Information scripts and the incentive compatibility of discrete choice experiments

Christian A. Vossler*
Department of Economics, University of Tennessee, Knoxville, TN 37996

David A. Keiser
Department of Resource Economics, University of Massachusetts, Amherst, MA 01003

Catherine L. Kling
Charles H. Dyson School of Applied Economics and Management and Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, NY 14853

Daniel J. Phaneuf
Department of Agricultural and Applied Economics, University of Wisconsin, Madison, WI 53706

The published version of this study is in the *Journal of the Association of Environmental and Resource Economists*. The final publication is available at https://www.journals.uchicago.edu/doi/10.1086/731527.

*Correspondence should be directed to Christian Vossler, Department of Economics, 527F Stokely Management Center, The University of Tennessee, Knoxville, TN 37996. E-mail: cvossler@utk.edu. Telephone: 865-974-1699. Fax: 865-974-4601. We thank Nick Kuminoff and three anonymous referees for their thoughtful feedback and suggestions. We are further grateful for comments provided by Daniel Rondeau, Kerry Smith, and seminar participants at the University of Alberta, University of Exeter, and London School of Economics. Approval was obtained from the Human Research Protection Office at the University of Massachusetts, under protocol 1774. This article was developed under Assistance Agreement No. 83616602 awarded by the U.S. Environmental Protection Agency. It has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication.

Information scripts and the incentive compatibility of discrete choice experiments

Abstract: This study examines whether information scripts can align respondent beliefs with conditions for incentive compatible elicitation in stated preference surveys. We embed tests of these scripts within a discrete choice experiment of large-scale surface water quality improvements in the midwestern US. An independence script, which encourages respondents to treat the multiple valuation scenarios presented to them in isolation, improves construct validity and increases the likelihood that stated beliefs are congruent with multiple, theory-relevant assumptions. There are positive spatial scope effects in willingness to pay for those receiving the script and negative scope effects, which we ascribe to strategic voting, for those that do not. Two payment scripts, hypothesized to enhance adherence to the payment mechanism, have an insignificant effect on stated beliefs. A payment script that implies a household's income and the cost of the policy are linked, however, appears to contribute to the finding of negative scope effects.

Keywords: stated preferences, willingness to pay, incentive compatibility, discrete choice experiment, payment vehicle, water pollution

JEL classifications: C93; H41; Q51

1. Introduction

Welfare estimates from stated preference (SP) surveys are frequently used as inputs to government regulatory and benefit-cost analyses involving nonmarket goods, and it is incumbent upon researchers to continually explore methodological advancements. In the context of valuing public goods, a mechanism design literature has emerged that has identified conditions under which a set of SP value elicitation mechanisms are incentive compatible in the sense that respondents can do no better than to truthfully reveal their preferences. In this study we focus on two potential sources of deviations between respondent beliefs and incentive compatibility conditions and examine information scripts designed to better align beliefs with theory. We frame our pursuits through the following questions. First, in surveys that present respondents with multiple valuation questions, can an information script prompt respondents to treat the questions independently? Second, can information on specific features of the payment vehicle increase perceptions among respondents that they would have to pay the stated amount if the policy were implemented? In the context of a discrete choice experiment (DCE) survey valuing large scale improvements in surface water quality, we embed tests of information scripts to provide insight on these questions. We examine a new information script intended to induce choice set independence and a new payment script devised to improve payment consequentiality. We also test an income-based payment script that has been used in a high-profile context but to our knowledge has not been experimentally evaluated.

A necessary condition for incentive compatible elicitation is that participants perceive that survey responses can influence an agency's actions and that they care about the outcomes of those actions (Carson and Groves 2007). Related to this, several studies have examined how information treatments that emphasize policy consequentiality – the perception of a link between

survey responses and public decision making – influence respondent beliefs and/or willingness to pay (WTP) (Bulte et al. 2005; Herriges et al. 2010; Czajkowski et al. 2017; Drichoutis et al. 2017; Oehlmann and Meyerhoff, 2017; Welling, Zawojska, and Sagebiel 2022). Other conditions identified by theory to be important have received less attention. For instance, incentive compatible elicitation using a DCE requires that participants perceive their response to one valuation scenario has no impact on the chance that a policy alternative considered in a different scenario is implemented (Vossler, Doyon, and Rondeau 2012). Though there is empirical evidence that repeated question formats are prone to strategic (i.e., non-truthful) behavior (e.g., Day et al. 2012; Whitehead, 2002), there is little evidence on whether information provision can mitigate strategic responses. Further, while beliefs over payment consequentiality – the extent to which respondents perceive they would have to pay the stated amount – can be a significant determinant of WTP (e.g., Carson, Groves, and List 2014; Zawojska, Bartczak, and Czajkowski 2019; Börger et al. 2021), there is a paucity of research that explores whether information can foster desired beliefs about a coercive payment vehicle.

Our study focuses on information treatments designed to promote beliefs over choice set independence and payment consequentiality, and in doing so helps fill the research gaps highlighted above.[2] For choice set independence, we introduce an *independence script*, which includes information given prior to each valuation scenario. This script encourages respondents to treat the multiple valuation proposals in isolation. We note that DCEs sometimes include a simple statement asking respondents to treat choices separately, and our baseline survey includes

[2] There is of course a long history of testing the effects of information in SP studies, such as the effects of varying the description of the commodity (Samples, Dixon, and Gowen 1986), providing perspective and relative expenditure information (Bergstrom, Stroll, and Randall 1989), and emphasizing the budget constraint and available substitute goods (Loomis, Gonzalez-Caban, and Gregory 1994). Relative to this broader literature, the distinguishing features of the present study are the development and testing of information scripts motivated by incentive compatibility conditions, along with the measurement of beliefs assumed to be important in theory.

a one-time, one-sentence statement encouraging independence. Our treatment therefore identifies whether an additional repeated and detailed script can address what has heretofore been a potential source of incentive incompatibility in studies using repeated questions.

The independence script also provides a rationale for why we would like respondents to consider proposals in isolation (to interpret their responses correctly), and attempts to dismiss the plausible belief that multiple proposals are being presented to identify which one has the highest support (since this belief would likely lead to strategic voting). The script is similar in concept to the information scripts proposed by Vossler and Holladay (2018), which are designed to diminish strategic responses in the context of open-ended and payment card elicitation formats.

We further test the effects of two payment scripts. The first script is new and references an existing natural resource management law, the Pittman-Robertson Wildlife Restoration Act. Like Pittman-Robertson, our script defines a tax collection mechanism through which funds are earmarked for an environmental policy. We refer to this as the *environmental law script*. The second is a payment script (the *income script*) similar to the one used in the Deepwater Horizon oil spill damage assessment (Bishop et al. 2017), which asked respondents to report their income prior to voting on a policy proposal. This creates the perception of a link between income and the cost share of a policy, which may increase the realism of the payment vehicle. While this script was originally designed around and used in a survey that elicited valuations through a single binary choice question, it is natural to ask whether the script is beneficial in the context of DCEs, which are more commonly employed. The impact of this script has not been tested, though other researchers have begun to use similar scripts in DCE contexts (e.g., Lupi et al. 2023).

We use two approaches to identify the effects of the information scripts. First, our survey includes a "belief inventory" to gauge respondent beliefs as they relate to the incentive

compatibility conditions for a binary DCE (Vossler et al. 2012). The belief inventory provides a measure of independence beliefs, and otherwise extends the increasingly common practice of asking policy and/or payment consequentiality questions. While stated beliefs are at best noisy signals of actual beliefs, they provide a useful diagnostic from which to examine whether or how the information treatments alter beliefs germane to incentive compatibility. Second, we estimate and compare WTP measures to assess construct validity for groups with different treatments and belief structures and to infer instances of strategic voting.

Our results show that the independence script has a dramatic influence on stated beliefs and WTP estimates. Recipients of the script are 22 percentage points more likely to report voting on the scenarios independently. More importantly, we identify construct validity failures for those not receiving the script. Respondents not receiving the script have a lower WTP for improving relatively larger policy areas (a failure of spatial scope) while those receiving the script exhibit intuitive responsiveness to spatial scale. These WTP differences are exacerbated when we more directly allow estimates to differ based on whether independence beliefs hold. We attribute WTP differences to a higher incidence of strategic voting among those not receiving the script (or not holding independence beliefs). That controlling for beliefs (and influencing them) can improve the validity of welfare estimates comports with prior literature (Herriges et al. 2010; Vossler et al. 2012; Vossler and Watson 2013; Vossler and Holladay 2018; Vossler et al. 2023a).

On a practical note, prior DCE surveys provide little information on how responses may be interpreted by authorities, and it is tenuous to assume that most respondents consider choice sets in isolation. Indeed, of the five beliefs we measure, in the absence of the independence script the baseline figures for the independence belief are the lowest. The results suggest that our independence script may provide a simple way to encourage choice independence and other

theoretically desirable beliefs in DCEs. The belief inventory provides additional information with which to assess and analyze beliefs.

We also show that our payment scripts have little effect on responses to the belief inventory items. This includes our payment consequentiality measure, for which the disparities across the environmental law script, income script, and no payment script groups differ by less than 3 percentage points. However, we find that the income script exacerbates the effects of strategic voting across repeated questions and contributes to construct validity failures among those who do not receive the independence script. Taken together, these findings suggest the income script does not improve payment consequentiality in our application and, in the context of DCEs, may encourage non-truthful responses.

2. Theoretical Framework

Our experiment is motivated by the small literature that uses mechanism design and implementation theory to better understand conditions under which survey-based value elicitation mechanisms are incentive compatible (e.g., Carson and Groves 2007; Vossler et al. 2012; Carson et al. 2014; Vossler and Holladay 2018; Vossler et al. 2023a). Directly relevant to the survey used in this study, Vossler et al. (2012) show that an elicitation involving a sequence of binary votes (choice sets), each involving a choice between a proposed policy and the status quo, is dominant-strategy incentive compatible under the following set of conditions:[3]

(i) The survey is consequential in the sense that votes can affect outcomes and utility.

(ii) The authority can enforce payments by voters.

(iii) At most one policy can be implemented.

[3] For ease of exposition, we have reworded some of the assumptions described in Vossler et al. (2012) to avoid more technical language and the need to provide details of their theoretical model.

(iv) The probability that a proposed policy is implemented is weakly monotonically increasing with the proportion of yes votes in favor of it.

(v) A vote on one proposed policy has no influence on the probability a different policy is implemented.

(vi) Votes can only influence the likelihood of implementing the possible outcomes described in the survey (i.e., the policies as indicated, including their cost; and the status quo).

As actual links between survey responses and policy decisions are typically unknown ex ante, the conditions establish a set of beliefs about the mechanism that participants must hold. Condition (i) is necessary for the theoretical analysis of SP mechanisms as it ensures that respondents are incentivized. Condition (ii) states that people perceive that they would have to pay upon policy implementation. Condition (iii) rules out the belief that multiple policies may be implemented, which could create a link between policies if the marginal value for one policy is conditional on others that could also be implemented.

Condition (iv) logically equates a vote in favor of a policy to an increase in the probability that policy would be implemented. Condition (v) is an "independence of choice sets" assumption that rules out strategic voting links across policies. If, for instance, people believe that authorities are trying to identify the policy with the most support, this may incentivize some to only vote for what they see as the most desirable policy, and to vote against other policies even if those other policies would be welfare-enhancing. Finally, condition (vi) rules out the possibility that a vote in favor of one policy influences whether some *other* policy – one not explicitly asked about in the survey – is implemented. For instance, if someone votes as if the actual cost to them is lower than what is described in the survey, this would lead to a loss of

incentive compatibility.

In our application we stress the consequentiality of the survey to participants (related to condition (i)), by stating that the survey is funded by the government, and that the results may be used to inform public policy. The valuation scenarios are framed as "advisory referenda", which may also encourage beliefs about consequentiality and further promotes the idea that a vote in favor of a policy would increase its chance of implementation (condition (iv)). We use a coercive payment mechanism (condition (ii)) and ask respondents to treat each referendum independently (condition (v)). These survey design elements are increasingly present in SP studies involving public goods, and consistent with SP best practice recommendations for studies designed to inform public decision making (Johnston et al. 2017). As the results from an SP study such as ours may be used by authorities in various and unanticipated ways, it is challenging to include information that explicitly encourages beliefs congruent with (iii) or (vi).

3. Information Scripts and Belief Elicitation

In this section we discuss information scripts intended to induce beliefs consistent with the incentive compatibility conditions. We then describe the belief inventory included in the survey to help gauge adherence to the incentive compatibility conditions. While some of the language is specific to our application to surface water quality improvements, the scripts and belief inventory can be easily adapted to other contexts.

3.1. Independence Script

To understand the effects of the independence script, the baseline survey includes the following statements prior to when participants vote on the policy proposals:

> "We will now present you with several proposals and ask you to vote yes or no on each of them. We are presenting you with multiple proposals because many policy options are available. **Please consider each proposal separately from the others**.
>
> Voting results will be shared with public authorities, and these authorities may consider this information when determining future environmental policy decisions."

Importantly, the baseline survey requests (in bold typeface) that respondents consider the proposals independently, consistent with condition (v). DCE surveys commonly include language directing respondents to treat valuation questions separately, though this practice is far from universal.[4] While understanding whether a simple independence statement like this improves preference elicitation relative to no statement would be interesting on its own, our experiment is designed to measure the effects of providing information that is additional to what is frequently done. To this end, the independence script has two components. First, we include the following passage immediately after the quoted passage above:

> "However, it is not the purpose of the survey to learn which proposal is preferred by the most people. Instead, should policymakers decide to seriously consider one of the proposals, the study results will let them know the percentage of people in favor of that specific policy over no policy at all."

The independence script includes the following narrative prior to every valuation scenario:

[4] Specifically, starting with a set of 100 studies employing DCEs that were published between 2018 and 2023 in *JAERE*, the *Journal of Environmental Economics and Management*, *Environmental and Resource Economics*, *Resource and Energy Economics*, and *Land Economics*, we identified 21 studies that provided complete information on the valuation scenarios. Of these, 12 studies (57 percent) included text to encourage respondents to treat the valuation questions independently.

"You will now be asked to vote on a specific policy. For us to interpret your votes correctly, we ask that you:

- Consider all characteristics of the proposal we present to you (impacted region, changes in water quality, cost to your household).
- Vote on the proposal based only on the information provided within that proposal (and not information in other proposals).
- Vote as if the proposal is the only one being considered.

If you follow these directions, we will be able to know how you would vote on this specific proposal if it appeared in a future election."

We hypothesize that the script engenders beliefs that are consistent with the independence of choice sets (condition (v)). The passages provide additional emphasis on considering the proposals separately and a rationale for why people are being asked to do so. This inducement is reinforced by dismissing a belief that may otherwise arise: that the authorities are asking about multiple scenarios to figure out which scenario or scenarios have the most support. The analysis of DCE data is predicated on the assumption that respondents are maximizing utility by selecting their preferred option within a choice scenario. Further, choices in one scenario are commonly modelled as being independent of those from other scenarios. Thus, the independence script is consistent with how most researchers interpret DCE data.

3.2. Payment Scripts

We consider two payment scripts relative to our baseline survey: a script modeled off the Pittman-Robertson Act, and an income script. Both scripts are presented prior to the valuation tasks. The baseline survey describes the coercive payment mechanism as follows:

> "A new policy would be funded through an increase in your federal, state, and local taxes. The increase would be set up so that households cannot avoid payment or alter the amount they are supposed to pay. The tax increases would last for five years and would end after that time. The funds would be used to maintain improvements even after the tax ends in five years."

The Pittman-Robertson Act style script provides an example of an existing law that involves the collection of taxes to fund natural resource improvements. In this treatment, the following passage is included prior to the last line of the baseline survey passage quoted above:

> "To be sure that the funds are used only for their intended purpose, the funds will be administered like the Pittman-Robertson Federal Aid in Wildlife Restoration Act (16 U.S. C. 669-669i; 50 Stat. 917) which provides funding for wildlife habitat and legislatively dictates that the funds cannot be diverted to other uses."

This official-sounding language is likely to enhance the credibility of the value elicitation exercise, which may increase beliefs that the elicitation is consequential (in a general sense) and that payment can be compelled if a policy were implemented (condition (ii)). Referencing a statute where the money is earmarked may also head off protest responses from those who would otherwise believe the government can collect the money but would spend it elsewhere. If the script is successful at inducing desirable beliefs, the impacts on WTP are unclear. On one hand, if the script enhances beliefs that respondents would have to pay, this should decrease WTP. On the other, a seemingly more trustworthy policy should increase the demand for it.

We also test the effects of an income script that creates a potential link between the respondent's household income and the cost of a policy to the household. A similar script was

included in the SP survey used to estimate damages associated with the Deepwater Horizon oil spill (Bishop et al. 2017). While the original application elicited values using a single binary choice question, which circumvents some of the incentive issues associated with asking multiple valuation questions, other researchers have begun to use similar language when describing the payment vehicle in DCEs. We are unaware of prior research studying the effects of such a script on stated preferences in a repeated choice context. Following the quoted passage included in the baseline survey, the income script adds:

> "Your annual payment would be determined by the pre-tax income for your household. So that we can provide accurate cost information when presenting the policy proposals, we need to ask about your household income at this time."

This passage was followed by a standard household income question. While we obtained demographic information on all respondents outside the survey, we nevertheless asked those receiving this treatment to indicate their household income for the prior year.

3.3. Belief Elicitation

It has long been standard practice in SP surveys to include debriefing questions to identify those who perceived the elicitation differently from what the researchers had intended. Motivated by theory, and popularized by Herriges et al. (2010), some researchers ask respondents whether they hold beliefs consistent with the incentive compatibility conditions. These questions are often labelled as "payment consequentiality" and "policy consequentiality" questions. The former is intended to gauge whether condition (ii) holds, and the latter is designed to understand adherence to condition (iv). An affirmative answer to either question implies condition (i) as it suggests that the participant perceives that her votes have consequences that

she cares about.

We developed a set of measures to take inventory of stated beliefs as they relate to the incentive compatibility conditions. The belief inventory, presented as Figure 1, asks respondents to select between options of "disagree", "neutral", and "agree" for each of five statements. The first statement may be considered as a "payment consequentiality" measure, and, together with the second and third statements, provide a gauge of whether respondents are taking the valuation scenarios at face value (tied to conditions (i), (ii) and (vi)). The third statement gives rise to a "policy consequentiality" measure (related to conditions (i) and (iv)). Responses to the fourth statement indicate whether people considered the votes independently (condition (v)). The final statement functions as a revealed preference benchmark, in that it indicates whether people voted in the survey as they would have had they faced a similar proposal on an election ballot.

While our belief inventory extends prior work on belief elicitation – our attempt to gauge choice set independence and revealed preference equivalence are new – it does not capture all deviations from the incentive compatibility conditions. The belief inventory could be extended to ask participants if they perceived multiple policies could be implemented simultaneously, and whether they believed the authorities might consider other policies not described in the survey. It is important to acknowledge that stated beliefs are not definitive measures of actual beliefs. Even if people respond truthfully to belief questions, elicited beliefs are subject to measurement error and related endogeneity concerns. Furthermore, some respondents may not hold a particular belief before being asked about it, and in this case the usefulness of the elicited belief indicator is less clear.

4. Survey Description and Design

The data used for this study were collected through a survey designed to elicit valuations for changes in surface water quality, as measured by the Biological Condition Gradient (BCG), in the Upper Mississippi, Ohio, and Tennessee River Basins. Figure 2 shows the study area. A description of the survey, methods for calculating BCG scores, and welfare estimates derived from the data (unconditional on information treatments) are presented in Vossler et al. (2023b). The earlier paper does not include analysis of the information treatments nor the elicited belief measures that are the focus here. In the following subsections we provide an overview of the survey and data. A representative survey instrument is included as online supplemental material.

4.1. Defining the Commodity

The BCG (Davies and Jackson, 2006; US EPA 2016) is an index of water quality that is grounded in ecological principles and is generalizable and transferable across space. It can capture traditional use mechanisms (boatable, fishable, swimmable) as well as nonuse mechanisms related to ecological integrity and biodiversity. The BCG consists of six levels, each associated with a different degree of departure from baseline ecosystem function and integrity. The degree of biological degradation represents the consequences of multiple co-occurring stressors such as nutrient pollutants, pesticides, sedimentation, and other physiochemical changes arising from human impacts.

To convey to survey respondents the ecological concepts underlying the BCG levels, we worked with a graphic artist to develop visual representations. These are shown in Figure 3. For each level, the upper panel provides a stylized visualization of water quality conditions, as defined by physical features of streams and rivers.[5] The bottom panel provides a depiction of

[5] We considered: water color/clarity; river channel shape (natural versus channelized); flow conditions (diverse riffles and pools versus homogenous flows); riparian condition (diversity/abundance of streambank vegetation),

biological diversity, with images of representative species that could be supported by the referenced biological condition.

The righthand border of the graphic displays three human use categories consistent with the traditional water quality ladder (Carson and Mitchell 1993), with the addition of a wading category to differentiate full and partial contact uses. A red circle and slash through the use graphic means that use is not supported. The graphics and the associated, nontechnical survey narrative corresponding to the six BCG levels define the water quality commodity.

As part of the broader project, taxonomic data and biological index score data were compiled for over 19,000 monitoring sites sampled over the last 20 years across the study region. The data are averaged across monitoring sites to create a BCG score at the 8-digit hydrologic unit code (HUC8) level, as defined by the US Geological Survey. There are 268 HUC8s in the study region, which we refer to as "sub-watersheds." We denote 4-digit HUCs (HUC4s) as "watersheds." Our study area includes 31 watersheds. Current water quality conditions largely consist of BCG Levels 3 (defined in lay terms as "Some Changes Noticeable") and 4 ("Many Changes Noticeable"), which constitute 42 percent and 49 percent of the study area, respectively. The remaining areas include 4 percent in Level 2 ("Close to Natural State") and 5 percent in Level 5 ("Major Degradation").

4.2. Choice Experiment Design

The survey included valuation scenarios designed to estimate the WTP of households for BCG level improvements at different spatial scales. Each valuation scenario is defined by the following attributes: (a) the spatial scale of the policy area; (b) the extent and spatial distribution

bank condition (eroded versus vegetated); and in-stream habitat (e.g., accumulated sediments versus gravel beds, submerged plants, and woody debris).

of water quality, as defined by the BCG level; (c) whether the policy area includes the home sub-watershed; and (d) an increase in household taxes if the policy were implemented. Table 1 shows the range of attribute values used to define specific valuation scenarios. Household cost is presented as an unavoidable tax increase that would be assessed if the referendum passed. Tax amounts are randomly assigned from the amounts in Table 1 and presented as annual for five years.[6,7]

A valuation scenario (see Figures 4 and 5) consists of a color-coded map showing the policy area and quality improvements, a table summarizing the policy area-wide average quality change, the change (if any) to the local (home) sub-watershed, the size of the policy area, household cost, and a vote solicitation framed as an advisory public referendum. Our scenario maps display BCG levels at the sub-watershed level (see Figure 4) and variation in baseline water quality levels stems from differences in actual conditions across our study region. Importantly, the BCG changes listed in Table 1 are therefore relative to baseline conditions.

We use experimental variation to identify how economic welfare changes with the size of the affected area. Our survey presented scenarios in which the water quality improvement was for a single watershed, three contiguous watersheds, and the full study region. To create the middle category, we divided our study area into 10 mutually exclusive, contiguous groupings of three watersheds.[8] To identify the effects of improving water quality, we include four different BCG change scenarios in the design (Table 1). These change scenarios, along with substantial variation in actual (current) conditions, allow identification of preferences for water quality

[6] Table entries reflect tax amounts in effect for 1875 of the 2000 respondents. We adjusted some of the tax amounts after a soft launch of the survey. See Vossler et al. (2023b) for details. Beyond these adjustments, data from the soft launch were similar in quality and therefore retained and in the final dataset.
[7] The decision to use a five-year timeline for payments and realizing the water quality improvements was made to promote realism and to coincide with a related survey the EPA was concurrently working on.
[8] In one case, the grouping is four watersheds.

improvements.

The four water quality change scenarios, along with the three local and two non-local spatial units, give rise to twenty potential voting proposals. To make this more manageable, and to reduce cognitive burden, we arranged the proposals into blocks. Within a block, either the water quality change (holding the spatial unit fixed) or the spatial unit (holding the water quality change fixed) varied. Participants faced two blocks of the same type, and therefore voted on two water quality change scenarios for the five spatial units or four water quality change scenarios for two spatial units. Prior to each block, we disclosed to respondents the attribute that would be held fixed (e.g., we stated that the set of scenarios involved their local watershed), along with the attributes that would vary. Overall, respondents voted on six to 10 proposals.[9]

We asked for the respondent's zip code at the beginning of the survey, which identifies the local sub-watershed, local watershed, and local group of three watersheds. In the policy scenarios we provide local water quality conditions at the sub-watershed level as an additional "attribute" in the scenario design. While the local spatial units are endogenous to the respondent, the non-local spatial units are randomly assigned. Specifically, we randomly assigned each zip code to a watershed different from their home watershed. This watershed, and a three-watershed grouping that contained the randomly assigned non-local watershed, are used to determine voting proposals.

4.3. Survey Development and Overview

The first section of the survey emphasizes the link to EPA funding and its purpose of

[9] Those assigned the varying spatial unit blocks voted on up to 10 proposals, and those assigned the varying water quality scenario blocks voted on up to 8 proposals. Some of the water quality change scenarios are redundant given the baseline water quality conditions in some regions, and as we avoided these in the design, some respondents faced as few as six proposals.

informing public decision making, and asks respondents about their experience with local water quality. The second part discusses human impacts on water quality in streams, rivers, and lakes, and describes and presents graphics defining the water quality scale (BCG) to participants. We also presented baseline water quality maps and asked questions to gauge respondents' understanding of the water quality metric and their ability to use the maps to identify water quality levels at points in space. The third section provides general information on the valuation scenarios (e.g., mechanisms for payment, water quality improvement, and potential links between the survey and policymaking), followed by voting on the scenarios, and ends with voting debriefing questions. The fourth and final part of the survey includes some questions on recreation behavior, and a limited set of demographic questions.

Extensive focus groups and classroom demonstrations, conducted in several locations across the study region, were used to develop the final graphics, maps, and the valuation scenarios. We coded the survey using the Qualtrics survey design platform and set it up to be completed by an online panel. The experimental design and survey functionality were tested using an online convenience sample obtained through MTurk. Once we were confident that the mechanics of the survey were working properly, and that the materials and questions were well understood, we piloted the survey with respondents from nine states within the study region to confirm the full survey functionality and to obtain preliminary results for informing the distribution of tax changes to use in the final survey.

A sample of 2,000 people residing in our study region, as verified by zip codes, successfully completed the survey between October 15 and November 16, 2021. The surveys were collected by Qualtrics in partnership with NORC at the University of Chicago, using

NORC's online probability based AmeriSpeak Panel.[10] Standard demographic information for all respondents was provided directly by Qualtrics, rather than elicited through the survey. Panel members are recruited rather than volunteer or opt-in to the panel, which increases response rates and sample representativeness, and circumvents issues with fraudulent responses. The survey completion rate was 27.9 percent.

Table 2 summarizes the characteristics of the survey sample. Our sample is slightly older and better educated than the population. The probability sample we used is likely superior to an opt-in sample in terms of minimizing the effects of unobservables and fraudulent responses. However, it will not necessarily yield a close match on observables without demographic targeting, which was not possible given the size of the available sample frame. Qualtrics nevertheless provided us with sampling weights that, when applied, lead to only trivial differences between the weighted sample and population in terms of age, race/ethnicity, educational attainment, sex, and census division.

5. Results

5.1. Effects of Information Scripts on Respondent Beliefs

Response distributions for the five belief inventory items are provided in Table 3 and Table 4. The percentage selecting "agree" is reasonably high for all items. The item associated with choice set independence has the lowest fraction in agreement, at 61.6 percent overall, whereas the figure is above 75 percent for the other items. Table 3 compares stated beliefs between those receiving and not receiving the independence script. The largest change is a 22 percentage point increase, from 51 to 73 percent, in the proportion agreeing that they voted on

[10] See https://amerispeak.norc.org/about-amerispeak/Pages/Overview.aspx for additional details on the panel.

the scenarios independently. This is not altogether surprising, as people may feel obliged to indicate their agreement given the multiple messages respondents received. Nevertheless, we emphasize that all respondents were asked to treat the scenarios separately, regardless of treatment, and that our independence script is additional to the baseline statement. Those treated with the independence script are statistically more likely to agree with statements related to policy consequentiality, the achievability of improvement goals, and the revealed preference benchmark. This is likely due to the independence language also implicitly encouraging policy consequentiality. Although the differences are modest at 4 to 5 percentage points, they are presumably meaningful given the baseline figures are already high at 74 to 78 percent in agreement.

Table 4 shows that the payment scripts do not generate any statistical differences across treatment groups for any belief inventory item. We hypothesized that both scripts enhance credibility of the payment mechanism, which is conceptually captured by the belief inventory item related to payment consequences. Nearly 81 percent of the group not receiving a payment script agreed that they voted as if they would have to pay the stated cost, which is already high. As such, even without the script most survey takers appeared to find the payment mechanism credible. Indeed, it is worth pointing out that the independence script alters stated beliefs for all items except for the payment mechanism, even though the independence script does ask respondents to pay close attention to the attribute levels for each scenario, including the cost.

5.2. Estimating Willingness to Pay

To explore how the information scripts affect WTP we estimate discrete choice models using scenario voting responses. Let the utility of respondent i derived from option j in voting

scenario t be given by

$$V_{ijt} = \gamma_i ASC_{ijt} + \boldsymbol{\beta}_i \mathbf{x}_{ijt} + \alpha c_{ijt} + u_{ijt}, \quad (1)$$

where utility V_{ijt} is additively separable in the cost of the option c_{ijt}, water quality attributes $\mathbf{x}_{ijt}$, and an alternative-specific constant (ASC) that equals 1 for the policy alternative and 0 for the status quo. The scalar constant α, and random vector $\boldsymbol{\beta}_i$ and parameter γ_i, are utility function parameters and u_{ijt} is an error term that is independent and identically distributed Type I extreme value.

The valuation scenario data are analyzed using mixed logit models specified in preference space for repeated choices (Revelt and Train 1998). The explanatory variables are listed in Table 5. We allow utility to vary based on the BCG score and the spatial scale and location of the policy area. Specifically, we estimate the marginal utility associated with an increase in the BCG score and allow this to vary freely across different spatial units and local versus non-local policies. For local policies, the model further captures changes in utility associated with changes in the BCG score specific to the sub-watershed (8-digit HUC) where the respondent lives. The included ASC allows for utility differences between any policy (regardless of its configuration) and the no policy option.

All model parameters, except for the one associated with the cost of the policy, which is fixed, are assumed to follow normal distributions.[11] Estimation was carried out via maximum simulated likelihood, using 500 Halton draws. To derive WTP, we calculate the difference in utility between an improvement scenario and current conditions and divide it by the marginal utility of income ($-\alpha$). This calculation is done separately for each respondent and then averaged

[11] The normal distribution assumption admits the possibility of marginal and total WTP distributions that are negative for some people. It is plausible that some respondents may have negative WTP, e.g., due to a disutility associated with any tax increase or from having to pay for improvements that occur sufficiently far away from their home (in either case irrespective of the amount paid).

over the sample, since current conditions vary across space:

$$WTP = N^{-1} \sum_{i=1}^{N} \left[(\mathbf{x}_i^1 \boldsymbol{\beta} + \gamma - \mathbf{x}_i^0 \boldsymbol{\beta}) \frac{1}{-\alpha} \right], \quad (2)$$

where $\mathbf{x}_i^k$ denotes attribute levels for the baseline ($k = 0$) and improved ($k = 1$) conditions, and $\boldsymbol{\beta}$ and γ are mean marginal utilities recovered from the mixed logit estimation. The delta method is used to calculate standard errors for the WTP estimates.[12]

To investigate the effects of the independence scripts, we estimate separate models for: the full sample (Model 1), the subsample that did not receive the script (Model 2), and the subsample treated with the script (Model 3). Likewise, we estimate models separately for those that did not receive a payment script (Model 4), those that were given the environmental law script (Model 5), and the subsample receiving the income script (Model 6). The estimates are reported in Table A1 of the online appendix. The full sample results replicate those reported in Vossler et al. (2023b). For all six models, all mean parameters, except for the ASC, are statistically significant at the one percent level. All parameters tied to BCG changes have negative signs since an increase in the BCG score reflects a decrease in water quality.

5.3. Effects of Information Scripts on WTP

We first discuss findings for the independence script. Table 6 reports WTP estimates associated with a one-level improvement in the BCG score separately for each spatial unit. These estimates reflect what the average household is willing to pay per year, over a period of five years, for the improvement. For the full sample, WTP estimates for the local watershed, group of three local watersheds, and study region scale policies are all around $300. WTP estimates for non-local policies are about half of the comparable local policy. As shown in the last two

[12] As the tables below show, the marginal utility of income in our models is precisely estimated. Given this, the delta method is expected to provide a reasonable approximation for the WTP distribution.

columns of the table, the WTP estimates for both local and non-local policies are insensitive to scope. This is not implausible because the three spatial scales are reasonably large. For instance, the average watershed in the study region is about 25,000 square kilometers. Additional discussion of the likely drivers of the observed scope effects in the full sample, and some comparisons with other studies that value water quality, can be found in Vossler et al. (2023b).

Results from the two subsamples show, however, that the full sample masks treatment heterogeneity. On average those not receiving the independence script have negative sensitivity to spatial scale for local policies whereas the opposite is true for those receiving the script. For the no independence script subsample, mean WTP is $319 for a watershed-level policy and this decreases to $249 for a study-region wide policy. As the quantity of the commodity has unambiguously increased, basic intuition suggests that WTP should be non-decreasing and so there is a clear failure of construct validity for this subsample. For those receiving the independence script, WTP is $308 for a watershed-level policy, and increases to $349 for a region-wide policy. For non-local policies, those not receiving the independence script have a WTP difference of just 28 cents. For those receiving the script, there is a positive and significant scope effect that is equal to $40 on average.

As noted above, outside of the dramatic effect on independence beliefs, the differences in stated beliefs among those who did and did not receive the independence script are modest. This motivates additional analysis of WTP that focuses on independence belief status.[13] While the overall evidence is mixed, there is a general concern in the literature that responses to belief

[13] We also considered indices that incorporate information from all belief inventory items. However, aside from the independence belief item, 91 percent of respondents answered "agree" or "neutral" to all other belief inventory items. As such, indices that aggregate information on all beliefs largely capture differences due to variation in independence beliefs.

questions are subject to measurement error or otherwise should be treated as endogenous in choice models (Börger et al. 2021). As such, we utilize instrumental variable (IV) methods.

Specifically, to explore the effects of independence beliefs on WTP, we estimate an IV probit model. Let y_{it}^* be the latent probability that the utility associated with the policy alternative ($j = 1$) is greater than the status quo ($j = 0$) in choice set *t*. Furthermore, let y_{it} be an indicator that equals 1 if the respondent voted for the policy alternative ($y_{it}^* > 0$) and is otherwise equal to 0 ($y_{it}^* \leq 0$). Given the utility function in (1), we can then write:

$$\text{Prob}(y_{it}^* > 0) = \text{Prob}(V_{it}^1 > V_{it}^0) = \text{Prob}(\gamma_i + \boldsymbol{\beta}_i \Delta \mathbf{x}_{it} + \alpha c_{it}^1 + u_{it}^1 - u_{it}^0 > 0) \quad (3)$$

If we drop the parameter heterogeneity, and assume that $\varepsilon_{it} \equiv u_{it}^1 - u_{it}^0$ follows a standard normal distribution, a binary probit regression of y_{it} on a constant, $\Delta \mathbf{x}_{it}$, and c_{it} can be used to recover the unknown utility parameters.

Let z_{it} be an indicator that equals 1 if the respondent received the independence script. To allow the marginal utility of the attributes $\mathbf{x}_{it}$ to vary with z_{it}, we extend the model to:

$$y_{it}^* = \gamma + \boldsymbol{\beta} \Delta \mathbf{x}_{it} + \boldsymbol{\theta} \Delta \mathbf{x}_{it} z_{it} + \alpha c_{it}^1 + \varepsilon_{it} \quad (4)$$

As there are six water quality variables, this specification includes six potentially endogenous variables. As instruments, we use indicators for whether the respondent received the (randomly assigned) independence script, and interactions between the independence script indicator and the six water quality variables. To facilitate estimation, we assume that the errors in the outcome model, and the errors in the six structural equations defining the endogenous explanatory variables as functions of the exogenous ones, are distributed multivariate standard normal. We estimate the set of equations by conditional maximum likelihood estimation.

Table A2 of the online appendix presents the estimated IV probit model (Model 8), along with a standard (non-IV) probit that excludes the belief interaction variables (Model 7). Table 7

presents WTP estimates derived from the two probit models. The first row of estimates is based on the standard (non-IV) probit. The fact that WTP estimates are very similar to those from Model 1 in Table 6 (which also utilizes the full sample) suggests that altering the error distribution and ignoring parameter heterogeneity have little impact on mean WTP estimates in this application.

The second and third row of WTP estimates in Table 7 are from the IV probit. Conditioning on independence beliefs has a dramatic impact, and the patterns exhibited when comparing subsamples with and without the independence script are further exacerbated. WTP estimates are similar for those with and without independence beliefs for a local watershed policy that would increase water quality by one level. However, increasing the spatial scale to the full study region increases WTP for those with independence beliefs from $328 to $488, which is nearly a 50 percent increase. For those without independence beliefs, the same increase in spatial scale sharply *decreases* the WTP estimate from $285 to just $12, with the latter estimate not statistically different from zero. Similar patterns emerge for non-local policies. The estimated WTP differences between those with and without independence beliefs should be interpreted as "local" treatment effects in the sense that they capture expected changes in WTP for someone for whom the independence script induces independence beliefs. As such, WTP estimates conditional on independence beliefs are unlikely to generalize to the broader population.

Turning to the payment scripts, Table 8 reports WTP estimates for the three payment script subsamples for each spatial unit. The focus again is on the WTP for a one-level across-the-board improvement in the BCG score. The results for the no payment script and environmental law script subsamples are similar, and both have statistically insignificant scope effects for both

local and non-local policies. For the income script we find a negative spatial scope effect for the local policy changes, with a higher magnitude relative to the no independence script subsample (see Table 6). This suggests that, since the subsamples are overlapping, one of the drivers of the negative scope effect in the no independence script subsample is the income script. To explore this, we estimated models and WTP for all subsamples defined by unique treatment variable combinations (see Tables A3 and A4). For respondents that received the income script but not the independence script, the difference in average WTP between a full study region policy and a local watershed policy is large: -$153 ($p$<0.01), which is a 41 percent decrease. The spatial scope effect is likewise sizable and negative for non-local policies at -$34, although the difference is not significant (p=0.33).

5.4. Interpretation of Findings

Following the theory, we posit that strategic voting may at least partially explain the WTP differences associated with information treatments and independence beliefs.[14] For local policies (improvements including the home area), a larger spatial scale implies the respondent is contributing to provision near their home and further away, and the latter may be less important. For a smaller spatial scale, payment is for more localized provision, which may also be subsidized by non-local contributors. When respondents do not hold independence beliefs, it is plausible that they may strategically vote for smaller policy areas and against larger ones, which serves to increase the perceived likelihood that a more desirable local policy is implemented.

The income script appears to exacerbate strategic voting in this context. We note that the income script implies that tax payments are tied to income. As evidence, our survey included the

[14] We acknowledge that alternative interpretations of our results are possible, but we deem the interpretations provided here to be the most plausible.

following question: "When voting on the policy proposals, how did you imagine that the tax increase would be distributed across households?" The response distributions varied significantly between those receiving the income script and the rest of the sample ($p < 0.01$). Forty-six percent of respondents with the income script selected the response option "Households with higher incomes would pay more in taxes than lower income households", which is 22 percent higher than for those not receiving the income script.[15] We speculate that this income effect strengthens the incentives for strategic voting based on spatial scale. For those with higher income, increasing the spatial scope means that not only would they have to pay for more distant changes that would largely benefit others, but they would have to pay a disproportionate share of the bill. The fact that those with higher incomes are generally willing to pay more for water quality improvements (see Vossler et al. 2023b) further suggests that we should see a larger effect of strategic voting among those with high incomes.

5.5. Supplemental Analysis

We have undertaken several supplemental analyses to confirm the robustness of our primary findings. We have analyzed respondent beliefs using linear regression and ordered probit models to assess interaction effects between the independence and payment scripts (Tables A5 and A6). For each belief inventory item, our specifications include indicators for the information scripts along with interactions between the independence and payment script treatments. Both interaction effects are statistically insignificant in every model, even when tested jointly. A caveat on these results is that our experiment is not powerful enough to identify anything but large interaction effects. Nevertheless, the point estimates of the interaction effects

[15] The other response options are: "No households would have to pay more in taxes"; "All households would pay about the same amount in taxes; and "I did not think about this when voting".

are small (a few percent at most).

Results from a range of alternative models of voting choices generally support the conclusions reported previously: evidence of negative spatial scope effects for those without the independence script, and evidence of positive scope effects for those with the script. Tables A7-A12 show parameter estimates and WTP estimates for conditional logit models, mixed logit models with correlated random parameters, and mixed logit models with ASCs that vary with the spatial scale of the policy. Tables A13 and A14 apply post-stratification weights.[16] Tables A15-A17 present WTP estimates from our baseline mixed logit models for additional BCG change scenarios: improving level 3 areas to level 2, minimum level 2, and minimum level 3.[17] Finally, we considered alternative IV estimators, including a GMM estimator for the IV-probit, and a two-stage least squares (2SLS) estimator that models voting choices as a linear probability model. Both alternative estimators yield similar WTP estimates and conclusions regarding the effects of independence beliefs (see Tables A18 and A19).

We have also examined mixed logit models that test the robustness of our findings to order effects. Tables A20 and A21 show parameter and WTP estimates for models that only use choice tasks 1 and 2 and tables A22 and A23 show estimates for models that only use choice tasks 7 and 8. For the early choice tasks we find negative local scope effects for the no independence script and income script subsamples. For later choice tasks, we find negative local and nonlocal scope effects for those not receiving the independence script. For those receiving the independence script, we find evidence of positive local scope effects (late choice sets only)

[16] For the independence script subsample, the non-local scope effect is statistically insignificant.
[17] For the level 3 to level 2 scenario, the non-local scope effect for the independence script sample is statistically insignificant.

and positive nonlocal scope effects (early choice sets only).[18] We do find that WTP is lower for the later choice tasks. Utility parameters, as well as WTP measures, are generally more precise when based on the later choice sets. This indicates that the decrease in WTP is unlikely to be an artifact of respondent fatigue, and instead arises due to preference learning or a higher incidence of strategic voting.

Next, we explored latent class models (LCMs) as an alternative approach to modelling preference heterogeneity.[19] Based on an LCM with five classes, and splitting the data into subsamples based on whether the respondent received the independence script, we obtain results that generally comport with those from the mixed logit analysis. For the case of no independence script (tables A24 and A25), there are two classes with estimated utility functions that display a statistically significant and negative sensitivity to an increase in local spatial scope. Their combined membership class probability is 61 percent. For the subsample with the independence script (tables A26 and A27), we find that one class (membership probability of 39 percent) exhibits positive local and nonlocal spatial scope effects. Among the other classes, there is no statistical evidence of negative spatial scope effects. These conclusions are robust to models with 2, 3, or 4 latent classes.

The LCMs also provide suggestive evidence that the independence script may reduce attribute non-attendance. For the model based on the no information script subsample, there are two latent classes with statistically insignificant coefficients on the cost parameter. For one class, nearly all utility parameters are statistically insignificant. The two classes have a combined

[18] The fact that spatial scope effects are not always significant in these models is likely due to the smaller sample sizes.

[19] Applying LCMs to our data introduces identification challenges, as individual-level preference information is somewhat sparse. Participants answer binary choice questions, and face a set of scenarios that conceptually identify only a subset of utility parameters.

membership probability of 27 percent. For the model based on the subsample receiving the script, the cost parameter, and most other utility parameters, are statistically significant in all classes.[20]

Finally, we explored the effects of socio-economic characteristics (defined in Table 2) on WTP and beliefs. First, interacting these variables with the ASC in the mixed logit models reveals that WTP generally increases with income, decreases with age, and is higher for those with a bachelor's degree. Second, based on linear regressions with elicited beliefs as the dependent variable, those with a bachelor's degree and higher income are more likely to hold any given belief, and racial/ethnic minorities are less likely to hold any given belief. Third, we explored whether the efficacy of information scripts on beliefs depends on socio-economic characteristics. While this analysis reveals scattered statistically significant relationships, no systematic patterns emerge.

6. Discussion

Information treatments can be used in stated preference (SP) studies to influence and better understand how respondents perceive and interact with the survey. This study examines how targeted information treatments can alter respondent engagement with the value elicitation mechanism in ways that align with or against incentive compatibility. Specifically, we show that a new independence script can shift stated beliefs in desirable ways, increase the probability that respondents treat repeated choice questions as independent, and improve the construct validity of

[20] As additional evidence related to attribute non-attendance, we asked survey respondents to indicate the extent to which each attribute influenced their votes. The degree of stated attribute non-attendance is low. The percentage selecting that an attribute has "little or no effect" on decisions is as follows: 11.6 percent (cost); 24.6 percent (spatial scope); 20.95 percent (local versus non-local); 6.85 percent (water quality level). Although the differences are not dramatic, information scripts did statistically impact stated non-attendance to the water quality attribute. Both the independence script and the environmental law script decrease non-attendance while the income script increases non-attendance.

willingness to pay (WTP) estimates. We also show that efforts to increase payment vehicle realism using an existing income-dependent payment script may exacerbate strategic voting across choice questions and jeopardize construct validity. Finally, we show that a new payment script referencing the Pittman-Robertson Act had little impact on stated beliefs, WTP, or validity properties of our estimates. Taken together, these findings suggest that valuation surveys that have a high baseline level of policy and payment consequentiality – perhaps owing to government sponsorship and an already-credible payment vehicle – may not benefit from information treatments that seek to influence beliefs that are already well-aligned with incentive compatibility conditions. However, for conditions that align less naturally with baseline beliefs – such as choice set independence – there may be substantial benefits to targeted information.

These findings nevertheless need to be placed into proper perspective. In particular, the effects of the information scripts are presumably closely tied to the case study, the experimental design, the mode-of-administration, and overall information content of the survey. First, we frame the valuation scenarios as advisory referenda and inform respondents that the survey is funded by the EPA and results will be shared with them. This appears to have promoted beliefs among most respondents that the survey could potentially be used to inform policy. It is therefore natural to suspect that details about the scenarios, including payment amounts, would also be viewed as credible. This provides a potential reason why the two payment scripts we tested did not have detectable effects on stated beliefs. In other applications, for instance studies that are not funded by the government or not associated with a prominent public policy issue, similar scripts may have a more pronounced effect.

Second, it is unlikely that the independence script would alter mean WTP in all applications. In fact, as WTP estimates for local watershed-level policies are largely invariant to

the presence/absence of the script (and the same is true conditional on beliefs), we may have uncovered a null effect had our study only focused on this spatial scale. Strategic voting, in particular responding to valuation scenarios in a manner inconsistent with truthful demand revelation, can take various forms depending on the application and characteristics of the experimental design. The development of methods for identifying and correcting strategic behavior, for instance through ancillary survey questions or econometric approaches (e.g., Day et al. 2012), remains an important research need.

Third, our experimental design is relatively simple: a binary choice, a small number of attributes defining the good, and a single-dimension water quality metric. We find that values elicited from the subsample not receiving the independence script indicate a negative spatial scope effect. This is a unique finding in the literature, although most studies investigating spatial scope have examined much smaller spatial areas. We speculate that the simplicity of the design may have made it easier for some respondents to devise strategic response patterns that manifest as negative spatial scope effects.

Last, as with most SP studies, we conducted the survey online. It may be particularly challenging to identify spatial scope effects when respondents view valuation scenarios and maps on their smartphones (43 percent of our sample). It stands to reason that both the magnitude of spatial scope effects, and the effect of information scripts on these spatial scope effects, may have been more pronounced had we conducted surveys by mail or in-person.

These caveats notwithstanding, the benefits of the independence script likely extend to other applications. In surveys where people are asked multiple valuation questions, it is unnatural for respondents to perceive that results for a specific question or choice set will somehow be considered in isolation. In the absence of the independence script, even though the baseline

treatment also asked respondents to treat scenarios separately, just 51 percent agreed that this is what they did. We speculate that if a similar question were asked in other surveys that low percentages would likewise be uncovered. As such, there is potential for our independence script to alter beliefs by providing an alternative explanation on how the results may be used in the policy process.

Turning to the income script, while asking participants about income prior to their voting on policy scenarios has conceptual appeal, our results suggest caution in using this approach in repeated choice contexts, as it may have unintended consequences, especially when not otherwise combined with the independence script. On one hand, we find dramatic results of using this income script on WTP, which implies that the script was salient. On the other hand, in the context of our application, we find that it exacerbated the effects of undesirable strategic voting.

The income script we test originated with Bishop et al. (2017). Their survey included a single binary choice mechanism, which is incentive compatible under relatively weak conditions, and reduces opportunities for strategic voting. Thus, our word of caution mainly applies to using the income script in the context of repeated DCEs. Nevertheless, the applicability of the income script presumably depends on whether it is reasonable to suspect that an actual policy would involve a cost that increases with household income. That said, most SP surveys are silent about whether payment would vary across households. This practice has its own consequences if WTP is sensitive to the payment distribution, and baseline survey beliefs are inconsistent with the potential policies that WTP estimates are being used to inform. Clearly, the use of the income script highlights important methodological issues that are beyond the scope of the current study but worthy of further investigation.

Our earlier paper (Vossler et al. 2023b) focused on methods for operationalizing the Biological Condition Gradient for use in valuation surveys and on generating estimates from a high-quality sample. In that work we provided WTP estimates from the same data set, but without conditioning on the information scripts or belief structures. As replicated here, WTP estimates based on the full survey sample are insensitive to spatial scale as we extend beyond a single watershed. The results reported previously are best thought of as valid for the local single watershed scale, but more defensible estimates of value for larger spatial units and non-local improvements are provided by the subsample that received the independence script. This is important for one of the policy objectives of our research agenda, which focuses on understanding the spatial extent of value generated by surface water quality improvements at points in space. In ongoing research, we are using the insights generated from the information treatments examined here to further investigate issues of spatial scale and extent of the market in surface water quality valuation.

References

Bergstrom, John C., John R. Stoll, and Alan Randall. 1989. Information effects in contingent markets. *American Journal of Agricultural Economics* 71(3): 685-691.

Bishop, Richard C., Kevin J. Boyle, Richard T. Carson, David Chapman, W. Michael Hanemann, Barbara Kanninen, Raymond J. Kopp, et al. 2017. Putting a value on injuries to natural assets: The BP oil spill. *Science* 356: 253-254.

Börger, Tobias, Tenaw G. Abate, Margrethe Aanesen, and Ewa Zawojska. 2021. Payment and policy consequentiality in dichotomous choice contingent valuation: Experimental design effects on self-reported perceptions. *Land Economics* 97(2): 407-424.

Bulte, Erwin, Shelby Gerking, John A. List, and Aart de Zeeuw. 2005. The Effect of Varying the Causes of Environmental Problems on Stated WTP Values: Evidence from a Field Study. *Journal of Environmental Economics and Management* 49(2): 330-342.

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.

Carson, Richard T., and Robert Cameron Mitchell. 1993. The value of clean water: The public's willingness to pay for boatable, fishable, and swimmable quality water. *Water Resources Research* 29(7): 2445-2454.

Czajkowski, Mikołaj, Christian A. Vossler, Wiktor Budziński, Anna Wiśniewska, and Ewa Zawojska. 2017. Addressing Empirical Challenges Related to the Incentive Compatibility of Stated Preferences Methods. *Journal of Economic Behavior and Organization* 142: 47–63.

Davies, Susan P., and Susan K. Jackson. 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications* 16: 1251-1266.

Day, Brett, Ian J. Bateman, Richard T. Carson, Diane Dupont, Jordan J. Louviere, Sanae Morimoto, Riccardo Scarpa, and Paul Wang. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management* 63(1): 73–91.

Drichoutis, Andreas C., Achilleas Vassilopoulos, Jayson L. Lusk, and Rodolfo M. Nayga, Jr. 2017. Consumer preferences for fair labour certification. *European Review of Agricultural Economics* 44(3): 455-474.

Herriges, Joseph, Catherine 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., Kevin J. Boyle, Wiktor (Vic) Adamowicz, Jeff Benett, Roy Brouwer, Trudy Ann Cameron, W. Michael Hanemann, *et al*. 2017. Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists* 4(2): 319-405.

Loomis, John, Armando Gonzalez-Caban, and Robin Gregory. 1994. Do reminders of substitutes and budget constraints influence contingent valuation estimates? *Land Economics* 70(4): 499-506.

Lupi, Frank, Joseph A. Herriges, Hyunjung Kim, and R. Jan Stevenson. 2023. Getting off the ladder: Disentangling water quality indices to enhance the valuation of divergent ecosystem services. *Proceedings of the National Academy of Sciences* 120(18): e2120261120.

Oehlmann, Malte, and Jürgen Meyerhoff. 2017. Stated Preferences Towards Renewable Energy Alternatives in Germany – Do the Consequentiality of the Survey and Trust in Institutions Matter? *Journal of Environmental Economics and Policy* 6(1): 1-16.

Revelt, David, and Kenneth Train. 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics* 80(4): 647-657.

Samples, Karl C., John A. Dixon, and Marcia M. Gowen. 1986. Information disclosure and endangered species valuation. *Land Economics* 62: 306-312.

U.S. Environmental Protection Agency. 2016. *A Practitioner's Guide to the Biological Condition Gradient: A Framework to Describe Incremental Change in Aquatic Ecosystems*. EPA-842-R-16-001, U.S. Environmental Protection Agency, Washington, D.C.

Vossler, Christian A., Stéphane Bergeron, Maurice Doyon, and Daniel Rondeau. 2023a. Revisiting the gap between the willingness to pay and willingness to accept for public goods. *Journal of the Association of Environmental and Resource Economists* 10(2): 413-445.

Vossler, Christian A., Christine L. Dolph, Jacques C. Finlay, David A. Keiser, Catherine L. Kling, and Daniel J. Phaneuf. 2023b. Valuing improvements in the ecological integrity of local and regional waters using the biological condition gradient. *Proceedings of the National Academy of Sciences* 120(18): e2120251119.

Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. 2012. Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics* 4(4): 145-171.

Vossler, Christian A., and J. Scott Holladay. 2018. Alternative value elicitation formats in contingent valuation: Mechanism design and convergent validity. *Journal of Public Economics* 165: 133-145.

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 and Organization* 86: 137-147.

Welling, Malte, Ewa Zawojska, and Julian Sagebiel. 2022. Information, consequentiality and credibility in stated preference surveys: A choice experiment on climate adaptation. *Environmental and Resource Economics* 82: 257-283.

Whitehead, John C. 2002. Incentive incompatibility and starting-point bias in iterative valuation questions. *Land Economics* 78(2): 285–97.

Zawojska, Ewa, Anna Bartczak, and Mikołaj Czajkowski. 2019. Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. *Journal of Environmental Economics and Management* 93: 63-84.

Please think about why you voted the way you did for the proposals. With this in mind, please indicate your level of agreement with the following statements.

	Disagree	Neutral	Agree
I voted as if my household would face the stated costs.	○	○	○
I voted as if the policies would achieve the stated improvements in water quality.	○	○	○
I voted as if the information collected in this survey will be used to inform policy makers.	○	○	○
I voted on each proposal without any consideration of the other proposals.	○	○	○
I am certain that I voted the same way I would if I were voting in a public election.	○	○	○

Figure 1: Belief inventory. A tool for measuring respondent beliefs about the value elicitation mechanism.

Figure 2: Study region including the Upper Mississippi, Ohio, and Tennessee River Basins.

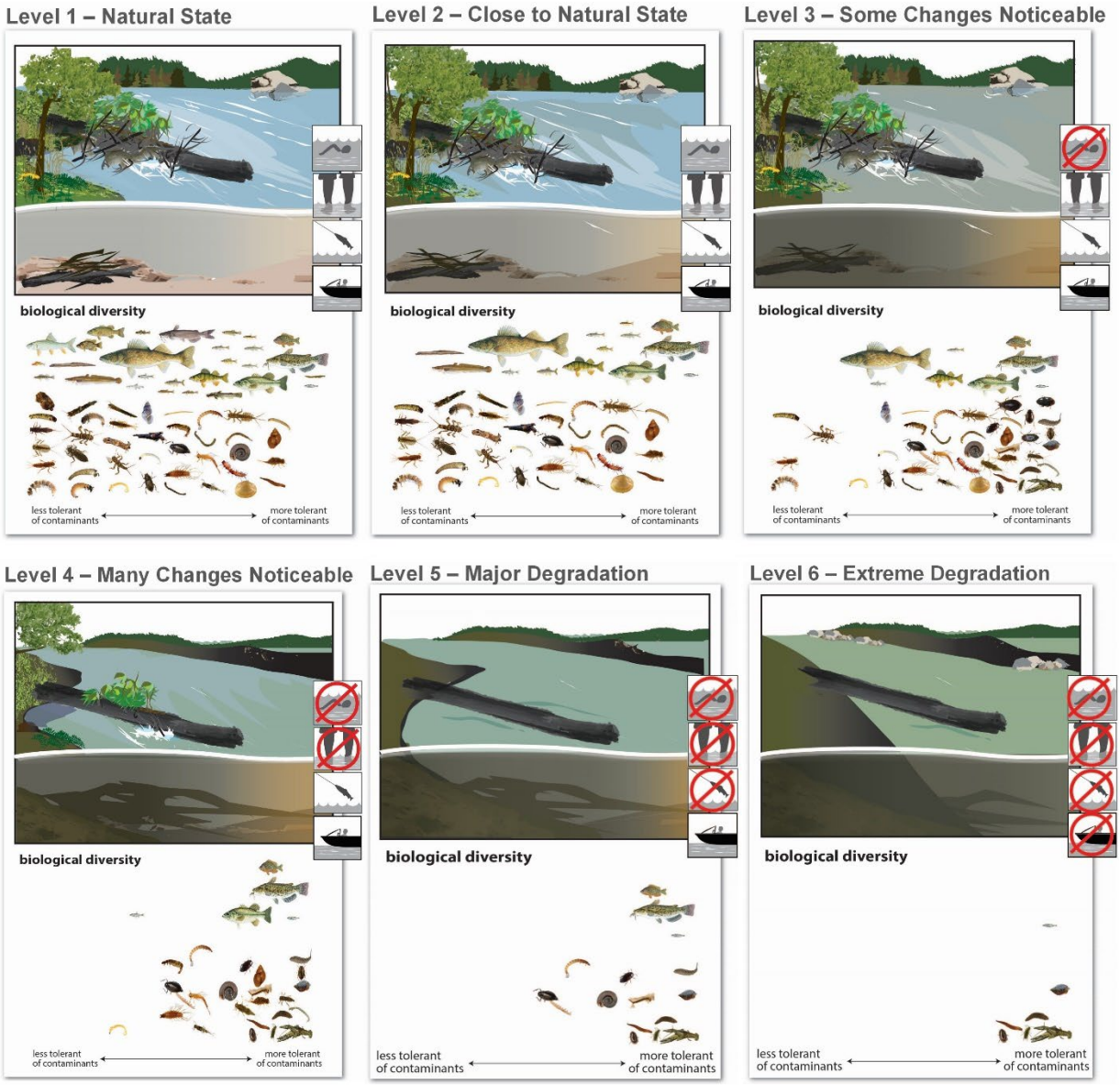
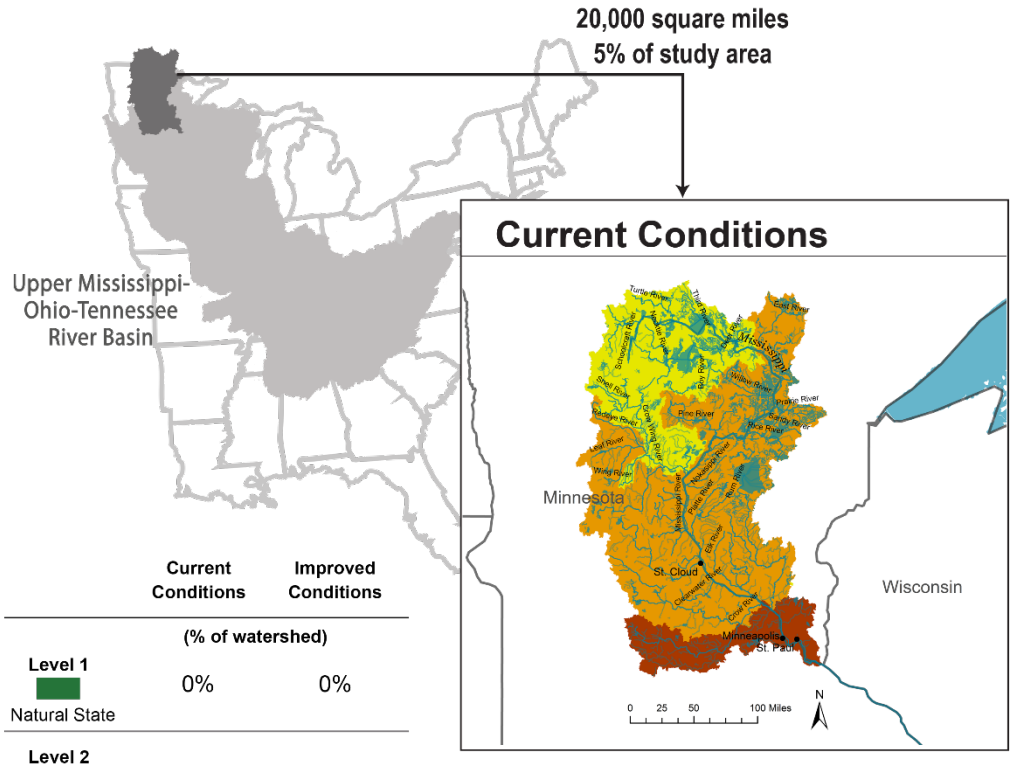


Figure 3. Graphics depicting the six BCG levels, and associated human uses supported, biodiversity, and visual conditions.



	Current Conditions	Improved Conditions
	(% of watershed)	
Level 1 Natural State	0%	0%
Level 2 Close to Natural State	0%	0%
Level 3 Some Changes Noticeable	26%	100%
Level 4 Many Changes Noticeable	63%	0%
Level 5 Major Degradation	11%	0%
Level 6 Extreme Degradation	0%	0%
Average Level	3.85	3.00

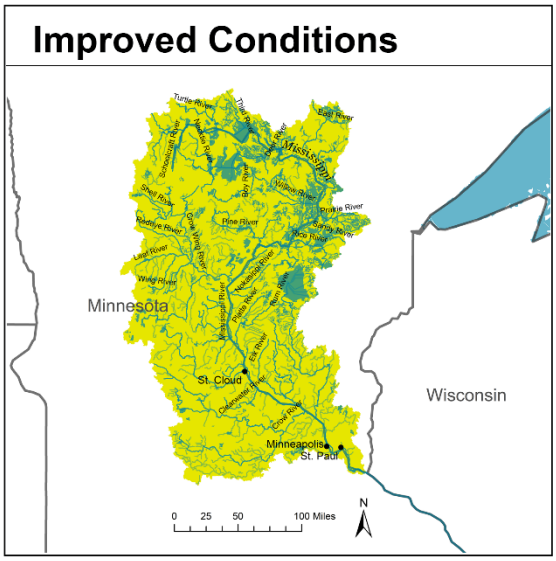


Figure 4. Water quality maps for an example valuation scenario (spatial unit: a single watershed; BCG change scenario: minimum BCG Level 3). The top panel displays current conditions; here, the northern section of the watershed is Level 3, the middle section is Level 4, and the bottom section is Level 5. The bottom panel displays improved conditions; here, the entire watershed is improved to Level 3.

Policy Summary

Description of policy region: Your **local** watershed.
Size of policy region: 20,000 square miles.

	No policy (current conditions)	Proposed policy (improved conditions)
Description of change	None	All areas improve so that **minimum** water quality is Level 3
Water quality near your home	Level 4 – Many Changes Noticeable	Level 3 – Some Changes Noticeable
Water quality throughout region (average)	3.85	3.00
Increase in taxes to your household (per year, for the next 5 years)	None	$50

Advisory Referendum

Should the authorities implement this proposed policy to improve water quality?

○ I vote "No" (against the proposed policy)
○ I vote "Yes" (for the proposed policy)

Figure 5. Tabular summary and voting question for an example valuation scenario (spatial unit: a single watershed; BCG change scenario: minimum BCG Level 3).

Table 1. Valuation scenario attribute levels

Attributes	Levels
Spatial scale	A single watershed (HUC4) Three contiguous watersheds Full study region
BCG change scenario	One-level BCG improvement in all sub-watersheds (HUC8s) Minimum BCG Level 2 Minimum BCG Level 3 Change all BCG Level 3 sub-watersheds to Level 2
Location	Policy area includes home watershed (local) Policy area does not include home watershed (non-local)
Annual tax increase, in effect for five years	$20, $50, $75, $100, $150, $200, $250, $350, $500, $750

Notes: A watershed corresponds with a 4-digit hydrologic unit code address (HUC4), as defined by the US Geological Survey. The full study region includes the Upper Mississippi, Ohio, and Tennessee River Basins (see Figure 2).

Table 2. Selected socio-economic characteristics of respondents

Variable	Description	Mean	Std. Dev.
Female	=1 if female	.529	.499
Age	Age, in years	52.905	16.602
Ethnic or racial minority	=1 if non-White or Hispanic	.18	.385
Bachelor's degree	=1 if has bachelor's degree	.446	.497
Married	=1 if married	.588	.492
Employed	=1 if employed	.577	.494
Retired	=1 if retired	.238	.426
Income	Household income, in $1000s	77.285	56.681
Metro	=1 if lives in metropolitan area	.743	.437
Homeowner	=1 if a homeowner	.747	.435
Household size	Number of people living in household	2.744	1.421

Notes: Summary statistics based on cross-section sample of *N*=2000 respondents.

Table 3. Effects of the independence script on belief inventory items

	Independence script	No independence script	Difference
Voted as if household would pay cost			
Agree	83.57	81.26	2.31
Neutral	13.89	15.98	-2.08
Disagree	2.54	2.76	-0.23
			***p* = 0.396**
Voted as if improvements achievable			
Agree	82.25	78.01	4.24
Neutral	15.31	18.24	-2.93
Disagree	2.43	3.75	-1.31
			p = 0.039
Voted as if survey would inform policy			
Agree	78.30	73.87	4.43
Neutral	18.86	22.49	-3.62
Disagree	2.84	3.65	-0.81
			***p* = 0.064**
Voted on proposals independently			
Agree	72.62	50.89	21.73
Neutral	16.94	27.42	-10.48
Disagree	10.45	21.70	-11.25
			***p* < 0.001**
Voted as they would in a public election			
Agree	80.22	75.54	4.68
Neutral	18.15	22.09	-3.94
Disagree	1.62	2.37	-0.74
			***p* = 0.035**

Notes: *p*-values correspond with a Fisher's exact test of the hypothesis of equal response distributions between the independence script and no independence script subsamples.

Table 4. Effects of the payment scripts on belief inventory items

	Environmental law script	Income script	No payment script
Voted as if household would pay cost			
Agree	83.99	82.26	80.98
Neutral	13.90	15.19	15.75
Disagree	2.11	2.56	3.27
	***p* = 0.253**	***p* = 0.716**	
Voted as if improvements achievable			
Agree	81.42	79.70	79.20
Neutral	16.16	17.44	16.79
Disagree	2.42	2.86	4.01
	***p* = 0.238**	***p* = 0.507**	
Voted as if survey would inform policy			
Agree	76.28	74.74	77.12
Neutral	19.94	22.41	19.76
Disagree	3.78	2.86	3.12
	***p* = 0.795**	***p* = 0.490**	
Voted on proposals independently			
Agree	61.78	61.80	61.22
Neutral	22.21	22.11	22.44
Disagree	16.01	16.09	16.34
	***p* = 0.978**	***p* = 0.978**	
Voted as they would in a public election			
Agree	77.49	78.20	77.86
Neutral	20.85	20.30	19.32
Disagree	1.66	1.50	2.82
	***p* = 0.301**	***p* = 0.249**	

Notes: *p*-values correspond to a Fisher's exact test of the hypothesis of equal response distributions between the indicated payment script subsample and the no payment script subsample.

Table 5. Variables included in baseline choice models

Variable	Definition	Mean (Std. Dev.)
ASC	Alternative specific constant that equals 1 for the proposed policy; =0 for current policy	0.50 (0.50)
BCG for local sub-watershed	BCG score for sub-watershed where the respondent lives; =0 for non-local voting scenarios	1.92 (1.73)
BCG for local watershed	BCG score for the respondent's home watershed; =0 if policy involved a different spatial unit	0.62 (1.27)
BCG for local 3-watershed group	BCG score for a local group of three watersheds; =0 if policy involved a different spatial unit	0.61 (1.27)
BCG for full study area	BCG score for the entire study area; =0 if policy involved a different spatial unit	0.62 (1.26)
BCG for non-local watershed	BCG score for a non-local watershed; =0 if policy involved a different spatial unit	0.60 (1.23)
BCG for non-local 3-watershed group	BCG score for the local group of three watersheds; =0 if policy involved a different spatial unit	0.61 (1.25)
Cost	Cost of the policy, an annual tax payable over five years; =0 for "no policy"	118.55 (194.68)

Notes: Summary statistics are based on the full data sample, where an observation is associated with a single option (policy or status quo) contained in a single choice set presented to a particular respondent (i.e., the total number of observations equals the number of respondents times the number of choice questions per respondent times 2 options).

Table 6. The effects of the independence script on willingness-to-pay

	WTP estimates by spatial unit					Spatial scope tests	
Model/Sample	Local Watershed (HUC4)	3 Local Watersheds (3 HUC4s)	Study Region	Non-Local Watershed (HUC4)	3 Non-Local Watersheds (3 HUC4s)	Local: Study Region = HUC4	Non-Local: 3 HUC4s = HUC4
Model 1: Full sample	$315.63 (13.00)	$302.20 (11.77)	$300.27 (11.69)	$165.20 (10.74)	$186.06 (12.43)	–$15.36 ***p* = 0.273**	$20.85 ***p* = 0.107**
Model 2: No Independence Script	$318.65 (20.32)	$275.42 (16.91)	$248.83 (16.77)	$149.77 (14.98)	$149.49 (18.10)	–$69.81 ***p* < 0.001**	–$0.28 ***p* = 0.988**
Model 3: Independence Script	$308.06 (17.26)	$331.48 (17.16)	$348.75 (16.40)	$179.48 (15.31)	$219.02 (17.22)	$40.69 ***p* = 0.043**	$39.55 ***p* = 0.035**

Notes: Table entries indicate the mean household willingness-to-pay (in 2021 dollars), per year over a period of five years, for a policy that would improve water quality throughout the indicated spatial unit by one BCG level. Cluster-robust standard errors in parentheses. All willingness-to-pay estimates are significantly different from zero at the 1 percent level. A 'local' spatial unit includes the watershed where the household lives, and a 'non-local' spatial unit does not. 'Study Region' refers to the Upper-Mississippi, Ohio, and Tennessee River Basins, and all participants reside in this region. Estimates are derived from mixed logit models reported in Table A1 of the online appendix.

Table 7. The effects of independence beliefs on willingness-to-pay

	WTP estimates by spatial unit					Spatial scope tests	
	Local Watershed (HUC4)	3 Local Watersheds (3 HUC4s)	Study Region	Non-Local Watershed (HUC4)	3 Non-Local Watersheds (3 HUC4s)	Local: Study Region = HUC4	Non-Local: 3 HUC4s = HUC4
Unconditional on beliefs	$308.53*** (12.43)	$298.88*** (12.57)	$302.50*** (12.57)	$171.63*** (12.15)	$190.56*** (12.57)	–$6.03 ***p* = 0.680**	$18.93 ***p* = 0.186**
Without independence belief	$285.34*** (57.77)	$136.55** (68.47)	$12.06 (68.24)	$119.29 (74.19)	-$17.59 (81.09)	–$273.29 ***p* < 0.001**	–$136.88 ***p* = 0.139**
With independence belief	$327.74*** (38.83)	$398.10*** (43.61)	$487.80*** (45.88)	$200.61*** (43.12)	$320.32*** (52.32)	$160.06 ***p* = 0.003**	$119.71 ***p* = 0.040**

Notes: Table entries indicate the mean household willingness-to-pay (in 2021 dollars), per year over a period of five years, for a policy that would improve water quality throughout the indicated spatial unit by one BCG level. Cluster-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. A 'local' spatial unit includes the watershed where the household lives, and a 'non-local' spatial unit does not. 'Study Region' refers to the Upper-Mississippi, Ohio, and Tennessee River Basins, and all participants reside in this region. Estimates are derived from binary probit models reported in Table A2 of the online appendix.

Table 8. The effects of payment scripts on willingness-to-pay

	WTP estimates by spatial unit					Spatial scope tests	
Model/Sample	Local Watershed (HUC4)	3 Local Watersheds (3 HUC4s)	Study Region	Non-Local Watershed (HUC4)	3 Non-Local Watersheds (3 HUC4s)	Local: Study Region = HUC4	Non-Local: 3 HUC4s = HUC4
Model 1: Full Sample	\$315.63 (13.00)	\$302.20 (11.77)	\$300.27 (11.69)	\$165.20 (10.74)	\$186.06 (12.43)	–\$15.36 ***p* = 0.272**	\$20.85 ***p* = 0.107**
Model 4: No Payment Script	\$304.28 (20.34)	\$326.53 (24.70)	\$312.32 (19.80)	\$170.70 (17.57)	\$198.66 (22.76)	\$8.04 ***p* = 0.682**	\$27.95 ***p* = 0.217**
Model 5: Environ. Law Script	\$330.62 (24.90)	\$295.13 (18.55)	\$319.26 (20.96)	\$163.05 (17.67)	\$189.82 (20.28)	–\$11.36 ***p* = 0.686**	\$26.77 ***p* = 0.216**
Model 6: Income Script	\$320.40 (23.83)	\$295.35 (20.87)	\$270.26 (21.49)	\$160.55 (20.52)	\$173.30 (21.56)	–\$50.14 ***p* = 0.053**	\$12.75 ***p* = 0.582**

Notes: Table entries indicate the mean household willingness-to-pay (in 2021 dollars), per year over a period of five years, for a policy that would improve water quality throughout the indicated spatial unit by one BCG level. Cluster-robust standard errors in parentheses. All willingness-to-pay estimates are significantly different from zero at the 1 percent level. A 'local' spatial unit includes the watershed where the household lives, and a 'non-local' spatial unit does not. 'Study Region' refers to the Upper-Mississippi, Ohio, and Tennessee River Basins, and all participants reside in this region. Estimates are derived from mixed logit models reported in Table A1 of the online appendix.