Alternative Value Elicitation Formats in Contingent Valuation: Mechanism Design and Convergent Validity

Christian A. Vossler and J. Scott Holladay

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

University of Tennessee, Knoxville, TN

The published version of this study is in the *Journal of Public Economics*. The final publication is available at https://doi.org/10.1016/j.jpubeco.2018.07.004.

*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. The authors thank Scott Gilpatric, Chris Knittel, seminar participants at the University of Kentucky, the 2016 AERE conference and the 2016 EAERE conference, and two anonymous reviewers for very helpful comments and suggestions. Thanks also go to the Institute for Policy Integrity at NYU, Michael Livermore and Howard Kunreuther for help designing and fielding the survey. This research was funded in part through a grant from the Alfred P. Sloan Foundation.

Alternative Value Elicitation Formats in Contingent Valuation: Mechanism Design and Convergent Validity

Abstract: To date, much of the theoretical work on the incentive properties of contingent valuation surveys has focused on the oft-recommended single binary choice (SBC), referendum format. This work has identified conditions under which an SBC elicitation is incentive compatible, and empirical evidence lends support to the predictive power of the theory. Nevertheless, researchers and practitioners commonly use alternative elicitation formats, and defend their design choices based on efficiency or other criteria. In this study, we demonstrate that it is possible to identify conditions under which alternative elicitation formats are incentive compatible, using as examples open ended (OE) and payment card (PC) question formats. 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 in mean willingness-to-pay 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 over methods 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. This guidance was reaffirmed recently by Johnston et al. (2017), who provide best practice recommendations for stated preference studies used to inform public decision making. Nevertheless, many alternatives, such as open-ended (OE) questions, are used in practice.[2] Researchers routinely adopt alternative formats, as they can reduce complications associated with experimental design (e.g. bid design) and increase the power of the experimental design (e.g. by asking about multiple goods in the same survey and/or by eliciting more precise information on preferences). However, alternative value questions are argued to be more complex and unfamiliar to respondents, and are hypothesized to give rise to strategic, untruthful responses.

A handful of recent papers have used mechanism design theory to establish conditions under which an SBC elicitation is incentive compatible (Carson and Groves, 2007; Carson, Groves and List, 2014; Vossler, Doyon and Rondeau, 2012; Vossler and Evans, 2009). This theory work has motivated refinements in survey design, such as emphasizing that surveys are inputs to public decision-making in order to incentivize responses. Moreover, since incentive compatibility pivots on unobserved beliefs, surveys now routinely include questions to measure

[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.

these beliefs. That theory is important to contingent valuation is further emphasized by empirical evidence. For instance, enhancing beliefs over policy consequences has been shown to increase construct validity (Herriges et al., 2010), predictions from theory are supported by controlled experiments (Carson et al., 2014), and evidence from field tests suggests that external validity can pivot on whether theoretical assumptions appear to hold (Vossler and Watson, 2013). Johnston et al. (2017) strongly recommend the use of incentive-compatible response formats, and indicate that the "most straightforward means" to accomplish this is with a SBC question.

In this study, we demonstrate that it is possible to identify conditions under which alternative elicitation formats are incentive compatible, using as examples OE and payment card (PC) question formats. In doing so, following in the footsteps of prior research on the SBC format, we hope to identify ways to improve survey design as well as enhance the validity of alternative formats. 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.

Alternative formats face at least 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. This is not only true for an OE elicitation, but other formats that present respondents with more than one possible cost. 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 for alternative

elicitation formats can give rise to untruthful responses. For example, as discussed by Carson and Groves (2007), respondents to an OE question 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. 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 continuous responses (bids) 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 no votes. Our designs further include a coercive payment vehicle and frame the elicitation as an advisory referendum. Some prior studies utilizing PC or OE elicitations 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 propose matter empirically, we include a more standard OE elicitation. Interestingly, we find that mean WTP from the SBC treatment is 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, it is possible that individuals form values based on the list of possible payment amounts included on the PC. Further, the PC elicitation produces an interval-censored signal of willingness-to-pay (WTP), which gives rise to speculation over how responses will be interpreted should the (exogenous) cost fall within this interval; as a result, the assumptions for incentive compatibility are stronger relative to OE, and may have been violated in practice.

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 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 OE and PC questions, and identify a set of conditions that as a group are sufficient to establish the incentive compatibility of the elicitations. 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 are as follows:[3, 4]

(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, which relates to the implementation of a single possible policy that is identical to the 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 vote, and requires that the proposal described in the survey must only influence whether an identically defined policy is implemented. As a possible violation of this, as supported by some empirical evidence and analyzed theoretically by Flores and Strong (2007), respondents may perceive that the actual cost to them would be higher or lower than the stated cost. Condition (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

[3] These conditions are identical to those in Vossler, Doyon and Rondeau (2012), with the exception of (iii) which is expanded upon for clarity.

[4] For our purpose, the use of the term "voter" here refers to a respondent who participants in the advisory survey referendum.

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".

A. Incentive compatibility of open ended questions

The BDM mechanism and RPVM are continuous, revealed preference response formats that are incentive compatible under expected utility. For either mechanism, the actual cost of the good is based on an exogenous process, implemented after respondents submit their bid for the good. For the BDM, if one submits a bid that is less than or equal to the randomly drawn cost, she purchases the good at this cost; otherwise, no purchase is made. For the RPVM, bids from a group of voters are converted into yes and no votes (based on whether the bid is higher or lower than the cost), and a majority-vote rule determines provision. In what follows, we apply similar logic to OE, stated preference questions.

Let c denote the cost of providing the public good, as perceived by voter, and let her true valuation be denoted by v. The utility from the policy is then $U(v-c)$. As the voter is not given precise information on cost, to her c is a random variable described by a density function $f(c)$ with positive density over the interval $[c^{min}, c^{max}]$.[5] If the elicitation is policy consequential, then voter i's stated preference, s_i, can influence the probability that a policy is implemented. Let $P(s_i, \boldsymbol{s}_{-i}; G)$ denote the probability that a policy is implemented, with P increasing in s_i. The vector $\boldsymbol{s}_{-i}$ denotes the stated preferences of other voters and G represents nuances of the policy process, including policymaker preferences and decision rules. With the expected utility framework, the voter wishes to maximize:

[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.

(1) $$EU(s_i) = \int_{c^{min}}^{c^{max}} [P(s_i, \boldsymbol{s}_{-i}; G)\, U(v - c) + (1 - P(s_i, \boldsymbol{s}_{-i}; G))U(0)] f(c) dc\,.$$

If a voter perceives her response will influence the particular cost paid upon implementation, in which case $c \equiv c(s_i)$, a tension arises. In particular, as long as $s_i < v$, increasing s_i can increase the chance a desirable policy is implemented (i.e. one where $v > c$) while also increasing the cost paid conditional on the policy being implemented. This tension can easily result in a loss of incentive compatibility, including cases where $s_i = 0$ is an optimal response when $v > 0$.

With cost uncertainty, voters may perceive there to be a set of policies under possible consideration, which correspond to the proposal described in the advisory referendum but that vary in terms of their cost. Following the logic of the BDM and RPVM, incentive compatibility is possible if the voter believes that the cost is *exogenous* to her stated valuation. Formally, we provide an alternative to condition (iii):

(iii´) the open-ended response s is interpreted as a vote on a single policy, 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 policies uniquely defined by their cost. *Ex post*, the OE response is a single yes or no vote for one policy defined by the exogenously determined cost. Condition (iii´) is unlikely to hold in standard OE applications where little or no information is given on the cost of the policy nor how responses would be interpreted. 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. We now provide a proof of the incentive compatibility of OE questions.

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

PROOF:

With OE responses interpreted as yes and no votes for any given cost, we can define the implementation probability as a function of the number of yes votes. Let $P(1_{\{c\leq s_i\}} + \sum_{j\neq i} 1_{\{c\leq s_j\}}; G)$ denote the probability that the policy is implemented, where $1_{\{c\leq s_i\}}$ is an indicator function that equals 1 when $c \leq s_i$. Then, we can write $P_1(1 + \sum_{j=i} 1_{\{c<s_j\}}; G)$ for $c \leq s_i$ and $P_0(\sum_{j\neq i} 1_{\{c\leq s_j\}}; G)$ for $c > s_i$. It follows from condition (iv) that $P_1 \geq P_0$. We can then write the expected utility of the voter as:

$$(2) \quad EU(s_i) = 1_{\{c\leq s_i\}} \int_{c^{min}}^{c^{max}} \left[P_1\big(U(v-c) + U(0)\big) + U(0)\right] f(c)dc$$
$$+ (1 - 1_{\{c\leq s_i\}}) \int_{c^{min}}^{c^{max}} \left[P_0\big(U(v-c) + U(0)\big) + U(0)\right] f(c)dc.$$

With expected utility written in this form, it is straightforward to see that the stated preference s_i determines the range of costs for which the voter increases or decreases the implementation probability.

The proof proceeds by considering deviations from the truthful strategy $s_i = v$. Consider first a strategy $s_i < v$. From (2), expected utility is increasing in s_i over the interval $[c^{min}, v]$. This is because an increase in s_i increases the probability that a desirable policy, one with $U(v - c) \geq 0$, is implemented. Therefore, the deviation decreases expected utility. Consider instead a strategy $s_i > v$. Expected utility is decreasing in s_i over the interval $(v, c^{max}]$. An increase in s_i beyond v increases the probability that an undesirable policy, where $U(v - c) < 0$, is implemented. Therefore, this deviation decreases expected utility. As long as $c^{min} < v < c^{max}$, it follows that deviating from the strategy $s_i = v$ decreases expected utility. Therefore $s_i = v$ 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 $f(v) = 0$, 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 $[v, c^{min}]$.

B. Incentive compatibility of payment card questions

A PC lists a set of possible cost amounts and, depending on the form of the PC, the respondent is asked to indicate (yes or no) whether she is willing to pay each cost or instead asked to circle the highest amount she would be willing to pay. Regardless of the form of the PC, one can interpret the response(s) as a set of yes or no votes on possible policies characterized by the non-cost attributes described in the survey and uniquely defined by cost. Let $c_1, c_2, \ldots, c_N$ denote listed PC cost amounts, rank-ordered from lowest to highest. Further, let the stated

preference $s = c_n$ denote that the voter indicates a no vote to $c_{n+1}, c_{n+2}, \dots, c_N$ and a yes vote to $c_1, c_2, \dots, c_n$. Given the voter's true valuation, v, denote as c_v the cost which bounds from below the true valuation such that $c_v \leq v < c_{v+1}$. For a PC mechanism to be incentive compatible, it must incentivize the voter to indicate $s = c_v$.

For the stated preference $s = c_n$, it is reasonable to believe that any cost for which $c \geq c_{n+1}$, including those not listed on the PC, will be interpreted as a no vote, and for any cost $c \leq c_n$ this will become a yes vote. This is consistent with the voter's stated preferences. For costs in the interval $c_n < c < c_{n+1}$, a complexity arises that is not present in the OE elicitation. Depending on the density $f(c)$ and how the PC response may be translated into a vote, it is possible for $s = c_v$, $s = c_{v-1}$ or $s = c_{v+1}$ to be an optimal response. In other words, the voter may under or over-reveal demand.

To see this, suppose that the voter who responds truthfully with $s = c_v$ believes that for any cost in the interval (c_v, c_{v+1}), her response will be interpreted as a no vote. The voter should instead indicate $s = c_{v+1}$ if the expected losses over the interval $(v, c_{v+1}]$ are less than the expected gains over the interval (c_v, v). This would occur if v is sufficiently close to c_{v+1}. Otherwise, she should indicate $s = c_v$. At the other extreme, the voter may perceive that for any cost in the interval (c_v, c_{v+1}), her response $s = c_v$ will be interpreted as a yes vote. In the case that v is sufficiently close to c_v, she is better-off voting $s = c_{v-1}$. By doing so, she only votes yes when $c < c_v$ and avoids voting yes for any cost in the interval (v, c_{v+1}) which would yield expected losses. Otherwise, if v is sufficiently close to c_v, she is better-off voting truthfully.

As another possibility, a voter who states $s = c_v$ may speculate that her stated preference will simply be ignored over the interval (c_v, c_{v+1}). This too can result in a loss of incentive compatibility. If, for instance, v is sufficiently close to c_{v+1} the voter may want to instead vote

$s = c_{v+1}$ so that she can help increase the probability a policy defined by a cost in this range is implemented. In doing so, her vote would then be ignored if $c_{v+1} < c < c_{v+2}$. Nevertheless, holding fixed her perceived influence over outcomes across these cost intervals, if the voter perceives that the actual cost of the policy is much more likely to fall in the interval (c_v, c_{v+1}) than in the interval (c_{v+1}, c_{v+2}) she would be willing to make this tradeoff.

Given the complications above, we need to place restrictions on $f(c)$ for a PC elicitation to be incentive compatible. One possibility is to assume $f(c) = 0$ for $c_v < c < c_{v+1}$. This assumption renders moot the undesirable effects of the speculations described above. For costs outside of this interval, the PC response is then interpreted identically to that of an OE response, and the Proposition applies. This assumption may be reasonable if, given the complexity of the decision task, voters do not even speculate over "what if" scenarios in the event that $c_v < c < c_{v+1}$. As another possibility, voters may believe that if the actual cost falls between c_v and c_{v+1}, or any two amounts listed on the PC for that matter, authorities will simply adjust the cost up or down to equal one of the listed amounts prior to interpreting responses. In practice, policymakers may be naturally inclined to avoid fractional costs, and amounts listed on PCs tend to represent focal points such as $10 and $25. If the process by which authorities adjust costs is exogenous to valuations, this in and of itself does not lead to a loss in incentive compatibility.

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 to12 inches of precipitation, 7 to 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. 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 in the high-risk flood zone. 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.

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 emphasizes the "advisory" nature of the survey and that results may or may not be taken into consideration by authorities.[6] 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

[6] To promote beliefs over policy consequentiality, an ideal experiment setting is one where authorities pre-commit to using stated preferences in the decision process in a transparent way. However, this does not reflect the political reality surrounding the vast majority of stated preference studies.

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).

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

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

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 address fully 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 propose are 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?"

As discussed previously, a theoretical complication arises because the PC elicits an interval-censored response. Although we could have, for instance, indicated that only those listed costs are probable, such a statement presumably would not be viewed as credible.

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 delineate officially the 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.

[8] We identify single family households by restricting our sample to parcels with ownership type "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".

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.

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 responses 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 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 20%.

To gain insight on possible differences due to sample self-selection, we make use of the PLUTO database. Table 2 compares means of key variables from the assessors' data on housing characteristics across respondents and non-respondents. Across house size, age, renovation status, renovation year and the assessed value of the lot and total structure, the two groups are nearly identical. Unfortunately, no available data source describes other socio-economic characteristics such as income and education specific to our targeted population.[10] While we

[9] Anonymized versions of the initial letter and the reminder post cards appear in the appendix.
[10] Census data on income, household size and education are not available for only single-family residences and Census tracts do not line up well with the high-risk zone.

cannot rule out that characteristics of respondents differ based on observable or unobservable factors, the survey respondent homes look remarkably like the homes of those who chose not to respond.

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 3 describes information on our respondents, using both assessor's data and survey data, for Experiment 1. Table 4 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.

Mortgage lenders typically require flood insurance on properties in the 100-year flood plain. This protects the value of their collateral in the event of a major flood. In the high-risk flood plain the average premium is over $3,000 a year and prices vary from house to house based on risk (Dixon et al., 2017). Many policies require a surveyor to provide detailed elevation data. Outside the 100-year flood plain, coverage is typically optional even for mortgage holders, relatively low cost and can be priced simply by entering an address and coverage level into an online form.

Table 5 presents the raw proportion of Experiment 1 respondents indicating a yes vote across the possible cost amounts included in the survey.[11] At a cost of fifty cents 68% of PC respondents vote yes. A handful of those respondents then stopped filling out the payment card. Beyond 50 cents, the survival function for the PC is monotonically decreasing, with 69% of respondents voting yes at $1 per month and 6% for $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. Each probability is based on a different random sample, however, and so violating monotonicity is somewhat expected. This is especially true given our bid design, which included many small dollar amounts with little separation between them.[12]

The PC and SBC 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 6 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

[11] For this table and all subsequent ones, we present the raw data including respondents who provided incomplete answers to other questions. The only exceptions are regression specifications that include a number of controls for which there may be missing data for some.

[12] We further note that the empirical survival functions for the SBC data in both experiments are better behaved if the sample is restricted to those indicating consequentiality.

OE, and 19% for theory-driven OE. Using a Fisher exact test, the proportion of zero responses is statistically different.[13] Carson and Groves (2007) argue that, in a handful of cases where incentive-compatibility is lost, theory predicts OE that respondents will 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, the proportions of yes votes are somewhat higher.

Finally, in Tables 7 − 10 we report the distribution of responses to two follow-up survey questions, one that targets beliefs tied to policy consequentiality and the other, to payment consequentiality. 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

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

"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. As a first-cut at data analysis, we estimate survival functions based on the nonparametric maximum likelihood estimator of Turnbull (1976) using the R package "icenReg" (Anderson-Bergman, 2017). The survival function for OE data is identical to the empirical survival function (i.e. the raw data). For the SBC data, the estimator imposes monotonicity smoothing out of the function relative to what is presented in Tables 5 and 6. Using a likelihood ratio test, we soundly reject equality of the PC and SBC distributions in Experiment 1 ($p < 0.01$). For Experiment 2, we fail to reject equality of either OE distribution when compared with SBC, although we reject equality of the two OE distributions ($p < 0.01$).

Nonparametric methods have the clear advantage of avoiding distributional assumptions. However, a rejection does not necessary signal that welfare estimates such as mean and median WTP differ. Nor does a fail to reject, especially since each distribution is censored, and the degree of censoring differs across elicitation formats. To facilitate welfare estimation, we turn to

parametric methods. 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, and 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

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

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. 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.

Table 11 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 asks 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 controlling 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.[14] We demean the socio-demographic and housing

[14] 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 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.

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 12 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, 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 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 the effect of inconsequentially to differ across treatments, and includes other controls. Similar to what we observed in Experiment 1, inconsequentiality 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 inconsequential respondents is higher for the standard OE, this serves as one explanation for the unconditional difference in WTP.

In Model 6, the sociodemographic and housing characteristics added to regression produce results that 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

Contemporary best-practice guidelines for stated preference studies advocate the use of incentive compatible formats (Johnston et al., 2017). To date, much of the theory focus has been on the single binary choice (SBC) format, and alternative mechanisms are often argued to give rise to untruthful responses. In this study, we identified sets of incentive compatibility conditions for two alternative and (near) continuous response formats: purely open-ended (OE) questions and payment cards (PCs). We therefore 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) on discrete choice experiments, 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 stated preference 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 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 framing the value elicitation as an advisory referendum. 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 plausible 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.

The results from the field study highlight that theory is a potentially important tool for understanding empirical results. We fail to reject convergent validity in mean WTP when comparing our single binary choice (SBC) and theory-driven OE elicitations, 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. Moreover, conditional on respondents holding beliefs that the elicitation is consequential, we find that the difference between the "standard" OE format and other treatments goes away. 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. One implication of the latter result is that, if the research goal is to maximize the number of respondents who view the elicitation as consequential, our theory-driven refinements may be useful.

Researchers using OE formats often defend their design choice based on lower sample size requirements. To our surprise, although there is work comparing SBC and double-bounded binary choice (e.g. Kanninen, 1993), we did not find any analytical or Monte Carlo analyses comparing the efficiency of OE and SBC formats. To explore this, we ran a simple Monte Carlo experiment, using the estimated WTP distribution for the SBC treatment in Experiment 2 as the population WTP distribution. We assumed the same SBC bid design, and analyzed the simulated data using an interval regression model. Based on 1000 replications, for SBC sample sizes of 200, 350 and 500, we find that the same level of precision for mean WTP is achieved with OE sample sizes of approximately 125, 250 and 355, respectively. We note that this is just

suggestive evidence of the possible efficiency gains, as the efficiency of the SBC format can depend critically on factors such as bid design.

The results related to convergent validity are not all encouraging, however, as we find that the SBC elicits higher values than our theory-driven PC. Unfortunately, the size of 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, as the discreteness of the response format leaves open the question of how responses are interpreted if the actual cost lies within a respondent's implied WTP interval. In our opinion, however, this explanation is unlikely. Indeed, if we take an extreme view that all respondents under-revealed demand, in particular that they voted truthfully with the exception of incorrectly voting no to the highest cost they would pay, this does little to bridge the observed 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.

The results have important policy implications. Respondents whose homes were damaged in Hurricane Sandy consistently reported higher WTP. If climate change increases the frequency of natural disasters, support for these types of controls is likely to grow. Interestingly, flood insurance coverage had no impact on WTP for residents in the 100-year flood plain, but had a positive impact on the WTP of residents outside the flood plain, comparable in magnitude to the effect of experiencing damage during Sandy. Households inside the 100-year flood plain frequently have flood insurance and might believe that their insurance will offset some of the damages during floods. Flood insurance is optional for households outside the 100-year flood

plain, so households that have purchased coverage there might be those that find the risk of flood particularly salient despite their relatively safe location.

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

Anderson-Bergman, Clifford. 2017. Bayesian regression models for interval-censored data in R. *The R Journal* 9(2), 487-498.

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 A. Kazmierczak. 2012. Are homeowners willing to adapt to and mitigate the effects of climate change? *Climatic Change* 112, 633-654.

Botzen, W.J. Wouter and Jeroen 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. Wouter and Jeroen 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. and Michael D. James. 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.

Dixon, Lloyd, Noreen Clancy, Benjamin M. Miller, Sue Hoegberg, Michael M. Lewis, Bruce Bender, Samara Ebinger, Mel Hodges, Gayle M. Syck, Caroline Nagy, and Scott R. Choquette. "The Cost and Affordability of Flood Insurance in New York City: Economic Impacts of Rising Premiums and Policy Options for One- to Four-Family Homes." Santa Monica, CA: RAND Corporation, 2017.

Flores, Nicholas E. and Aaron Strong. 2007. Cost credibility and the stated preference analysis of public goods. *Resource and Energy Economics* 29(3), 195-205.

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, Joseph, Catherine L. Kling, Chih-Chen Liu and Justin Tobias. 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.

Kanninen, Barbara J. 1993. Optimal experimental design for double-bounded dichotomous choice contingent valuation. *Land Economics* 69(2), 138-146.

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.

Turnbull, Bruce W. 1976. The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society Series B (Methodological)* 38(3), 290-295.

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 participate in the survey. All 13,432 households in the 100-year flood plain were recruited as part of either the pilot or the main wave of the survey.

Table 2. Housing characteristics for survey respondents and non-respondents

	Residential Area	Num. Floors	Year Built	Ever Altered	Year Altered	Land Assessment	Total Assessment
Non-respondents	1648.9 sf.	1.9	1954.9	0.064	2002.2	$13,624.5	$28,156.0
Respondents	1632.7 sf.	1.9	1954.8	0.063	2003.2	$13,718.1	$28,157.3

Notes: Table entries correspond to sample means. Ever Altered is an indicator equal to 1 if the residence was renovated since built, and Year Altered is the average year of alterations conditional on it having been renovated.

Table 3. Summary Statistics by Treatment, Experiment 1

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

Note: Summary statistics by treatment type are for Experiment 1 households that live inside the 100-year flood zone. Flood insurance is an indicator for households indicating 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 stating either that 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. Summary Statistics by Treatment, Experiment 2

	Single binary choice	Theory-driven Open-ended	Standard Open-ended
Income (thousands $)	116	113	110
	(91)	(116)	(82)
Female	0.50	0.46	0.49
	(0.50)	(0.5)	(0.50)
Environmental Org. Member	0.09	0.10	0.08
	(0.3)	(0.3)	(0.30)
Flood Insurance	0.43	0.41	0.45
	(0.50)	(0.49)	(0.50)
Damaged in Hurricane Sandy	0.47	0.50	0.45
	(0.50)	(0.50)	(0.50)
Evacuated in Hurricane Sandy	0.23	0.19	0.20
	(0.42)	(0.39)	(0.40)
Basement	0.71	0.73	0.76
	(0.45)	(0.44)	(0.43)
Extreme Sport Participant	0.16	0.13	0.15
	(0.37)	(0.34)	(0.36)
Inconsequential	0.24	0.32	0.38
	(0.43)	(0.47)	(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 indicating 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 stating either that 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 5
Proportion of Yes Votes by Cost Level, Experiment 1

Cost	Theory-driven payment card (PC)	Single binary choice (SBC)
0.5	0.68	0.68
1	0.69	0.76
2	0.58	0.86
3	0.52	0.49
5	0.49	0.56
7	0.35	0.60
10	0.28	0.49
15	0.16	0.43
20	0.13	0.53
35	0.06	0.43
50	0.05	0.28

Notes: PC treatment is based on *N*=261, and SBC treatment utilizes *N*=535 with between 45 and 51 responses at each cost level.

Table 6
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 7

Experiment 1 Payment Consequentiality

Treatment	Definitely No	Maybe No	Maybe Yes	Definitely 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 8

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 9

Experiment 2 Payment Consequentiality

Treatment	Definitely No	Maybe No	Maybe Yes	Definitely 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 10

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 11
Willingness-to-pay regressions for Experiment 1.

	(1)	(2)	(3)
Theory-driven payment card (PC)	-11.58***	-11.42***	-20.18***
	(3.10)	(3.51)	(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***	20.46***	30.59***
	(2.93)	(3.34)	(4.84)
Standard deviation function (σ)			
Theory-driven PC	-31.38***	-34.41***	-30.34***
	(8.63)	(10.46)	(9.39)
Constant	47.37***	50.56***	45.74***
	(8.49)	(10.34)	(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 12
Willingness-to-pay regressions for Experiment 2.

	(4)	(5)	(6)
Theory-driven open-ended (OE)	1.52	-1.56	0.71
	(3.41)	(3.74)	(4.83)
Standard OE	-15.18**	-13.75*	-0.18
	(6.90)	(7.22)	(6.83)
Inconsequential × Single binary choice (SBC)			-29.31***
			(7.42)
Inconsequential × Theory-driven OE			-30.26***
			(9.13)
Inconsequential × 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***	14.54***	21.31***
	(2.53)	(2.67)	(3.04)
Standard deviation function (σ)			
Theory-driven OE	12.91	18.32	21.10**
	(14.71)	(16.40)	(10.70)
Standard OE	80.14**	63.08	67.95*
	(36.44)	(39.07)	(37.15)
Constant	38.32***	34.94***	28.45***
	(7.27)	(6.74)	(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)