NYC, we develop theory-driven OE and PC formats, which emphasize cost uncertainty and suggest the possible interpretation of responses as yes or vote votes. Our designs further include a coercive payment vehicle and frame the elicitation as an advisory referendum. Some prior studies utilizing PC or OE elicitation have incorporated one or both features, but they are not systematically included by practitioners or in the academic literature.

As primary evidence on the theory-driven mechanisms, we implement two complementary field survey experiments. In the first, we use a split-sample approach to test for convergent validity between our theory-informed PC and a parallel SBC mechanism. We find that, consistent with prior comparisons involving PCs, SBC values are statistically higher (see

Champ and Bishop, 2006). In a second experiment, we compare a theory-driven OE mechanism with SBC. As a third treatment in this experiment, designed to provide an indication of whether the modifications we made matter empirically, we include a more standard OE elicitation. Interestingly, we find that SBC values are statistically higher than with the standard OE question, but statistical equivalence between SBC and our theory-informed OE mechanism. The frequency of zeroes is much lower in the theory-driven elicitation, which provides suggestive evidence that the zero-response strategies discussed by Carson and Groves (2007) may have been dampened by survey refinement. The above results are robust to the inclusion/exclusion of control variables, including whether respondents indicated beliefs consistent with both payment and policy consequentiality, which coincide with incentive compatibility assumptions.

Taking an appropriately designed SBC elicitation as a yardstick from which to measure alternative elicitation approaches, our results for the OE elicitation support the notion that truthful demand revelation using OE questions pivots on whether theory-based enhancements are implemented. The results for the PC suggest that convergent validity is rejected. Although additional work is needed to decipher the drivers of the result, one possibility is that the inclusion of the set of possible cost amounts included in the PC influences how individuals form values for the proposed policy.

The valuation of flood protection measures is important in its own right. Climate change is predicted to alter the frequency and severity of flood events. While equity issues surrounding paying for climate change adaptation have been the subject of some attention in the literature (Bichard and Kazmierczak, 2012) the willingness-to-pay (WTP) for adaptation remains unclear. We take advantage of detailed flood maps and parcel-level data to identify households within

and just beyond the 100-year flood plain to evaluate how WTP varies with exposure to flood risk.

Our results suggest that the WTP for flood control systems varies markedly with risk. Households just outside the 100-year flood plain are willing to pay significantly less (half as much) to install a flood control system in the city. We further find that factors correlated with actual or perceived attitudes towards risk (e.g., whether the household has flood insurance) influence WTP in expected ways. These results complement those found in a related literature that focuses on estimating the WTP for flood insurance (see Botzen and van den Bergh, 2012a, 2012b).

2. Theoretical framework

In this section we discuss the incentives facing respondents to an OE or PC elicitation, and identify a set of sufficient conditions under which the elicitations are incentive compatible. Similar to prior mechanism design work in this area, embedded in these conditions are beliefs assumed to be held by respondents in terms of how responses will be interpreted, aggregated and used in the context of public decision making. While we can speculate on the reasonableness of these beliefs, there are of course challenges to identifying whether they hold in practice or whether they are important empirical drivers of stated preferences.

In our analysis we build upon the conditions proposed by Carson and Groves (2007) for an SBC elicitation, and later formalized and expanded upon by Vossler and Evans (2009), Vossler, Doyon and Rondeau (2012) and Carson, Groves and List (2014). Proceeding in this

fashion allows us to highlight differences across formats. These conditions, as presented in Vossler, Doyon and Rondeau (2012), are that:³

- (i) the participants care about the outcome;
- (ii) the authority can enforce payments by voters;
- (iii) the elicitation involves a yes or no vote on a single project; and
- (iv) the probability that the proposed project is implemented is weakly monotonically increasing with the proportion of yes votes.

Briefly, condition (i) rules out indifference, which presumably would be an issue for predicting responses to any elicitation mechanism. Condition (ii) suggests that respondents consider both the potential benefits and costs of the policy when formulating votes. This assumption is clearly violated for voluntary contributions mechanisms. We refer to an elicitation that meets this condition as maintaining “payment consequentiality”. Condition (iii) restricts the survey elicitation to a SBC question, and (iv) requires the respondent to have a positive subjective probability that her vote will be pivotal to the outcome. In other words, the voter needs to envision that at least one possible scenario exists where her vote will matter, considering the possible distributions of votes from others and possible ways in which votes translate into a policy decision. We will refer to an elicitation that meets this condition as maintaining “policy consequentiality”.⁴

The BDM mechanism and RPVM are continuous response formats that are incentive compatible under expected utility. For both mechanisms, the possible strategic link between one’s bid and the cost of obtaining the good is broken by the decision rule. For the BDM, the

³ For our purpose, the use of the term “voter” here refers to a respondent who participates in the advisory survey referendum.

⁴ The literature often refers to a survey that meets condition (iv) simply as “consequential”.

cost of the good is drawn from an exogenous distribution, and the bid is converted into a binary choice of buy or not buy at the random cost. The decision rule is similar for the RPVM, with the exception that bids from a group of voters are converted into yes and no votes, and a majority-vote rule determines provision. We apply the same logic to continuous response formats.

Let c denote the cost of providing the public good, as perceived by the voter. In the case the voter is not given precise information on cost, to her c is a random variable with density function $p(c)$ with positive density over the interval $[c^{min}, c^{max}]$.⁵ If the elicitation is policy consequential, then the voter's stated valuation s can influence the probability that a policy is implemented. Let $q(s, \mathbf{s}_{-i}; G)$ denote the probability that (some) policy is implemented, where $\mathbf{s}_{-i}$ is a vector of stated valuations from other voters and G represents nuances of the policy process, including policymaker preferences and decision rules.

A voter may perceive her response to influence the particular cost paid upon implementation, in particular that c is increasing in s . Such beliefs would violate the “one-to-one correspondence” condition of Vossler, Doyon and Rondeau (2012). That condition implies that the “project” described and voted on in the survey corresponds to exactly one “policy”, defined identically to the “project”, that is possibly considered by the authorities. As the authors discuss, this condition rules out that the SBC response has an influence over whether some alternative policy is implemented. Otherwise, for instance, a voter with positive expected net benefits for the project may nevertheless vote no if she believes this will increase the chance that an alternative policy with higher net benefits is implemented. For an OE or PC elicitation, without clear information on how cost is determined, it is reasonable that voters may perceive there to be a set

⁵ It is uncommon for studies that use an OE or PC format to provide information on cost. Even with an explicit cost, as in a standard SBC elicitation with a coercive payment mechanism, respondents may nevertheless have uncertainty over the actual cost they may pay.

of policies under possible consideration, which correspond to the project described in the advisory referendum but that vary in terms of their cost (and cost to others).

To address this, and focusing first on an OE elicitation, we propose the following condition:

- (iii') the open-ended response s is interpreted as a vote on a single project, consisting of non-cost attributes as described in the survey and a cost c determined exogenously from s . That $s \geq c$ ($s < c$) signals a yes (no) vote.

Condition (iii') eliminates s as a determinant of c , and from an ex ante perspective maps the OE response into a continuum of binary choices on possible projects uniquely defined by their cost. Ex post, the OE response is a single yes or no vote at a specific cost. Condition (iii') is unlikely to hold in standard OE applications where little to no information is given on how the cost (distribution) would be determined. However, beliefs consonant with (iii') may arise with information provision, in particular that cost is not known with certainty, and that cost would be the actual cost of implementing the project and not based on stated valuations. Relaying uncertainty over cost further provides a rationale for eliciting a continuous response as a means of having available relevant preference information in the event uncertainty is later resolved and the policy is considered seriously by authorities. In turn, such information provision has the potential to prevent the onset of undesirable beliefs on the part of OE respondents.

In addition to (iii'), we rely on the expected utility framework. Carson, Groves and List (2014) relax the expected utility assumption and show that a SBC is incentive compatible under mixture monotonicity. This highlights another theoretical advantage of an SBC question. We now provide a proof that an OE question (can be) incentive compatible.

PROPOSITION 1: An OE question is incentive compatible if conditions (i), (ii), (iii') and (iv) hold.

PROOF:

Let the true valuation of an individual be denoted by v . With OE responses interpreted as yes and no votes for any given cost, we can write $q(s, \mathbf{s}_{-i}; G)$ as $q(R(c); G)$, where $R(c)$ is the total number of yes votes from the sample of survey respondents (i.e., the number of voters with $s \geq c$). For an individual, the total number of votes can be written as $R_{-i}(c) + 1$ if she votes yes and $R_{-i}(c)$ for a no vote. The expected utility of the voter is then:

$$(1) \quad \int_{c^{min}}^s [q(R_{-i}(c) + 1; G)U(v - c) + (1 - q(R_{-i}(c) + 1; G))U(0)]p(c)dc + \\ \int_s^{c^{max}} [q(R_{-i}(c); G)U(v - c) + (1 - q(R_{-i}(c); G))U(0)]p(c)dc.$$

The first term is the expected utility associated with cost realizations for which $s \geq c$, in which case the stated valuation indicates a yes vote and the overall number of yes votes is $R_{-i}(c) + 1$. The second term is the expected utility for cost realizations where the OE response is construed as a no vote. Taking the first-order condition with respect to s , and rearranging, yields:

$$(2) \quad [q(R_{-i}(s) + 1; G) - q(R_{-i}(s); G)][U(v - s) - U(0)]p(s) = 0.$$

That the first bracketed term is positive follows from condition (iv). For an interior solution on the interval $[c^{min}, c^{max}]$, in which case $p(s) > 0$, the voter maximizes expected utility with $s = v$. In other words, the mechanism is incentive compatible, and truthful preference revelation is a weakly dominant strategy.

An interior solution requires that the voter perceives there to be a chance the actual cost of the policy is equal to her valuation. This may not hold for certain projects if the voter has a very low or very high valuation and/or well-defined beliefs over the cost distribution. If $v > c^{max}$, in which case $p(v) = 0$, the voter is indifferent between stating any value in the range $[c^{max}, v]$ as they all yield the same expected utility. Otherwise, if $v < c^{min}$, the voter is indifferent between stating any value in the range $[v, c^{min}]$.

It turns out that incentive compatible elicitation involving PCs requires stronger assumptions, although the elicitation extracts less precise information on valuations. In practice, a PC elicitation involves presenting a set of possible payment amounts and asking the respondent to either indicate whether they would pay each amount separately or to circle the highest amount they would be willing to pay. For either case, let s indicate the highest amount selected on the PC. If $s = c_h$ is the highest amount indicated, standard econometric approaches assume that $c_h < v < c_{h+1}$, where c_{h+1} is the next highest amount.

If the first part of (iii') holds, specifically that cost is exogenous to responses and responses are interpreted as yes or no votes, a complexity arises when the actual cost falls between c_h and c_{h+1} . Depending on the density $p(c)$ and how the PC response may be translated into a vote, it is possible that either c_{h+1} or c_h is an optimal response. For instance, suppose the authority considers the response as a no vote for any c above the highest amount indicated on the PC. The voter should select c_{h+1} if the expected losses over the range $[v, c_{h+1}]$ are less than the expected gains over the range $[c_h, v]$, or otherwise select c_h . A heuristic such as considering the vote a yes if c is closer to c_h does not solve the problem, without placing additional assumptions on $p(c)$ and $q(\cdot)$. Given this issue, we propose the following condition:

(iii'') the PC response s is interpreted as a vote on a single project, consisting of non-cost attributes as described in the survey, and a cost c that is determinedly exogenously from s but taken from the set of cost amounts included on the PC. That $s \geq c$ ($s < c$) signals a yes (no) vote.

From the respondent perspective, condition (iii'') translates the PC response into a set of yes or no votes to projects defined by the set of costs included on the PC. This condition implies that expected utility is defined over a discrete rather than a continuous set of outcomes defined by the cost distribution:

$$(3) \quad \sum_{j=c_{\min}}^s \left[q(R_{-i}(c_j) + 1; G) U(v - c_j) + (1 - q(R_{-i}(c_j) + 1; G)) U(0) \right] p(c_j) + \sum_{j=s}^{c_{\max}} [q(R_{-i}(c_j); G) U(v - c_j) + (1 - q(R_{-i}(c_j); G)) U(0)] p(c_j)$$

Condition (iii'') is slightly stronger than what is needed, as the issue raised above only occurs when the cost falls in the same range of the voter's valuation $[c_h, c_{h+1}]$. Thus, we could instead restrict $p(c)$ to have positive density everywhere but within this range. Condition (iii'') may nevertheless be a reasonable approximation. Indeed, respondents may not focus on such interpretation subtleties when formulating a response. Alternatively, respondents may simply believe that if an intermediary cost arises, it will simply be adjusted up or down prior to interpreting responses (and, in turn the total number of payments would be adjusted). In practice, policymakers may be naturally inclined to avoid fractional costs or to, say, adjust \$23 to \$25.

PROPOSITION 2: An PC question is incentive compatible if conditions (i), (ii), (iii'') and (iv) hold.

PROOF: The proof follows from that for Proposition 1.

3. Survey description and experimental design

A. Background

On October 29, 2012, Hurricane Sandy hit the northeastern U.S. coastline. Based on storm surge predictions, mandatory evacuations were ordered on October 28, including for NYC’s Evacuation Zone A, the coastal zone at risk for flooding from any hurricane. By October 31, the region had 6–12 inches of precipitation, 7–8 million customers without power and approximately 20,000 persons in shelters. According to mortality tracking by the American Red Cross, there were 117 Sandy-related deaths, with nearly 75% of deaths occurring in New York and New Jersey.

In December 2012, Mayor Bloomberg launched the Special Initiative for Rebuilding and Resiliency and charged it with recommending steps the city should take to protect against the impacts of climate change. This culminated into the June 11, 2013 report “A Stronger, More Resilient New York” (NYC Special Initiative for Rebuilding and Resiliency, 2013), which provides recommendations for protecting neighborhoods and infrastructure from future climate events. Included as Chapter 3 of the report is a Coastal Protection Plan for NYC, which forms the basis of the flood protection proposal described in our survey.

B. Description of survey

The development of the survey was informed by extensive pretesting with focus groups convened at New York University, and under consultation with flood insurance experts and academics. Based on this information, we implemented as a pilot the theory-driven PC, using

cost amounts of \$1, \$2, \$3, \$5, \$7, \$10, \$15, \$20, \$25, \$35, \$50, \$70 and \$100. The pilot utilized just over 5% of our sample frame for Experiment 1. Nearly one-quarter of pilot respondents indicated a WTP of less than \$1, and fewer than 3% indicated they were willing to pay more than \$50 per month. Based on these results, in the subsequent PC and SBC bid designs we included a lower cost amount (50 cents), and omitted an intermediate amount (\$25) and the two highest amounts.⁶

The introduction to the survey emphasized the potential policy consequences of the survey, by relaying that the results will be used to inform policymakers about the opinions of NYC residents, and that results will be distributed to NYC’s Department of Environmental Protection, FEMA and other agencies interested in coastal flooding in NYC and elsewhere. After agreeing to participate, respondents were asked a series of questions about their home, insurance coverage and then presented with a flood protection proposal.

The flood protection proposal, consistent with the city’s proposed plan, described the construction of flood control measures including raised berms and deployable seawalls. To provide context for these systems we included in the survey the images displayed in Figure 1. We emphasized that while the images show particular places in the city, the flood control measures would be deployed “in the areas of the city most at risk of flooding”. The payment vehicle chosen was the household water bill, and the proposal involved an additional monthly fee on the water bill to recover construction and maintenance costs.

Following the proposal, an advisory referendum was described. This description necessarily varies across experimental treatments, and we provide the details in the next subsection. Respondents are reminded of their budget constraint, and the included language

⁶ We include pilot responses in our estimation sample, noting that the estimated WTP distribution from the pilot sample is statistical equivalent to the non-pilot sample for the theory-driven PC treatment.

emphasizes the “advisory” nature of the survey and that results may or may not be taken into consideration by authorities. The value elicitation question is framed as a vote for all treatments, with minimal adjustments in language made to accommodate the different question formats.

Follow-up questions to the value elicitation were included to probe respondents about the reasons underlying their decision(s), and, measured using Likert-scales, beliefs regarding both policy and payment consequentiality. After the valuation exercise respondents participated in an experiment where they were induced to search for flood insurance through a randomly presented subsidy. This aspect of the survey required using the online software to dynamically respond to home ownership and flood insurance status.⁷ For this reason all respondents were required to use the online survey, and no mail version of the survey was available. Finally, respondents participated in a risk preference elicitation and answered some demographic questions.

C. Experiment treatments

The main objective of the field experiments is to provide primary empirical evidence on theory-driven implementations of the PC and OE value elicitation questions. In an ideal world, we would be able to compare the different formats with a parallel measure that reflects actual demand. This is not possible for our application as the policy we analyze was merely in the proposal stage. Further, the large scale of the proposed flood control system prevents the use of experimental methods to devise an appropriate revealed preference benchmark. As a commonly used alternative, we include a SBC elicitation in our experimental design. Field validity tests that compare SBC surveys and parallel public referenda involving local environmental policies suggest a close correspondence (e.g., Johnston, 2006; Vossler and Watson, 2013).

⁷ Results of that experiment will be presented in a separate study.

Taking as given the inclusion of a SBC elicitation as a treatment, the design is further motivated by the question of whether the theory-driven enhancements matter empirically. There are many differences between theory-driven mechanisms we develop and common implementations. Importantly, the theory-driven mechanisms (a) introduce cost uncertainty, (b) suggest that cost is exogenous to stated valuations, (c) suggest one way in which responses may be interpreted, and (d) include in the proposal that the cost is the same across households, and that money collected would only go towards the proposed project. Our theory-driven versions also use a coercive payment vehicle and frame the elicitation as an advisory referendum. Although most recent CV surveys utilizing continuous response formats do not include either of the latter two features, their use is consistent with NOAA Panel recommendations, and there are now examples in the literature of studies with one or both of these attributes.

To fully address the research questions as presented above requires myriad treatments. As we were limited by sample size and budget, we instead converged on a smaller design that we view as a starting point to addressing the questions posed. There are five treatments, implemented across two experiments. In Experiment 1, we compare what we label as a theory-driven PC format with a SBC format. To yield an informative comparison, the SBC format is also informed by mechanism design theory. Thus, the experiment allows for a test of convergent validity between SBC and PC formats.

In Experiment 2, which utilizes a different sample, we compare a theory-driven OE format with SBC to test for convergent validity. The SBC elicitation is identical to that in Experiment 1. As a third treatment, we include what we label as a standard OE format. Comparing the two OE formats thus provides insight into whether the alterations we make is

meaningful in practice. We now highlight the specific ways we used the theory to inform our survey design.

Theory-driven OE format

The theory assumptions presented in Section 2 highlight that respondents must hold specific beliefs regarding how cost would be determined and how responses might be interpreted. To convey this in the survey, in describing the advisory referendum we state (Passage 1):

“Given the scope and cost of this project, it is important for us to learn the opinions of New York City residents. Some people might be willing to pay for these measures while others might not. At this point, the cost of the proposed flood control system is uncertain. Until detailed designs are completed and evaluated by engineers and architects, the monthly fee needed to fund the project is not known for sure. For this reason we are going to present you with an advisory referendum and ask you to indicate the highest amount you would be willing to pay for the project. This way, if cost information becomes available, we will be able to compare the necessary monthly fee with the amount you (and others) indicate you are willing to pay. We will then be able to know the percentage in favor and against the proposal at the resulting monthly fee.

Voting results from this study are not binding, but instead are advisory in nature. Results will be shared with local authorities, and these authorities may or may not take this information into consideration.

Please consider the advisory referendum below. Keep in mind that the purpose of asking you to indicate the highest amount you would pay is that, at this time, the fee needed to fund the project is uncertain. Consider that New York City has over 500 miles of shoreline. When considering your decision, please bear in mind that there may be other things that you would rather spend your money on. Think about how much, if anything, you are willing to pay before entering your decision. In the space below, write in the highest dollar amount at which you would still vote in favor of the program.”

Thus, the passage suggests that cost is (rightfully) uncertain at this time, uses cost uncertainty to motivate the value question format, and suggests that responses can be used to determine the percentage in favor at the realized cost amount. We do not explicitly mention that the local authorities would use information in this exact manner, as this of course cannot be guaranteed.

We speculate however that by mentioning this possible interpretation that respondents would be less likely to devise alternative interpretations, importantly including interpretations that would threaten the incentive compatibility of the elicitation.

Important for the theory is that cost is exogenous to responses. Passage 1 is not definitive about this, but we emphasized this point through the discussion of the payment mechanism (Passage 2):

“The seawall would be funded by adding a mandatory fixed fee to every New York City households’ monthly water bill for the foreseeable future. This fee would be the same for every household. The Department of Environmental Protection would use the money only to pay for the construction and maintenance of seawalls.”

Finally, the exact wording of the referendum is as follows:

“What is the highest mandatory fixed fee that City authorities could include on every New York City households’ monthly water bill, for the foreseeable future, for which you would still vote in favor of funding the proposed flood control system for New York City?”

Directly following the referendum, we included the following passage to further emphasize cost uncertainty and a possible interpretation of the OE responses (Passage 3):

“Remember, the cost of this project is very uncertain at this time, and this is why we are asking you to indicate the highest amount at which you would still vote in favor. This way, when the construction and maintenance costs are known, and the necessary monthly fee is calculated, we will be able to know the percentage in favor and against at this amount.”

Standard OE format

Relative to the theory-driven OE format, paragraphs 2 and 3 of Passage 1 are included verbatim, and all but the first, second and sixth sentence of the first paragraph is excluded. Passage 3 is excluded, and Passage 2 is replaced with a vague description of the payment vehicle:

“The seawall would be funded with money collected through fees added to New York City household monthly water bills.”

As such, any differences between the two OE formats may be driven by the presence/absence of the set of attributes (a) – (d) described above.

Theory-driven PC format

The description of the advisory referendum deviates only slightly from that used for the theory-driven OE format. Passage 2 from the OE elicitation is included verbatim, with only slight adjustments to Passage 1 and Passage 3 to account for the different value elicitation question. Importantly, we present the PC as a series of yes or no votes, as in Bateman et al. (2005) and Vossler and McKee (2006). This contrasts with the more commonly used PC format, where respondents are asked to circle the highest amount they would pay. Although this is merely speculative, framing the PC in this manner may better convey that, upon resolving cost uncertainty, the votes corresponding to actual cost will be viewed in isolation. This implementation of the PC also allows for possible “mistakes” in the form of switching from “no” to “yes” as cost increases, and thus gives rise to anecdotal evidence of whether the elicitation format is understood. The PC referendum reads:

“Should City authorities introduce a mandatory fixed fee of \$_____ to every New York City households’ monthly water bill, for the foreseeable future, to fund the proposed flood control system for New York City?”

We note that information provided in the survey is not fully consistent with theory assumption (iii’), in the sense it was not made explicit that, if implemented, the actual cost would be one of the amounts listed on the PC. This design choice was deliberate, as we felt such a statement would not be viewed as credible. Further, as highlighted in the theory section condition (iii’’) is

stronger than what is needed for incentive compatibility.

SBC format

We include a SBC elicitation that follows conditions (i) – (iv) presented in Section 2. Passage 2 is included verbatim, and there is no discussion of cost uncertainty. The advisory referendum is worded as for the PC, with the exception that a randomly selected cost amount is presented (rather than a blank). In particular, one of the 11 cost amounts included in the PC is selected, with equal probability.

D. Survey Samples

We constructed a sample of households living in and near the 100-year flood plain using data from FEMA Flood Insurance Rate Maps (FIRMs) and New York City's Primary Land Use Tax Lot Output (PLUTO) database. FIRMS are created using detailed hydrological and topographical studies. FEMA uses FIRMS to officially delineate flood risk zones nationwide. FEMA provides digital versions of these maps suitable for mapping in Geographic Information Systems through the National Flood Hazard Layer.

To identify households, we employed New York City's PLUTO database. PLUTO contains New York City Assessor's data for each parcel in the city as well detailed latitude and longitude data. Using a Geographic Information System, we mapped each NYC parcel to a FEMA flood zone. The parcel data also include a rich set of information about the parcel and structures including structure type, assessed value, square footage, number of floors and whether

the structure has a basement. We employ this data to confirm our randomization of treatments was effective and to identify heterogeneity in elicited values.

We restrict our sample frame to single family residences in New York City for two reasons.⁸ First, in the survey the payment vehicle is a fee in the household water bill. Households in single family residences (owner or renter occupied) typically pay their own water bill while many residents in multifamily buildings have their water, oil and electric bills folded into their rent or paid through some type of collective assessment. Additionally, the assessor's data includes the parcel address, but very limited information on the number of units in multifamily dwellings, and no apartment numbers, making it difficult to reach these households by mail.

Table 1 summarizes the number of single family households in our data by flood zone and NYC borough. Just over one-third of the parcels in the city are single family dwellings and only four percent of parcels are in the 100-year flood plain. Relative to the number of tax parcels in the city our flood plain sample has more parcels in Staten Island and the Bronx is underrepresented.

Each of the 13,342 single family households in the high risk zone were selected into the sample for Experiment 1. We also used GIS to identify parcels that lived within 500 meters beyond the 100-year flood zone boundary, which generated an additional set of just over 23,500 households. Of these, 12,000 were randomly selected to be included in the sample for Experiment 2. These households, as well as much of the rest of the city are inside the 500-meter buffer zone. Figure 2 maps the high risk flood plain across the city.

⁸ We identify single family households by restricting our sample to parcels with ownership type of "P" or null in the PLUTO data. The majority of records in PLUTO are blank and the documentation states that null indicates "Unknown (Usually Private Ownership)". To assure that we have restricted the sample to single family homes we also include only parcels with Building Class Codes beginning with A, which signifies "One Family Dwellings".

In the summer and fall of 2015, selected households were sent a letter inviting them to direct their web browser to a specific URL and complete the survey. Each letter included a unique four-character code that allows us to map survey response to assessor's data and flood zones. The PLUTO extract includes the name of the parcel owner according to tax documents, but we cannot be certain which households are owner occupied and which are rentals. There are also some concerns that the owner name field might not be updated in our extract of the assessor's data. For this reason, each mailing is addressed to "Flood Zone Resident". Upon accessing the survey website respondents are provided information on the study and prompted to enter their unique access code. The survey itself was created and hosted in Qualtrics, an online survey software suite. The survey was presented in eight sequential pages and took an average of just over ten minutes to complete. After completing a page respondents clicked on a button and were unable to return. Respondents were asked in which flood zone their home was located, but they were not told their flood zone (as mapped using the FEMA flood maps and PLUTO data) until after completing the value elicitation exercise.

Ten days after sending the initial letter a follow-up post card was sent to all households who had not yet responded. A final reminder was sent one week later, and households were informed that we would not contact them further.⁹ The letter and each of the follow up post cards reminded households that ten respondents would be randomly selected to receive a \$100 gift card. To minimize mailing processing fees, the two samples were surveyed approximately one month apart.

Across the two experiments we mailed 24,640 letters and received 1,719 complete responses for a response rate of 7%. For the households in the 100-year flood plain we mailed

⁹ Anonymized versions of the initial letter and the reminder post cards appear in the appendix.

12,640 letters of which 88 were returned as undeliverable. The mailings led to 1,128 surveys being started and 900 completed, for a dropout rate of just over 20%. In the second experiment we randomly selected 12,000 households living within 500 meters of the 100-year flood plain to recruit by mail. Only 15 of those letters were returned as undeliverable. 1,019 respondents began the survey and 819 completed it, for a dropout rate of 25%.

4. Results

In this section we describe the results of our two experiments. We begin by describing the data. We follow this with a formal econometric analysis of WTP, with a focus on testing not only convergent validity but on assessing the extent to which theory is important for stated preferences.

Table 2 describes information on our respondents, using both assessor's data and survey data, for Experiment 1. Table 3 provides parallel data for Experiment 2. As evident from summary statistics, the respondents were properly randomized into treatments within each experiment. The main difference across the two experiment samples is that a higher proportion of the Experiment 1 sample has flood insurance, experienced property damage from Hurricane Sandy and were forced to evacuate. Of course, this is to be expected given these respondents are in the 100-year flood plain.

Table 4 presents the raw proportion of Experiment 1 respondents indicating a yes vote across the possible cost amounts included in the survey. The survival function for the PC is monotonically decreasing, with 70% of respondents voting yes at the lowest cost level (\$0.50 per month) and 6% for the highest cost level (\$50). Percentages for the SBC treatment vary from a high of 86% at a cost of \$2 per month to a low of 28% at \$50 per month. Although the

percentage of yes votes, as expected, is generally decreasing with cost, this relationship is not monotonic. Of course each percentage is based on a different random sample, and such violations of (aggregate) monotonicity are common in the literature.

The distributions are similar across treatments at lower cost levels, but as the presented cost rises the proportion of yes votes falls more quickly for the PC. Indeed, for a range of higher costs there are stark differences of over 20% and 30%.

Table 5 reports the proportion of yes votes in Experiment 2. To compare the OE distributions with the SBC data we construct empirical survival functions with the raw data and report the acceptance rates as they correspond with SBC cost levels. The percentage of yes votes at the lowest cost level varies somewhat by treatment, with the theory-driven OE producing 81% yes votes, the SBC 75% yes votes, and the standard OE 68% yes votes. The two OE formats diverge by roughly 10% for low cost amounts, but are similar at costs of \$5 and higher. In examining the raw data, the percentage of reported zero valuations is different: 32% for standard OE, and 19% for theory-driven OE. Using a Fisher exact test, the proportion of zero responses is statistically different.¹⁰ Carson and Groves (2007) argue that, in a handful of cases where incentive-compatibility is lost, theory predicts OE respondents to strategically report zero. As such, this finding provides some suggestive evidence that the theory-driven elicitation may serve to circumvent strategic zeros. Compared to the OE treatments, the SBC begins with roughly the same proportion of yes votes, but at the highest costs levels the proportions of yes votes are somewhat higher.

Finally, in Tables 6 – 9 we report the distribution of responses to two survey questions, one that targets beliefs tied to policy consequentiality and the other, to payment consequentiality.

¹⁰ Unless otherwise noted, in drawing conclusions we use a 5% significance level throughout the analysis.

Payment consequentiality was elicited using a four-point Likert scale ranging from “definitely no” to “definitely yes” as to whether participants would have to pay if a flood control system were built. For the policy consequentiality question, responses were on a five-point Likert scale ranging from their responses having “no effect” on the probability the flood control system is built to being “absolutely crucial”.

In Experiment 1, for both consequentiality questions there are similar response distributions across the SBC and PC treatments. A majority of respondents answered “Maybe Yes” to the payment consequentiality question. Around 80 percent of respondents in the PC treatment indicated the elicitation to be at least weakly consequential in both dimensions, whereas this figure is 75 percent with SBC.

In Experiment 2 the differences in stated consequentiality across treatments are more pronounced. A majority of respondents still choose “Maybe Yes” as their answer to the payment consequentiality question in each treatment, but around five percent more respondents answer “Definitely No” for the standard OE elicitation relative to the theory-driven OE. Interestingly, an important difference arises when we focus on the percentage of respondents indicating at least weak agreement with both consequentiality questions. These figures are 72% and 69% for SBC and theory-driven OE formats, respectively, but just 59% for the standard OE. In both pairwise comparisons, the standard OE treatment yields a statistically different percentage based on Fisher exact tests. This provides suggestive evidence that the theory enhancements help in fostering beliefs that are theoretically relevant for truthful demand revelation.

A. Econometric analysis

The data generated from the three types of value questions give rise to continuous, left-censored, right-censored or interval-censored signals of WTP. The maximum likelihood estimator we specify, which nests the estimators of Cameron and James (1987) and Cameron and Huppert (1989), accommodates the different types of information elicited across question formats in a unified way, which allows for straightforward tests of treatment effects. The estimation allows for the possibility that some individuals have negative WTP. This is plausible as, for instance, respondents may experience disutility due to obstruction of views and modifications to the coastal landscape. As suggestive evidence that some valuations may be negative, we note again that there are many zeros for the OE elicitations, and a fair fraction of PC and SBC respondents who vote no to the lowest cost amount (50 cents).

Let WTP_i denote respondent i 's willingness to pay for the proposal. WTP_i is not directly observed, except for a subset of OE respondents, but instead can be treated as a censored dependent variable. For a PC elicitation, we obtain the signal $c_{i,l} \leq WTP_i < c_{i,u}$ where $c_{i,l}$ is the highest cost for which the participant votes yes and $c_{i,u}$ is the next highest amount. For the case where the respondent votes no to the lowest amount, $c_{i,l} = -\infty$; similarly, $c_{i,u} = \infty$ if she votes yes to the highest amount. For a SBC elicitation, we obtain the signal $WTP_i < c_i$ if the agent votes no to the stated cost c_i and $WTP_i \geq c_i$ for a yes vote. As such, the SBC data represent a special case of PC data, where c_i defines only an upper or lower bound on the WTP interval (e.g., $c_{i,l} = c_i$ and $c_{i,u} = \infty$ for a yes vote). For OE responses, WTP_i is directly observed, with the exception of zero valuations, which we interpret to be left-censored to allow for possibly negative WTP, which is consistent with our treatment of SBC and PC data. These left-censored observations are accommodated into the PC framework by defining the WTP interval with $c_{i,l} = -\infty$ and $c_{i,u} = 0$.

Assume WTP_i is a linear function of a row vector of covariates, $\mathbf{x}_i$, such that $WTP_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$, where $\boldsymbol{\beta}$ is a column vector of unknown parameters and ε_i is a normally distributed mean-zero error term with standard deviation σ_i . With the linear conditional mean function, assuming the error term has a normal distribution is analogous to assuming a normal distribution for WTP_i . Moreover, interpretation of estimated parameters is the same as for a standard linear regression model that treats WTP_i as a directly observed (i.e. uncensored) dependent variable. Let $D_i = 1$ if the response is censored (PC, SBC; OE zero responses), and $D_i = 0$ otherwise (uncensored OE responses). Then, the log-likelihood function is

$$(4) \quad \ln \mathcal{L} = \sum_{i=1}^N \left\{ D_i \cdot \ln \Phi \left(\left(\frac{c_{iu} - \mathbf{x}_i\boldsymbol{\beta}}{\sigma_i} \right) \right) - \left(\frac{c_{il} - \mathbf{x}_i\boldsymbol{\beta}}{\sigma_i} \right) \right\} + (1 - D_i) \cdot \ln \left(\frac{1}{\sigma_i} \phi \left(\frac{WTP_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma_i} \right) \right),$$

where Φ and ϕ denote the CDF and PDF of the standard normal distribution, respectively. The first term corresponds with the log-likelihood for an interval regression whereas the second corresponds to that of a normal regression model.¹¹ The estimator for OE data is thus a Tobit with left-censoring at zero.

Studies such as Haab, Huang and Whitehead (1999) highlight the importance of allowing for different error variances when pooling preference data from different experimental treatments. To allow for possibly different error variances across response formats, for instance PC and SBC, we can define $\sigma_i = \sigma_0 + PC_i\sigma_1$, where $PC_i = 1$ for PC observations. In this formulation, σ_0 is the standard deviation of the errors for the SBC data and $\sigma_0 + \sigma_1$ is the standard deviation for the PC data. This is easily extended for Experiment 2 data, where there are three treatments.

¹¹ Given how we code the upper and lower bounds, the contribution to the log-likelihood for a SBC respondent is mathematically equivalent to that of a probit model.

Table 10 reports the results of the WTP regressions for Experiment 1. Model 1 allows for a simple unconditional test of the treatment effect. The coefficient on the indicator for the PC treatment is negative and statistically significant, indicating a difference in mean WTP between the two treatments. In fact, SBC respondents indicate they are willing to pay twice as much. Mean WTP for the SBC elicitation is \$20.25 (SE = 2.93), and is \$8.67 (1.01) for the PC. The standard deviation of the error term is notably smaller for the PC treatment, indicating that the WTP distribution for this format has less dispersion along with a lower mean.

The survey asked a number of questions about experience with Hurricane Sandy as well as sociodemographic characteristics. We merged in data on house characteristics from the county assessor's data file. We employ answers to some of these questions as controls, but several of the survey questions suffer from item non-response. Model 2 is estimated using exactly the same specification as Model 1, but on a sample restricted to respondents with complete data. This represents about 90% of the sample. There are no notable differences between the estimates across Models 1 and 2.

Model 3 includes the socio-demographic and housing variables. These covariates may explain some of the variation in WTP across individuals, while also serving to control for unintended differences due to sampling. We further include interactions between treatment-specific indicators and the dummy variable *Inconsequential*, which equals 1 if the respondent indicated “definitely no” to the payment consequentiality or “no effect” to the policy consequentiality question. The model thus allows for differences for those stated beliefs inconsistent with two of the incentive compatibility assumptions.¹² We demean the socio-

¹² Given there is mixed evidence on whether measures of consequentiality can be treated as exogenous in WTP regressions, we also estimated instrumental variables models that parallel those reported as models 3 and 6, using the control function approach described in Wooldridge (2010, p. 784). Using as IVs various measures found to be strongly correlated with consequentiality but insignificant determinants of WTP, such as response certainty, SBC

demographic and housing variables, which allows the estimated intercept, which is \$30.59, to be directly interpretable as mean WTP for SBC treatment respondents who perceive consequentiality.

From this model, there is a large and statistically significant effect of consequentiality on both PC and SBC values. This effect is more pronounced for the SBC elicitation, where the model estimates that WTP is \$35.52 lower for those who do not perceive consequentiality to hold. The direction of this finding mirrors results from other recent studies utilizing a SBC format (e.g., Herriges et al., 2010; Vossler and Watson, 2013). Conditioning on consequentiality does not alter the previous result of a significant difference between elicitation formats. In fact, the difference in WTP nearly doubles from roughly \$11 to over \$20.

Turning to the effects of other control variables, we find that the income effect is positive and statistically significant, which can be construed as evidence of construct validity. Females are willing to pay \$4.03 less than males, and participating in extreme sports (a proxy for risk-taking behavior) decreases WTP. Respondents indicating property damage from Hurricane Sandy and those living in houses with a basement have a statistically higher WTP. The direction of these effects are intuitive. As most households living in the 100-year flood plain have flood insurance, it is not overly surprising that this indicator variable has no statistically significant effect.

Table 11 reports parallel WTP regressions for Experiment 2 data. The baseline is again the SBC treatment. Recall that the samples are different across the two experiments, with Experiment 1 respondents drawn from the 100-year flood plain and Experiment 2 respondents from just outside this flood plain. From Model 4, mean WTP for the SBC treatment is \$10.72,

cost amounts and short-term flood risk perceptions, we fail to reject exogeneity of the consequentiality-related variables for both experiments. Additional details of these models are available from the authors upon request.

which is roughly one-half the corresponding Experiment 1 estimate. As respondent characteristics (aside from insurance and effects from Hurricane Sandy) are similar across the two samples, this result can be interpreted as evidence of scope sensitivity to exogenous differences in flood risk. Mean WTP for standard OE is \$15.18 (6.90) lower than the SBC and the difference is statistically significant. In contrast there is no statistical difference between the SBC and theory-driven OE formats, and there is less than a 10% difference in point estimates. Further, as indicated by the estimated standard deviation parameters, there is no statistical difference in the variances either for these two formats. In contrast, the standard OE variance is substantially higher.

Turning to Model 5, the same basic conclusions arise when restricting the sample to the 78% of respondents with complete sociodemographic characteristics. The mean WTP estimate from SBC does increase modestly, and the difference between SBC and standard OE becomes only weakly significant ($p=0.06$). Model 6 again allows for in consequentiality to have a different effect across each treatment, and includes other controls. Similar to what we observed in Experiment 1, in consequentiality has a large and negative effect on WTP for all treatments. Conditional on consequentiality holding, mean WTP for the three treatments are: SBC, \$21.31 (3.04); theory-driven OE \$22.02 (4.20); standard OE \$21.13 (6.21). Interestingly, there are no pairwise differences in these estimates and further the point estimates are very similar. Recalling that the fraction of in consequential respondents is higher for the standard OE, this serves as one explanation for the unconditional difference in WTP.

Turning to the sociodemographic and housing characteristics added to Model 6, the results are generally consistent with expectations. As with Experiment 1 these variables have been demeaned to allow us to interpret the treatment indicators as WTP measures at the mean of

these covariates. Income is again positive and statistically significant. Households with active flood insurance policies, and those who had property damage from Hurricane Sandy, have a higher WTP. Similar to the Experiment 1 findings, being female and participating in extreme sports decreases WTP. Persons indicating membership in an environmental organization interestingly have much lower WTP (about \$14 less).

5. Discussion

In this study we identified sets of incentive compatibility conditions for two continuous response formats: purely open-ended (OE) questions and payment cards (PCs). These continuous response formats are often cited as *not* being incentive compatible, and a careful discussion on this issue is provided by Carson and Groves (2007). Thus, with this paper we highlight that incentive compatibility may not be an elusive goal when considering alternative elicitation formats. Our theory work, when considered along with the work of Vossler, Doyon and Rondeau (2012), suggests that incentive compatibility is not necessarily lost in surveys that either elicit a continuous response or include two or more value elicitation questions. The work to date thus encompasses many of the contingent valuation question formats in use.

The theory assumptions we identify are unlikely to have been met in prior studies, however, and with the hope of informing practice we illustrate one possible execution of the theory. One particular issue with the OE and PC formats, in their standard implementations, is that they leave much to the respondents' imaginations. An OE format for instance may remind us of a situation where we are haggling over price. In speculating about how survey results might be aggregated, one would naturally think that valuations would be summed up or averaged, which can lead to extreme stated valuations to influence the outcome in a discernable way. The survey

refinements we implement, in particular information relaying how responses can be interpreted as votes in an advisory referendum, are likely to restrict these undesirable speculations by providing what has traditionally been missing information.

On a related note, such speculations over the interpretation and use of values elicited through continuous response formats may vary wildly across case studies and populations. Even a person confronted with the same valuation scenario at different points in time may provide different responses depending on what particular ideas and beliefs have entered her mind during the valuation exercise. To the extent that survey refinements are widely adopted, this is thus likely to enhance reliability by eliminating nuisance variables.

Important for practitioners, the survey refinements we highlight can be applied broadly and amount to adding a few paragraphs to the value elicitation scenario and by framing of the value elicitation as an advisory referendum. Moreover, although we limited our attention to PC and OE formats in this paper, we speculate that similar theoretical assumptions and associated empirical implementations can be identified for other question formats that ask the same respondent about different cost amounts, such as the double-bounded dichotomous choice or multiple-bounded discrete choice formats.

Central to the theory is that the cost to the respondent upon project implementation is not known with certainty but would be exogenously determined based on actual project costs if plans moved forward. That cost is uncertain should be palatable to respondents, and in early planning stages of a project there typically is considerable cost uncertainty. Those skeptical about the policy process could still suspect the survey information may be used to extract additional revenue. There is certainly merit in gauging beliefs related to the plausibility of this cost

determination process in focus groups, along with other aspects of the valuation scenario. In our case study at least, no issues were identified.

One important takeaway message from our field experiment is that we fail to reject convergent validity between our single binary choice (SBC) and theory-driven OE elicitation, all the while finding significantly lower values for a counterfactual “standard” OE format included in the design. That the standard elicitation yields lower values is consistent with the stylized fact from the literature, whereas equivalence is rare. One interpretation of our results is thus that “theory matters”. Related to this, we find that the proportion of respondents who view the elicitation as consequential is similar between the theory-driven OE and SBC formats, but statistically lower for the standard OE elicitation. Moreover, to the extent that a SBC elicitation reveals demand, the result suggests that it is possible to use an OE mechanism to achieve this. If such a finding can be broadly replicated, this sheds a bright light on the use of this format, and would allow researchers to avoid the well-cited disadvantages of SBC elicitation – the sparse preference information elicited and complications associated with bid design.

The results related to convergent validity are not all positive, however, as we find that the SBC elicits higher values than our theory-driven PC. Unfortunately, our sample frame did not allow for a “standard” PC to be included in the design, and so it is not known whether our theory enhancements to the PC had an empirical impact. The theory assumptions are stronger for the PC relative to the OE format, given an ambiguity that arises when the actual cost falls between two amounts listed on the PC, and this is a possible driver of results. In our opinion, however, this explanation is not likely. Indeed, if we take an extreme view and code all WTP intervals up a level, this does little to bridge the WTP gap. A second possible explanation is that there is a behavioral effect of providing respondents with a set of possible payment amounts. Such

possible value cues are largely absent from both the SBC and OE formats. A third explanation is that the PC may be an inherently difficult incentive mechanism to decipher. However, using an induced-value experiment with financial incentives, Vossler and McKee (2006) find that both SBC and PC questions are demand revealing, with very few responses inconsistent with induced valuations. Thus, we believe the more palatable explanation is that PC and SBC formats may give rise to inherently different value formation processes that stem from the differential information provided on possible costs.

On a final, policy note, the results from our experiment bode ill for efforts to fund extensive flood protection systems in the New York City area. The mean WTP of households in the 100-year flood plain, conditioned on respondent beliefs that the elicitation is consequential, is estimated to be around \$30. Even taking the upper-end of the 95% confidence interval associated with this estimate, perceived benefits do not appear sufficient to support the construction of new flood control systems. Only 36 thousand households live in the 100-year flood plain in New York City. A fee of \$40 a month from these households would generate less than \$15 million a year in revenue. Expanding to include the 23 thousand families in the 500-meter buffer zone only increases the flow of revenue to \$25 million a year. While, as noted earlier, detailed costs estimates are not available, the city has projected that a full flood control and climate resiliency plan could cost around \$20 billion. While the benefits of such a program can be argued to exceed the costs, the results presented here suggest that NYC residents in and around the 100-year flood plain may not be willing to pay for the system.

References

- Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner and Howard Shuman, 1993. Report of the NOAA Panel on Contingent Valuation. *Federal Register* 58, 4601-4614.
- Bateman, Ian J., Philip Cooper, Stavros Georgiou, Stale Navrud, Gregory L. Poe, Richard C. Ready, Pere Reira, Mandy Ryan and Christian A. Vossler. 2005. Economic Valuation of Policies for Reducing Acidity in Remote Mountain Lakes. *Aquatic Sciences* 67(3), 274-291.
- Becker, Gordon M., Morris H. DeGroot and Jacob Marschak, 1964. Measuring Utility by a Single-Response Sequential Method. *Behavioral Science* 9(3), 226-232.
- Bichard, E. and Kazmierczak, A. 2012. Are homeowners willing to adapt to and mitigate the effects of climate change? *Climatic Change* 112, 633-654.
- Botzen, W J.W., and J.C.J.M. van den Bergh. 2012a. Monetary Valuation Of Insurance Against Flood Risk Under Climate Change. *International Economic Review* 53(3), 1005-1026.
- Botzen, W J.W., and J.C.J.M. van den Bergh. 2012b. Risk attitudes to low-probability climate change risks: WTP for flood insurance. *Journal of Economic Behavior & Organization* 82, 151-166.
- Cameron, Trudy A. and Daniel D. Huppert. 1989. OLS versus ML estimation of non-market resource values with payment card interval data. *Journal of Environmental Economics and Management* 17, 230-246.
- Cameron, Trudy A., James, Michael D. 1987. Efficient estimation methods for use with “closed-ended” contingent valuation survey data. *The Review of Economics and Statistics* 69, 269-276.
- Carson, Richard T. and Theodore Groves. 2007. Incentive and informational properties of preference questions. *Environmental and Resource Economics* 37, 181-210.
- Carson, Richard T., Theodore Groves and John A. List. 2014. Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists* 1(1-2), 171-207.
- Champ, Patricia A. and Richard C. Bishop. 2006. Is willingness to pay for a public good sensitive to the elicitation format? *Land Economics* 82(2), 162-173.
- Dillman, Don A., Jolene D. Smyth and Leah Melani Christian, 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th Edition*. Hoboken, New Jersey: John Wiley & Sons.
- Haab, Timothy C., Ju-Chin Huang and John C. Whitehead, 1999. Are hypothetical referenda incentive compatible? A comment. *Journal of Political Economy* 107(1), 186-196.

- Herriges, J., Kling, C., Liu, C.-C., Tobias, J. 2010. What are the consequences of consequentiality? *Journal of Environmental Economics and Management* 59(1), 67-81.
- Johnston, Robert J. 2006. Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum. *Journal of Environmental Economics and Management* 52, 469-481.
- Kling, Catherine L., Daniel J. Phaneuf and Jinhua Zhao. 2012. From Exxon to BP: Has Some Number Become Better Than No Number? *Journal of Economic Perspectives* 26(4): 3-26.
- Messer, Kent D., Gregory L. Poe, Daniel Rondeau, William D. Schulze and Christian A. Vossler, 2010. Social Preferences and Voting: An Exploration Using a Novel Preference Revealing Mechanism. *Journal of Public Economics* 94: 308-317.
- NYC Special Initiative for Rebuilding and Resiliency. 2013. A stronger, more resilient New York. Last accessed January 26th, 2016: <http://www.nyc.gov/html/sirr/html/report/report.shtml>.
- Vossler, Christian A., Maurice Doyon and Daniel Rondeau. 2012. Truth in consequences: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics* 4(4), 145-171.
- Vossler, Christian A. and Mary F. Evans. 2009. Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. *Journal of Environmental Economics and Management* 58(3), 338-345.
- Vossler, Christian A. and Michael McKee, 2006. Induced-Value Tests of Contingent Valuation Elicitation Mechanisms. *Environmental & Resource Economics* 35: 137-168.
- Vossler, Christian A. and Sharon B. Watson. 2013. Understanding the Consequences of Consequentiality: Testing the Validity of Stated Preferences in the Field. *Journal of Economic Behavior & Organization* 86: 137-147.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*, Second Edition. Cambridge, MA: The MIT Press.

Table 1. Number of households by New York City borough

	Total	Single family	100-year flood plain	Single family & 100-year flood plain	Single family & 500-meter buffer zone
Brooklyn	283,913	62,313	7,644	2,478	2,702
Bronx	94,784	23,316	3,505	1,241	3,822
Manhattan	44,948	2,016	1,968	21	86
Queens	335,450	158,713	12,881	4,417	7,804
Staten Island	131,488	80,224	10,085	5,275	9,087
Total	890,583	326,582	36,083	13,432	23,501

Notes: 12,000 of the 23,501 households in the single family-500-meter buffer zone sample frame were randomly selected to be recruited into the survey. All 13,432 households in the 100-year flood plain were recruited as part of either the pilot of the main wave of the survey.

Table 2. Summary Statistics by Treatment, Experiment 1

	Single binary choice	Theory-driven payment card
Income (thousands \$)	100.1 (79)	99.1 (83)
Female	0.54 (.50)	0.52 (0.50)
Environmental Org. Member	0.10 (0.30)	0.09 (0.29)
Flood Insurance	0.77 (0.42)	0.82 (0.39)
Damaged in Hurricane Sandy	0.81 (0.39)	0.84 (0.37)
Evacuate in Hurricane Sandy	0.41 (0.49)	0.48 (0.50)
Basement	0.76 (0.43)	0.75 (0.43)
Extreme Sport Participant	0.18 (0.38)	0.18 (0.39)
Inconsequential	0.29 (0.46)	0.25 (0.43)
N	478	244

Note: Summary statistics by treatment type for Experiment 1 households that live inside the 100-year flood zone. Flood insurance is an indicator for households that answered that they have an active flood insurance policy. Damaged and evacuated in Hurricane Sandy are indicators for respondents who answered yes to those questions. Inconsequential is an indicator for respondents who answered that either they did not believe their vote would not affect the probability that the flood control system was built or that if it was built they would not have to pay. See the main text for details.

Table 3. Summary Statistics by Treatment, Experiment 2

	Single binary choice	Theory-driven Open-ended	Standard Open-ended
Income (thousands \$)	116 (91)	113 (116)	110 (82)
Female	0.50 (0.50)	0.46 (0.5)	0.49 (0.50)
Environmental Org. Member	0.09 (0.3)	0.10 (0.3)	0.08 (0.30)
Flood Insurance	0.43 (0.50)	0.41 (0.49)	0.45 (0.50)
Damaged in Hurricane Sandy	0.47 (0.50)	0.50 (0.50)	0.45 (0.50)
Evacuated in Hurricane Sandy	0.23 (0.42)	0.19 (0.39)	0.20 (0.40)
Basement	0.71 (0.45)	0.73 (0.44)	0.76 (0.43)
Extreme Sport Participant	0.16 (0.37)	0.13 (0.34)	0.15 (0.36)
Inconsequential	0.24 (0.43)	0.32 (0.47)	0.38 (0.49)
N	300	193	193

Note: Summary statistics by treatment type for experiment 2 households that live just outside the 100-year flood zone. Flood insurance is an indicator for households that answered that they have an active flood insurance policy. Damaged and evacuated in Hurricane Sandy are indicators for respondents who answered yes to those questions. Inconsequential is an indicator for respondents who answered that either they did not believe their vote would not affect the probability that the flood control system was built or that if it was built they would not have to pay. See the main text for details.

Table 4

Proportion of Yes Votes by Cost Level, Experiment 1

Cost	Theory-driven payment card (PC)	Single binary choice (SBC)
0.5	0.70	0.68
1	0.69	0.76
2	0.57	0.86
3	0.53	0.49
5	0.52	0.56
7	0.33	0.60
10	0.32	0.49
15	0.16	0.43
20	0.15	0.53
35	0.06	0.43
50	0.06	0.28

Notes: Payment card treatment N=231 and dichotomous choice treatment N=535 with between 45 and 51 responses at each cost level.

Table 5

Proportion of Yes Votes by Cost Level, Experiment 2

Cost	Single binary choice (SBC)	Theory-driven open ended	Standard open ended
0.5	0.75	0.81	0.68
1	0.63	0.80	0.68
2	0.63	0.75	0.64
3	0.63	0.69	0.63
5	0.57	0.66	0.63
7	0.53	0.52	0.53
10	0.46	0.52	0.52
15	0.18	0.30	0.31
20	0.29	0.27	0.29
35	0.35	0.13	0.14
50	0.21	0.13	0.13

Notes: Open ended responses were tabulated at levels consistent with costs included in the SBC design. Sample sizes: N=382 (SBC); N=249 (standard OE); N=244 (theory-driven OE).

Table 6

Experiment 1 Payment Consequentiality

Treatment	Definitely		Maybe Yes	Definitely	
	No	Maybe No		Yes	N
Single binary choice	0.10	0.08	0.56	0.27	531
Payment card	0.11	0.08	0.57	0.24	263

Notes: Cell entries indicate the fraction of respondents who selected the indicated response option to the question “If the City goes forward with a plan to construct the seawall, do you think New York City households will have to pay for it?”.

Table 7

Experiment 1 Policy Consequentiality

Treatment	No Effect	Small Effect	Moderate Effect	Large Effect	Absolutely Crucial	N
Single binary choice	0.21	0.29	0.26	0.14	0.10	534
Payment card	0.17	0.29	0.29	0.14	0.11	263

Notes: Cell entries indicate the fraction of respondents who selected the indicated response option to the question “To what degree do you believe that the advisory referendum decision from you and other survey participants will affect whether a flood control system is built?”.

Table 8

Experiment 2 Payment Consequentiality

Treatment	Definitely		Maybe Yes	Definitely	
	No	Maybe No		Yes	N
Single binary choice	0.17	0.09	0.57	0.18	381
Standard open-ended (OE)	0.20	0.10	0.50	0.21	246
Theory-driven OE	0.14	0.15	0.58	0.14	237

Notes: Cell entries indicate the fraction of respondents who selected the indicated response option to the question “If the City goes forward with a plan to construct the seawall, do you think New York City households will have to pay for it?”.

Table 9

Experiment 2 Policy Consequentiality

Treatment	No Effect	Small Effect	Moderate Effect	Large Effect	Absolutely Crucial	N
Single binary choice	0.14	0.31	0.32	0.12	0.10	381
Standard open-ended (OE)	0.25	0.30	0.22	0.16	0.07	247
Theory-driven OE	0.22	0.27	0.31	0.15	0.05	237

Notes: Cell entries indicate the fraction of respondents who selected the indicated response option to the question “To what degree do you believe that the advisory referendum decision from you and other survey participants will affect whether a flood control system is built?”.

Table 10

Willingness-to-pay regressions for Experiment 1.

	(1)	(2)	(3)
Theory-driven payment card (PC)	-11.58*** (3.10)	-11.42*** (3.51)	-20.18*** (4.97)
Inconsequential × Single binary choice (SBC)			-35.52*** (9.30)
Inconsequential × Theory-driven PC			-7.16*** (2.54)
Income (thousands US\$)			0.03** (0.01)
Female			-4.03** (1.94)
Environmental Org. Member			0.45 (2.55)
Flood insurance			0.65 (2.87)
Damaged in Hurricane Sandy			6.90** (3.03)
Evacuated in Hurricane Sandy			2.25 (2.16)
Basement			6.77*** (2.27)
Extreme Sport Participant			-4.31* (2.26)
Constant	20.25*** (2.93)	20.46*** (3.34)	30.59*** (4.84)
<i>Standard deviation function (σ)</i>			
Theory-driven PC	-31.38*** (8.63)	-34.41*** (10.46)	-30.34*** (9.39)
Constant	47.37*** (8.49)	50.56*** (10.34)	45.74*** (9.21)
Log-L	-1006.61	-937.20	-902.32
N	796	720	720

Notes: “Inconsequential” is an indicator equal to 1 for respondents with stated beliefs inconsistent with payment and/or policy consequentiality as described in the main text. All sociodemographic and house characteristics have been demeaned to ease interpretation. Robust standard errors reported in parentheses. *** significant at the 1% level, ** significant at the 5% level and * significant at the 10% level.

Table 11

Willingness-to-pay regressions for Experiment 2.

	(4)	(5)	(6)
Theory-driven open-ended (OE)	1.52 (3.41)	-1.56 (3.74)	0.71 (4.83)
Standard OE	-15.18** (6.90)	-13.75* (7.22)	-0.18 (6.83)
Inconsequential $\times$ Single binary choice (SBC)			-29.31*** (7.42)
Inconsequential $\times$ Theory-driven OE			-30.26*** (9.13)
Inconsequential $\times$ Standard OE			-58.52** (27.20)
Income (thousands US\$)			0.09** (0.04)
Female			-9.59** (4.01)
Environmental Org. Member			-14.01** (6.72)
Flood insurance			7.38* (4.40)
Damaged in Hurricane Sandy			7.78* (4.46)
Evacuated in Hurricane Sandy			-3.47 (4.55)
Basement			-5.32 (4.53)
Extreme Sport Participant			-8.31* (4.98)
Constant	10.72*** (2.53)	14.54*** (2.67)	21.31*** (3.04)
<i>Standard deviation function (σ)</i>			
Theory-driven OE	12.91 (14.71)	18.32 (16.40)	21.10** (10.70)
Standard OE	80.14** (36.44)	63.08 (39.07)	67.95* (37.15)
Constant	38.32*** (7.27)	34.94*** (6.74)	28.45*** (5.24)
Log-L	-2450.50	-1930.92	-1880.07
N	875	686	686

Notes: Inconsequential is an indicator equal to 1 for respondents with stated beliefs inconsistent with payment and/or policy consequentiality as described in the main text. All sociodemographic and house characteristics have been demeaned to ease interpretation. Specification allows for heterogeneous variances across treatments. Robust standard errors reported in parentheses. *** significant at the 1% level, ** significant at the 5% level and * significant at the 10% level.



Figure 1. Flood wall and berm images

Notes: Images of a deployable seawall (left) and bridging berm (right) submitted as part of a contest to design a flood control system for New York City hosted by Rebuild by Design. Source: <http://www.rebuildbydesign.org> [last accessed August 10, 2016]



Figure 2. New York City's 100 year flood plain (source: FEMA flood maps distributed by WNYC)