CIRANO
Allier savoir et décision

Revisiting the Gap Between the Willingness-to-Pay and Willingness-to-Accept for Public Goods

CHRISTIAN A. **VOSSLER**
STÉPHANE **BERGERON**
MAURICE **DOYON**
DANIEL **RONDEAU**

2020S-48
CAHIER SCIENTIFIQUE
CS

CIRANO
Knowledge into action

Center for Interuniversity Research and Analysis on Organizations

The purpose of the **Working Papers** is to disseminate the results of research conducted by CIRANO research members in order to solicit exchanges and comments. These reports are written in the style of scientific publications. The ideas and opinions expressed in these documents are solely those of the authors.

Les cahiers de la série scientifique visent à rendre accessibles les résultats des recherches effectuées par des chercheurs membres du CIRANO afin de susciter échanges et commentaires. Ces cahiers sont rédigés dans le style des publications scientifiques et n'engagent que leurs auteurs.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

ISSN 2292-0838 (online version)

Revisiting the Gap Between the Willingness-to-Pay and Willingness-to-Accept for Public Goods

Christian A. Vossler [*], *Stéphane Bergeron* [†], *Maurice Doyon* [‡], *Daniel Rondeau* [§]

Abstract/Résumé

Comparisons of willingness-to-pay (WTP) and willingness-to-accept (WTA) compensation measures have raised concerns over the validity of stated preference methods, and have motivated researchers to predominantly elicit WTP measures even when existing property rights or entitlements would make WTA measures more appropriate. Extending the insight of Plott and Zeiler (2005) to the case of public goods, we argue that past results may in part be driven by experimental design choices, including the use of non-incentive compatible elicitation methods. Using the conservation of wetlands in northern Quebec (Canada) as a case study, we find that WTA/WTP ratios are poorly identified unless estimation procedures control for the beliefs of respondents regarding the consequentiality of their responses. Beliefs over consequentiality are directly tied to sufficiency conditions for the incentive compatibility of stated preference surveys. We find that when respondents express at least "moderate" beliefs over the consequentiality of the survey, resulting WTA/WTP ratios are close to unity.

Keywords/Mots-clés: Stated Preferences, Incentive Compatibility, Willingness-to-Pay, Willingness-to-Accept, Consequentiality

JEL Codes/Codes JEL: C93, D6, Q51

[*] 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 session participants at the Association of Environmental and Resource Economists, European Association of Environmental and Resource Economics, and Society for Cost-Benefit Analysis annual conferences. We gratefully acknowledge funding from the Ministère de l'environnement et de la lutte contre les changements climatiques (MELCC), as well as Fonds de recherche du Québec – Société et culture (FRQSC) for graduate student funding on this project. Special thanks to Monique Poulin for sharing her expertise on wetlands.

[†] Département d'économie agroalimentaire et des sciences de la consommation, Université Laval, Québec City, Québec, Canada.

[‡] Département d'économie agroalimentaire et des sciences de la consommation, Université Laval, Québec City, Québec, Canada and CIRANO.

[§] Department of Economics, University of Victoria, Victoria, British Columbia, Canada and CIRANO.

1. Introduction

There remains much intrigue over the relationship between the willingness-to-accept (WTA) and willingness-to-pay (WTP) for the same good. Standard theory suggests that when substitutes for the good being valued are readily available, the difference between the two measures should be small. However, as Hanneman (1991) suggested, many of the public goods subject to valuation studies do not have good substitutes and in these circumstances, substantial differences between WTA and WTP measures can arise. Experimental evidence supports this theoretical prediction, showing that the difference is generally higher for public than for private goods (Tunçel and Hammitt, 2014).

Yet, there remains considerable debate over the large magnitudes of observed disparities for both public and private goods. In the contingent valuation of public goods, the empirical evidence has been interpreted as an indication that "willingness to pay questions measure preferences but willingness to accept questions do not" (Horowitz and McConnell, 2003, p. 544), or as more general evidence that "answers to contingent valuation questions do not actually reflect stable or well-defined preferences" (Hausman, 2012, p. 47). Our study provides new empirical evidence on the WTA-WTP disparity for public goods. We demonstrate that incorporating recent advances into the design of stated preference studies can reduce the WTA-WTP gap.

There are good reasons to accumulate new data on the welfare gap. From a methodological perspective, as detailed in Johnston et al. (2017), stated preference methods have evolved in meaningful ways over the last few decades. Much of the evidence on the welfare gap comes from older studies, which do not reflect the current state of practice. As we discuss in more detail later, our examination of the stated preference literature reveals that commonly used research methods

raise concerns over the validity of prior WTA and WTP estimates, as well as WTA-WTP disparities.

From a policy perspective, stated preference methods are the only approach for estimating both use and passive use values for changes in public good provision, as well as use values that are beyond the scope of revealed preference data. Extant evidence of the WTA-WTP gap, along with the influential report of the NOAA Panel on Contingent Valuation (Arrow et al., 1993), has led to a heavy reliance on WTP estimation, even when property rights favor estimation of WTA. Lloyd-Smith and Adamowicz (2018) make a similar claim. Moreover, they searched the Environmental Valuation Reference Inventory and found that there are fourteen times more WTP studies in the database. From the purview of Kaldor-Hicks compensation, and based on the stylized fact that WTA exceeds WTP, the currently accepted practice of estimating WTP when the policy setting warrants WTA leads to conservative welfare estimates and the under-provision of public goods. New data generated through improved research designs has the potential to change perceptions over the validity of contingent valuations and, in turn, encourage WTA estimation when dictated by property rights.

Our investigation draws from the insights of Plott and Zeiler (2005), who critically examine the WTA-WTP literature and suggest that the persistent and large welfare gaps observed may be due to "subject misperceptions" tied to the preference elicitation methods employed, rather than a particular theory of preferences. They highlight that many studies employ incomplete controls over misconceptions, and for instance may inadequately explain elicitation procedures or use non-incentive compatible methods. Through an application involving the valuation of private goods

(coffee mugs), they demonstrate that using experimental procedures designed to control for subject misperceptions can lead to meaningful differences in the estimated welfare gap.[1]

The meta-analysis of Tunçel and Hammitt (2014) provides additional evidence that the welfare gap may be sensitive to research methods. They find that the use of incentive-compatible methods and the publication year of the study (which may be a proxy for study quality) statistically decrease WTA/WTP ratios. As the meta-analysis pooled data across all types of goods, including private goods and experimental lotteries over money, it is not clear whether these results hold for specific types of goods.

In this paper, we study the WTA-WTP gap using a contingent valuation survey that values the protection of vast areas of wetlands in northern Quebec (Canada). This region is important for mineral extraction and hydroelectricity production, and further development poses threats to water quality, biodiversity and natural resources. Congruent with recent methodological guidelines (Johnston et al., 2017) and supported by mechanism design theory (Carson and Groves, 2007; Vossler, Doyon, and Rondeau, 2012), we elicit values using a single binary choice (SBC) question, framed as an advisory referendum, with a coercive payment vehicle. WTA and WTP are elicited using a between-subjects experimental design. Further, using follow-up questions, we obtained respondent's perceptions of the consequentiality of the survey. We specifically asked respondents how much they thought their responses would influence the government's decision with respect to implementing the policy presented in the survey. These perceptions of the consequentiality of their

[1] Specifically, with controls over misconceptions in place, the authors find that the welfare gap is not statistically different from zero. Subsequent research utilizing similar experimental procedures reveals that a finding that WTP equals WTA does not generalize to other goods and subject pools (e.g., Isoni, Loomes, and Sugden, 2011; Fehr, Hakimov and Kübler, 2015).

responses are important for assessing the incentive compatibility of the survey's value elicitation methods.[2]

Our results depend critically on respondents' beliefs regarding the consequentiality of their decisions. When respondents think that there is at least a moderate chance that the survey results are consequential, our estimated WTA/WTP ratios are in the 1.0 to 1.2 range across several model specifications. These ratios are considerably lower than those reported in prior contingent valuation studies of large-scale public goods (see Tunçel and Hammitt, 2014). In contrast, when the consequentiality data is omitted, the estimated WTA response function is flat over a large range of values, giving rise to implausibly large and imprecise point estimates and WTA/WTP ratio. Similar results are obtained when welfare estimates are computed only for respondents who perceive the survey unlikely to be consequential. These findings suggest that WTA/WTP ratios elicited from stated preference surveys can be quite different when the elicitation methods meet incentive compatibility conditions.

2. Mechanism design theory and prior WTA-WTP comparisons

In their seminal paper, Carson and Groves (2007) discuss conditions under which a SBC value elicitation question can be incentive compatible. A careful reading of their article identifies four key conditions: (i) the participants care about the outcome; (ii) the authority can compel payments by voters if the good is provided; (iii) the elicitation involves a yes or no vote on a single

[2] Although we highlight here recent developments relevant for value elicitation, it is important to note that methodological advancements extend beyond these, and include refinements in survey development and implementation (e.g., sampling, scenario development, etc.), data analysis, validity assessment and study reporting (Johnston et al., 2017).

project; and (iv) the probability that the proposed project is implemented is weakly monotonically increasing with the proportion of yes votes.

Condition (i) and (ii) require project attributes, including the proposed cost, to enter the respondent's utility function. Condition (iii) specifies a SBC mechanism but its importance comes from ensuring that the elicitation mechanism does not give rise to beliefs that the respondent can influence other outcomes not specified in the policy being evaluated.[3] Condition (iv) implies that the respondent has some opportunity to influence decision making, which in turn means that survey choices can materially influence her future utility. Condition (ii) is sometimes referred to as "payment consequentiality", and (iv) as "policy consequentiality". Vossler, Doyon, and Rondeau (2012) develop a game-theoretic model and, under the assumption that agents are expected-utility maximizers, show that the above four conditions are together sufficient for an elicitation to be incentive compatible. Carson, Groves and List (2014) relax the expected utility assumption and prove incentive compatibility under mixture monotonicity.

Empirical evidence from WTP studies supports the notion that elicitations that correspond with the sufficient conditions enhance both construct and external validity (Carson, Groves, and List 2014, Herriges et al., 2010; Vossler, Doyon, and Rondeau, 2012; Vossler and Watson, 2013). Moreover, accumulated evidence from a large number of studies demonstrate that estimates of WTP vary based on a respondent's beliefs about whether survey results can potentially inform policy (the policy consequentiality of condition iv), and whether they would have to pay (or receive) the amount stated in the survey if the proposed policy is enacted (the payment

[3] For example, Vossler and Holladay (2018) identify (stronger) assumptions, which replace condition (iii) above, under which open-ended and payment card formats are incentive compatible. Nevertheless, they argue that past studies using these formats are unlikely to be incentive compatible, and that considerable information needs to be provided to respondents to induce theoretically-desirable beliefs related to how responses will be interpreted.

consequentiality of condition ii) (see Börger et al., *forthcoming*, and references therein). A lab experiment by Lloyd-Smith and Adamowicz (2018) confirms that incentive compatibility and consequentiality are important for WTA elicitation as well. However, the literature provides few links between WTA elicitation and theory.

Table 1 summarizes the design characteristics of prior stated preference studies that compare WTA and WTP. This table includes 45 studies identified from the meta-analyses of Tunçel and Hammitt (2014) and Koń and Jakubczyk (2019).[4] In our review, we identified specific design choices that may give rise to a loss of incentive compatibility.[5] First, nearly all studies (91%) solely rely on comparisons using formats other than a SBC. Second, in about three-quarters of the studies (73%), the elicitations exclusively involve a private provision mechanism, such as voluntary payment. Third, almost two-thirds of the studies (62%) only compare WTA and WTP based on within-subject experimental designs. Though not recorded in the table, several of the studies using between-subject comparisons nevertheless ask participants multiple valuation questions, which raises similar issues. Below we elaborate on how these design choices may influence stated preferences.

[4] As the starting point of our literature review, we considered all studies referenced in Tunçel and Hammitt (2014) and characterized as valuing "environmental", "health and safety", and "other public or non-market" goods. Of these, seven are excluded. One study was excluded because it only elicits WTP, another one because we could not gain access to it, and five other studies were excluded because they do not use stated preference methods. This is supplemented with more recent studies (those published between 2012 to 2018) included in Koń and Jakubczyk (2019) and characterized as valuing "non-market" goods.

[5] In fact, none of the studies reviewed are immune to these potential design criticisms.

Table 1
Characterization of prior literature

Study	Type of good	Survey mode	Elicitation format	Provision mechanism	Comparison type
Aabø 2005	Other public or non-market	Field	Format 1: DM + OE; Format 2: MBDC + OE	Public	Between
Amigues et al. 2002	Environmental	Field	WTP: SBC or OE; WTA: OE	Private	Between
Bishop, Heberlein, and Kealy 1983	Other public or non-market	Field	SBC + OE	Private	Within
Bowker and MacDonalød 1993	Environmental	Field	PC	Public	Between
Brookshire, Randall, and Stoll 1980	Other public or non-market	Field	Iterative bidding	Private	Within
Brookshire and Coursey 1987	Environmental	Field	Survey 1: PC Survey 2: Smith Auction	Public	Between
Chiwaula et al. 2016	Health and safety	Field	OE	Private	Within
Carthy et al. 1999	Health and safety	Field	Chained approach	Private	Within
Chapman and Johnson 1995	Environmental; Health and safety; Other public or non-market	Lab	OE	Private	Within
Chilton et al. 2012	Health and safety	Lab	OE*	Private	Within
Del Saz-Salazar et al. 2009	Environmental	Field	WTP: SBC + OE; WTA: OE	Public	Within
DuBourg, Jones-Lee, and Loomes 1994	Health and safety	Field	Stage 1: Iterative bidding; Stage 2: PC	Private	Within
Flachaire, Holland, and Shogren 2013	Other public or non-market	Lab	OE*	Public	Between
Garbacz and Thayer 1983	Other public or non-market	Field	OE	Private	Between
Gerking, De Haan, and Schulze 1988	Health and safety	Field	PC	Private	Between
Gleason-Comstock et al. 2017	Health and safety	Field	OE	Private	Within
Griffin and Mjelde 2000	Other public or non-market	Field	OE	Public	Between
Guria et al. 2005	Health and safety	Field	Iterative bidding	Private	Within
Hajek and Stejskal 2015	Other public or non-market	Field	OE	Private	Within
Hartman et al. 1991	Other public or non-market	Field	OE	Private	Within

Huang et al. 2013	Environmental	Field	PC	Private	Within
Kufeoglu and Lehtonen 2015	Other public or non-market	Field	OE	Private	Within
Martin-Fernandez et al. 2017	Health and safety	Field	Double PC	Private	Within
Martin-Fernandez et al. 2013	Health and safety	Field	Double PC	Private	Within
McDaniels 1992	Health and safety	Field	Exp 1: OE; Exp 2: SBC	Private	Exp 1: Within Exp 2: Between
Merkle et al. 2017	Other public or non-market	Field	OE	Private	Between
Nataf and Wallsten 2013	Other public or non-market	Field	OE	Private	Between
Navrud and Mungatana 1994	Environmental	Field	WTA: OE; WTP: OE and PC	Private	Within
Nepal et al. 2018	Environmental	Field	SBC	Private	Between
O'Brien et al. 1998	Health and safety	Field	Iterative bidding	Private	Within
Painter et al. 2002	Other public or non-market	Field	OE	Public	Within
Petrolia and Kim 2011	Environmental	Field	SBC	Public	Between
Rowe et al. 1980	Environmental	Field	Iterative bidding	Public	Within
Schulze et al 1986	Environmental	Field	OE	Private	Between
von Selasinsky et al. 2017	Other public or non-market	Field	OE	Private	Within
Shefrin and Caldwell 2001	Other public or non-market	Lab	OE	Private	Study 1: Between; Study 2: Within
Sun et al. 2005	Other public or non-market	Field	SBC	Private	Within
Tanrivermis 1998	Environmental	Field	OE	Public	Within
Tomohara 2005	Other public or non-market	Field	PC	Private	Within
Van de Berg et al. 2005	Health and safety	Field	Sample 1: SBC + OE; Sample 2: OE	Private	Within
Van Kooten and Schmitz 1992	Other public or non-market	Field	OE	Private	Within
Venkatachalam and Narayanamoorthy 2012	Other public or non-market	Field	OE	Private	Between
Viscusi, Magat, and Huber 1987	Health and safety	Field	OE	Private	Within
Viscusi and Huber 2012	Health and safety	Field	Iterative bidding	Public	Between
Whynes and Sach 2007	Health and safety	Field	Iterative bidding	Private	Within

Notes: *Type of good*: categories as defined in Tunçel and Hammitt (2014). *Survey mode*: all studies except those conducted in an experimental lab labeled as "Field". *Elicitation format*: form of the valuation question(s) used, which includes dissonance minimizing (DM), open ended (OE), open ended with random price mechanism (OE*), multiple-bounded discrete choice (MBDC), payment card (PC), single binary choice (SBC). *Provision mechanism*: "Private" describes settings where payment is voluntary or can otherwise be avoided (e.g., donation, household purchase); "Public" refers to settings where, if the good is provided (or taken away), payment (compensation) is mandatory/coercive. *Comparison type*: "Within" means that both WTA and WTP collected from the same respondents; "Between" means that WTA and WTP collected from different samples. Some studies incompletely describe research methods, and in these cases, we did our best to characterize them accurately. Any errors are unintentional. Complete references are provided in the online supplement to Koń and Jakubczyk (2019).

Most studies rely on open-ended questions, or closely related iterative bidding games, payment cards, or multiple-bounded discrete choice formats designed to elicit a narrow value interval or point estimate from each respondent. The ability to "name your own price" gives rise to the possibility that respondents may strategically over or understate demand in attempt to either influence the amount paid (received) or influence whether a good is provided.[6] In the WTP context, Carson and Groves (2007) argue that open-ended questions (and related formats) lead to underestimates of demand.[7] There are good reasons to suspect that these alternative elicitation formats may not only bias welfare estimates but also increase the WTA-WTP gap. First, consider a case where the respondent perceives she can influence the price paid if a policy is implemented through her stated valuation. In a WTP setting, a strategic respondent should under-reveal demand to lower the price; in contrast, she should inflate the minimum compensation demanded in the WTA setting. Second, consider the case where a respondent believes that the price is fixed but her response can influence whether a policy is implemented. In a WTP frame, the respondent should bid zero when her valuation is less than the expected price as this has the potential to exert the greatest influence. On the other hand, if her WTA exceeds the expected price she should instead demand a very high compensation that exceeds her valuation.

The use of private provision rules can also give rise to a loss of incentive compatibility. Examples include donations towards the provision of a public good, or the purchase of quasi-

[6] There are a few notable exceptions. Brookshire and Coursey (1987) compare WTA and WTP using a Smith Auction, which is an open-ended format. Chilton et al. (2012) and Flachaire, Holland, and Shogren (2013) combine an open-ended question with a random price threshold. These stated preference elicitations resemble incentive-compatible mechanisms that have been used in incentivized lab experiments. In the lab setting substantial training is needed for people to understand these mechanisms (see Plott and Zeiler, 2005).

[7] In general, beliefs can give rise to over- or under-revelation of demand. The stylized fact from the literature, however, supports the notion that WTP elicited from SBC questions exceeds that elicited from open-ended questions or payment cards (Champ and Bishop, 2006).

public goods such as hunting permits, optional car safety features, or medical treatments. Under such provision rules, Carson and Groves (2007) argue that respondents have incentives to over-reveal demand in the WTP setting. For instance, respondents may think that expressing a higher than truthful WTP in a survey deploying a donation mechanism might increase the likelihood that a real fund-raising drive will take place, where they are not bound to make the same payment (and with others being asked to contribute). Similarly, studies designed to value quasi-public goods can fail to be incentive compatible through several channels because they strategically link survey responses to future prices or the availability of products without imposing a commitment to actually purchase the good in the future. Such studies risk confounding, for instance, the actual value of an optional car safety feature, with the desire of respondents to influence the likelihood that the safety option will become available by over-revealing demand, or the price at which it might be offered by under-revealing demand. In a WTA setting, incentives are even murkier. It seems very unlikely that survey results would have much influence on the decision to compensate people for *not* making a quasi-public good available. If these are the respondent's beliefs, payment consequentiality may be entirely lost.

An entirely different methodological criticism of previous studies stems from concerns over within-subject comparisons (see Charness, Gneezy, and Kuhn, 2012). When a survey or experiment contains two value elicitations, the first valuation question creates a reference point and unwanted psychological sources of variation, resulting in the non-independence of tasks. The resulting incentives for respondents to a survey containing both WTP and WTA questions for the same good are complex. Even if both elicitations are incentive compatible when taken in isolation, presenting them in succession raises the possibility that the survey results might not only result in the provision of the good, but also affect property rights. This easily leads to a loss of incentive

compatibility. In deciding whether to vote "yes" for a policy in the WTP setting (presented after an initial WTA question), a respondent who is willing to pay the stated amount may instead vote "no" if she thought that this will increase the chance that the previous (and more desirable) WTA policy might be enacted. As a practical matter, one or both elicitations are likely to be viewed as unrealistic (e.g., hypothetical) as citizens do not usually have a say in whether they should pay for a new policy or instead be compensated if it is foregone.

3. Survey description and design

A. Background

The empirical data for this study was collected from a contingent valuation survey on the preservation of wetlands in Northern Quebec, Canada. The study was mandated by the Government of Quebec to estimate how the population valued a vast zone of remote wetlands situated north of the 49th parallel, and spanning 1.2 million square kilometers or approximately 70% of the province. Despite its size, only 2% (120 000) of the Quebec population inhabits this land, primarily in villages along the north coast of the Gulf of St-Lawrence. The region largely remains in its natural state, although it is home to logging operations and several active mines.

This territory is known for its high potential for hydroelectric power generation and its natural resources including gold, diamonds, titanium, uranium and rare earth elements. In 2009, the Quebec government announced the pursuit of an ambitious program of hydroelectric development projects in the north, with the objective of increasing provincial production by 4,500 megawatts. This was part of a provincial economic stimulus package that would inject 80 billion dollars into the region over the next 25 years. These development projects raised concerns

regarding the integrity of the natural environment, especially with respect to their potential impacts on wetlands and their role in maintaining water quality.

Most of the population of Quebec lives in the southern part of the province, and few travel to the policy region for the purpose of tourism or recreation. However, previous province-wide surveys conducted by private firms found that the Quebec population places a high priority on preserving water quality, including the large quantities of water naturally stored or flowing through northern Quebec. To help structure the survey and valuation exercise, we conducted 52 face-to-face interviews with residents of the village of Havre-St-Pierre. This village is located near the most recently completed hydroelectric project and where significant uranium deposits have been found.

No respondent had specific concerns relating to direct use activities (fishing, berry picking, duck hunting, etc.) since the vast quantities of undeveloped space provides ample substitutes. However, and in line with province-wide opinion polls, serious concerns were expressed over water quality issues. It was thus deemed appropriate to conduct a contingent valuation survey to estimate both the use and passive use values for wetlands in northern Quebec, with a focus on their role in regulating and purifying water flows.

B. Description of survey and data

The development of survey instruments was informed by discussions with experts, focus groups, and substantial pretesting. The policy considered by respondents was the proportion of the northern territory to be legally protected from development. Two conservation targets were used, 35% and 50%. Both are significant increases relative to the status quo of 14%, which is the mandated regulatory level of conservation across the province. The 35% level is advocated by

ecologists as the minimum threshold necessary for maintaining water quality. The 50% level was motivated by the political discourse since the Premier of Quebec had suggested that it was a plausible conservation target. These targets give rise to four valuation scenarios, two for each welfare frame: WTP to increase conservation from 14% to 35% or 50%; and, WTA compensation to forego an increase from 14% to 35% or 50% conservation levels.

We utilize a between-subjects design wherein each respondent is randomly selected into one of the four valuation scenarios (treatments), and then randomly assigned a bid or offer amount. In all cases, an SBC elicitation is used, framed as an advisory referendum. The proposal for WTP involves a vote for an increased level of conservation accompanied by a fixed annual tax increase that would be invested in a Northern Quebec conservation fund. The WTA proposal is to forego protection of wetlands, thus increasing government revenues that would be used to compensate households with a uniform refundable tax credit.[8]

Determining the relevant range of annual prices (i.e., bids and offers) required to properly identify the WTP and WTA distributions proved to be challenging. Initial price ranges were informed by pilot tests with an intercept sample (n=18) and online surveys (n=75) utilizing open-ended questions. A subsequent online survey with a large sample (n=576) and an SBC elicitation revealed that much higher prices were needed to identify WTA.[9] The final survey implementation utilized prices from the set {10, 50, 100, 150, 200, 250, 300, 400, 500, 750, 1000} for WTP, and WTA compensation amounts {100, 150, 200, 250, 300, 400, 500, 750, 1000, 1500}.

[8] A refundable tax credit implies that citizens who do not pay taxes still receive a payment. These types of programs already exist across Canada and in the province of Quebec, in particular.
[9] This data is not used for the analysis presented in this paper. The survey was very similar to the final version used for this paper but did not include the consequentiality question that is central to the present analysis.

Participants were informed that their answers would remain anonymous and that the survey was financed and approved by the Ministry of Development, Environment and Parks.[10] The content of the online survey consisted of a narrative video outlining the area and briefly describing potential impacts of hydroelectric and mining activities on wetland ecosystems.[11] It explained the benefits to water quality of increasing wetland conservation as well as the drawbacks to the provincial revenue stream caused by limiting highly profitable industrial development. A brief written summary reminded the respondent of key points before asking them to vote on the proposal.

Following the video and summary points, respondents voted on one of the four possible proposals. After their vote, participants answered the following question to gauge survey consequentiality: "When the Quebec government decides whether or not to implement the land conservation proposal you just voted on, how likely do you think it is that the government will take into account your vote and that of the other respondents to this study in its decision-making?" Possible response options included "very likely", "somewhat likely", "somewhat unlikely" and "very unlikely".

At the time the survey was administered, prior studies concerned with consequentiality tended to include one fairly vague question related to policy consequentiality.[12] Vague in the sense that it asks about whether survey responses will be considered by decision makers rather than whether votes on the survey proposal will be used to inform an actual decision regarding the same

[10] The agency is now known as the Ministry of Sustainable Development, Environment, and Fight Against Climate Change.
[11] The Reviewer Appendix includes a translation of the survey instrument, and a link to the video.
[12] Of the studies reviewed in Section 2, only Petrolia and Kim (2011) asked respondents about perceived consequentiality. More recent studies, such as Vossler and Holladay (2018) and Zawojska, Bartczak, and Czajkowski (2019), use more carefully worded questions, and elicit separately beliefs about policy consequentiality and payment consequentiality.

proposal. As our proposal makes explicit the cost (or compensation), this single question is intended to encompass both payment as well as policy consequentiality. For those with little or no agreement with the question we asked whether they believed the actual cost (compensation) or conservation goal considered would be different, whether the government would not use the survey results to make any conservation decision, and whether the government is actually considering any conservation policy. The remainder of the survey consisted of standard demographic questions, and attitudinal questions regarding respondent opinion on environmental protection and economic development.

The survey was conducted during the first three months of 2014. It was administered by a private firm who sent email invitations to a representative sample of the Quebec population taken from their private panel. In total, 1048 surveys were completed and all are included in our analysis.

4. Results

Table 2 provides a comparison of selected respondent characteristics with the adult population of the Quebec province, where the latter is drawn from 2016 census data. As can be seen in Table 2, the sample and population match up reasonably well, although the survey sample is somewhat younger and better educated. While we cannot rule out possible differences in unobservable characteristics between our sample and the general population, this is of secondary importance for our immediate purpose since our goal is to compare WTA and WTP across randomly determined sub-samples, and not to conduct a cost-benefit analysis or advise policy-making.

Table 2
Socioeconomic comparison of the survey sample with those of the Quebec population above 18 years of age

Socioeconomic variables	**Sample (%)**	**Population (%)**
Gender		
Male	55	49
Female	45	51
Age		
18-24	10	10
25-34	16	15
35-44	20	16
45-54	25	17
55-64	19	18
65 and over	10	23
Education		
No diploma	2	17
High school or trade school	33	40
College or CEGEP	27	18
University	38	25
Location		
Quebec City	14	10
Montreal	51	50
Other	35	40
Household income (Can$)		
Less than 15 000	5	6
15 000 to 24 999	7	10
25 000 to 34 999	9	10
35 000 to 54 999	19	20
55 000 to 74 999	19	16
75 000 to 99 999	17	14
100 000 and over	24	24

Note: Population statistics compiled from Statistics Canada, 2016 Census.

Table 3 describes key variables collected in the survey and used in the econometric analysis. Central to our analysis are stated beliefs over consequentiality. About 24% of participants responded "very unlikely" to the consequentiality question, indicating a lack of consequentiality. This percentage is nearly identical across frames: 24.5% for WTP versus 24.3% for WTA. Around 52% of respondents selected the "somewhat likely" option, which we interpret at connoting a "weak" level of consequentiality. We characterize the remainder of respondents (23%) as having a "moderate" level of consequentiality.[13] We asked those who selected "very unlikely" or "somewhat unlikely" about the drivers of this belief. However, the heterogeneity in responses and the small number of respondents limits our ability to make WTA-WTP comparisons across these subgroups.

[13] We characterize as "moderately consequential" those selecting the "very likely" or "somewhat likely" response options to the consequentiality question. Just 4% selected the former, and so we lump these together in the analysis.

Table 3
Data description

Variable	Description	Mean	Std. Dev.
WTP	= 1 if willingness-to-pay (WTP) frame	0.51	0.50
WTA	= 1 if willingness-to-accept (WTA) frame	0.49	0.50
Vote	= 1 if respondent voted "yes"	0.38	0.49
Bid / Offer	Amount of tax increase (WTP) or compensation offered (WTA), in Can$	393.16	359.30
Scope	= 1 for 50% conservation scenario	0.50	0.50
Charity	= 1 if the respondent donates to charities or non-profit organizations	0.70	0.46
Env Org	= 1 if the respondent is a member of an environmental organization	0.04	0.21
Male	= 1 if respondent is male	0.55	0.50
College	= 1 if respondent has a college certificate or diploma	0.64	0.48
Homeowner	= 1 if respondent is a homeowner	0.72	0.45
HH Size	Number of people living in the household	2.81	1.23
Retired	= 1 if respondent is retired	0.09	0.28
Student	= 1 if respondent is a student	0.18	0.39
Quebec	= 1 if respondent lives in the Quebec metropolitan census area	0.14	0.35
Montreal	= 1 if respondent lives in the Montreal metropolitan census area	0.51	0.50
Inconsequential	= 1 if believes that survey is "not at all" consequential	0.24	0.43
Weakly consequential	= 1 if believes that survey is "weakly" consequential	0.52	0.50
Moderately consequential	= 1 if believes that survey is "somewhat" or "strongly" consequential	0.23	0.42
Date IV	Day respondent completed survey, 1 to 25 (=1 for first day of survey returns)	8.24	7.39

For the econometric analysis of the WTA-WTP disparity, we consider several model specifications, each applied separately to the WTP and WTA data. In all cases, we model the yes/no voting choice using a probit estimator. As detailed in Cameron and James (1987), mean/median WTP (or WTA) estimates can be obtained through a nonlinear transformation of probit coefficients. We use the delta method to compute standard errors for WTA, WTP, and the WTA/WTP ratios. For models that include control variables, welfare estimates are calculated using the means of these variables computed from the full survey sample.

Specification 1 is the simplest possible specification. It models the voting choice as a function of the bid/offer. The remaining specifications control for consequentiality, which is important in theory. In Specification 2, both the location and the scale of the welfare distributions are allowed to vary across the three consequentiality levels: inconsequential, weakly consequential, and moderately consequential.[14] Interestingly, for the WTA model, there is only statistical support for varying scale. On the other hand, for the WTP model, there is only support for varying location.[15]

While there is mixed evidence in the literature that stated beliefs of consequentiality are endogenous in choice models (see Börger et al., *forthcoming*), we nevertheless include three specifications to help address this issue. In Specification 3, we include control variables (defined in Table 3) to help adjust for differences in observed characteristics across consequentiality levels.

[14] In a probit model, the negative inverse of the coefficient on the bid/offer variable is a point estimate of the standard deviation of the WTA (WTP) function. To vary scale, the consequentiality level indicators are interacted with the bid/offer variable. To vary location, consequentiality level indicators enter directly as additional covariates in the probit.

[15] Welfare estimates are obtained from restricted versions of the varying location/scale specification, as justified by statistical tests and documented in the appendix.

Specification 4 fully interacts the control variables with the consequentiality indicators, an approach known as "regression adjustment" in the program evaluation literature (see Wooldridge, 2010). Specification 5 introduces an instrumental variables approach. The political climate in Quebec was dynamic during the four-week period of our data collection. As a result, it is likely that beliefs over consequentiality varied over time.[16] As the response date to the survey is plausibly random (i.e., potential survey takers were invited incrementally), and uncorrelated with WTA (WTP), we use the response date to construct an instrumental variable. We jointly model a two-equation system consisting of the voting equation and a consequentiality equation. This approach is similar to Herriges et al. (2010), although we use maximum likelihood instead of Bayesian methods. We assume consequentiality is a latent dependent variable and model it using an ordered probit since we have three consequentiality levels with a natural ordering.

When calculating welfare, we compute a weighted average of values across the 35% and 50% conservation scenarios.[17] For both WTP and WTA, the welfare estimates across the two conservation scenarios are not statistically different for any specification. One might interpret this result as a failure of the "scope test", which is an important test of construct validity.[18] However, this finding was not unexpected in our setting. We again note that our 35% scenario was suggested by ecologists as necessary to maintain water quality, whereas the 50% scenario was included based on political interests. In all valuation scenarios, respondents were provided information regarding

[16] A general election for members to the National Assembly of Quebec took place a few weeks after our survey ended. During the survey period there was considerable uncertainty over whether the Quebec Liberal Party or the Parti Québécois had a realistic prospect of winning a majority government.

[17] To be clear, we include an indicator for the 50% conservation scenario as a control variable in the econometric models, and plug in the mean of this indicator when calculating WTA and WTP (as we do for other control variables).

[18] A standard assumption of consumer choice theory is that the utility function is *weakly* increasing in the quantity of a good. As such, this scope insensitivity result does not necessarily imply a violation of theory. See Whitehead (2016) for a recent discussion of the scope test.

the ecologists' assessment of the 35% threshold. This threshold may have served as a signal of the socially desirable level of conservation, leading to perceptions that the benefits from 50% conservation were not larger than those obtained from preserving 35% of the area. A related finding of scope insensitivity was found by Bateman et al. (2005), in a similar setting where respondents were evaluating a return in water quality from current ecologically sound levels to natural pre-industrial levels.

Table 4 presents welfare estimates based on the five econometric specifications. Full estimation results are presented in Tables A1-A5 of the appendix. Using Specification 1, which ignores consequentiality beliefs, the WTA/WTP ratio is extremely large. Driving this result is an extremely high mean WTA estimate. In fact, the estimated WTA function is very "flat" in the sense that the probability of votes in favor of the proposal varies little across compensation amounts. Under an ideal scenario, we would be able to precisely estimate WTA regardless of beliefs over consequentiality. This in turn would provide an indication of whether WTA increases or decreases with consequentiality. In retrospect this may have been accomplished by extending the range of compensation amounts considerably.[19] Even then, there is no guarantee that that this strategy would have proven successful. Indeed, it is likely that we would have needed unrealistically high offer amounts, which could have reduced the credibility of the value elicitation exercise.

[19] As the WTA/SE(WTP) ratio is just 0.09, it is very unlikely that even a tripling or quadrupling of the sample size would have led to a precise (unconditional) WTA estimate.

Table 4
Welfare estimates (Can$)

	WTA	WTP	WTA/WTP
Specification 1: fixed location/scale model			
Unconditional	55742.80 (599 738.1)	315.80*** (63.09)	176.52 (1899.46)
Specification 2: varying location/scale model			
Inconsequential	–5993.57 (11 704.19)	–228.55 (195.99)	26.22 (55.93)
Weakly consequential	–3516.86 (3 173.60)	308.21*** (90.87)	–11.41 (10.83)
Moderately consequential	1033.02*** (293.56)	874.44*** (227.86)	1.18*** (0.46)
Specification 3: varying location/scale model, with control variables			
Inconsequential	–4461.75 (6 190.57)	–159.28 (177.64)	28.01 (49.87)
Weakly consequential	–4591.53 (4 879.10)	259.33*** (89.30)	–17.71 (19.78)
Moderately consequential	964.04*** (234.95)	886.90*** (224.52)	1.09*** (0.38)
Specification 4: varying location/scale model, with regression adjustment			
Inconsequential	–1941.08 (1 419.17)	–156.96 (172.65)	12.37 (16.33)
Weakly consequential	–10404.67 (22 824.85)	248.44*** (89.66)	–41.88 (93.11)
Moderately consequential	951.40*** (215.73)	836.30*** (209.55)	1.14*** (0.38)
Specification 5: IV probit			
Inconsequential	–4686.70 (8 194.93)	–232.37 (1 498.62)	20.17 (134.77)
Weakly Consequential	–4596.23 (4 883.99)	258.43*** (92.37)	–17.79 (19.94)
Moderately Consequential	973.55*** (317.18)	960.07 (1500.15)	1.01 (1.62)

Notes: *, **, and *** denote estimates that are statistically different from zero at the 10%, 5%, and 1% significance levels, respectively. Standard errors in parentheses. See text for description of underlying econometric models.

In contrast, WTP is precisely identified, with a point estimate of $315.80 per person per year. To place this estimate in perspective, a meta-analysis of the value of wetlands that provide water quality services (Brander et al., 2006, p. 25) report annual median and mean values of approximately $30 per hectare and $1500 per hectare (in 1995 US$) respectively. Converting our estimate yields an annual mean WTP of about $29 per hectare (in 1995 US$).[20]

Specifications 2 through 5 highlight that conditioning on stated consequentiality can alter conclusions significantly. Mean WTA is statistically insignificant for the "inconsequential" and "weakly consequential" respondents. However, for respondents with "moderately consequential" beliefs, the welfare estimates are precise and range from $951 to $1033 across specifications. For WTP, estimates are likewise imprecise for the "inconsequential" respondents. Mean WTP is statistically significant for the "weakly consequential" respondents, and ranges from $248 to $308. For the "moderately consequential" respondents, WTP ranges from $837 to $887 across Specifications 2 to 4. Although the literature provides mixed results on the relationship between beliefs over consequentiality and WTP, that WTP increases with consequentiality is a common finding in field surveys (e.g., Herriges et al., 2010).

The IV probit estimate of WTP is somewhat higher at $960. Although the point estimates for most coefficients are very close to those obtained from non-IV models, the WTP estimate itself is imprecise. The IV probit estimation results suggest that, for both the WTA and WTP data, the IV is statistically significant in explaining (latent) consequentiality. On the other hand, demographic variables do a poor job at explaining consequentiality beliefs. In neither case is there statistical evidence to support the use of an instrumental variables approach. Specifically, the

[20] We arrived at this figure by first multiplying the mean WTP estimate by the number of adults living in the province, and then dividing by the number of hectares conserved under the 35% scenario. This figure was then converted to US$, and finally deflated to reflect the value of year 1995 currency.

estimated correlation coefficients are not statistically different from zero, suggesting that the voting and consequentiality equations are statistically independent. The point estimate of the WTA/WTP ratio is smaller, although very close to those estimated from the non-IV models.[21]

Overall, for respondents holding "moderately consequential" beliefs, the WTA/WTP rations are very close to one, with point estimates ranging from 1.0 to 1.2. Importantly, but with the exception of the WTP estimate in the IV probit model, the WTA and WTP measures are precisely estimated, along with the corresponding WTA/WTP ratios. These estimates are quite small in comparison to what is typical in the broad literature, including for revealed preference studies. To place our estimates in perspective, the meta-analysis by Tunçel and Hammitt (2014) generates a geometric mean WTA/WTP ratio of 6.23 for studies that elicit values for environmental goods.

5. Discussion

Plott and Zeiler (2005) argue that the observed disparities between the willingness-to-accept (WTA) and willingness-to-pay (WTP) for a private good may be an artifact of poor experimental designs, including the use of non-incentive compatible mechanisms. They also provide evidence that adopting mechanisms with desirable theoretical properties can significantly reduce the estimated difference between WTA and WTP. For public goods, research in mechanism design for stated preference surveys has only recently identified conditions for incentive

[21] As a robustness check, given the effect of the response date on consequentiality may be highly nonlinear, we explored an alternative specification where we included a set of date indicator variables as IVs. This yields similar results (moderately consequential): mean WTA = 904.20 (Std. Err. = 294.47); mean WTP = 939.24 (1474.90); and WTA/WTP = 0.96 (1.54).

compatible elicitation. Thus, it is natural to ask whether prior findings from the WTA-WTP literature hold when incentive compatible mechanisms are introduced.

Using as a case study the elicitation of WTA and WTP for the conservation of wetlands in northern Quebec – a large-scale public good – we provide evidence that the WTA/WTP ratio decreases with incentive compatible elicitation methods. In fact, based on our literature review, our estimated ratios of 1.0 to 1.2 are the smallest reported from studies that elicit values for large-scale environmental public goods. The vast majority of related studies have used open-ended and other continuous mechanisms, and we provide theoretical arguments that this can lead to bias, and in some cases an upward bias, in WTA/WTP ratios. Our empirical results thus provide suggestive evidence that theory and methodology matter. They challenge the prior view that credible WTA elicitation is not possible or that stated preference surveys are generally flawed. A direct implication of our findings is that (some) prior results may be artifacts of flawed designs. As existing evidence on WTA-WTP disparities has diminished the use of WTA elicitation, our results provide some new optimism that eliciting WTA using contemporary methods can be considered, especially when WTA corresponds with the existing assignment of property rights.

Of course, a finding of a relatively small WTA/WTP ratio does not in and of itself suggest that this ratio is "accurate" or "valid". It remains an open question whether our estimated WTA/WTP ratios accurately reflect true preferences. Neoclassical economic theory suggests that this relationship depends on income and substitution effects, and we do not have reasonable estimates of these. Hanemann's (1991) numerical simulations demonstrate that ratios of 5-to-1 (in the range of many empirical studies), can arise for goods that have a low elasticity of substitution, or for goods with a moderate elasticity of substitution that are highly valued relative to income. We suspect that the substitutability between the public good in our study (preservation of wetlands)

and other goods is fairly high. As revealed through a focus group, the largest component of preservation values is tied to the ability of the wetlands to maintain the quality of water flowing to the south. Municipalities throughout the province have the ability to mitigate water quality issues, and households can otherwise use filtration devices and consume bottled water. A high elasticity of substitution, along with somewhat modest ratios between WTA (or WTP) and income, suggests that a WTA/WTP ratio near unity is at least plausible in this case.

It would be dangerous to attempt generalizations based on the findings of a single study. We hope, however, that our investigation will renew interests in WTA studies and motivate others to accumulate additional data points. Aside from merely replicating our methods in a different setting, a number of approaches could be deployed to test the robustness of our general findings. First, one might use split-sample comparisons of WTA/WTP ratios obtained from different elicitation methods, e.g. single binary choice versus open-ended. Second, given that over half of the prior studies of public goods involved within-subject comparisons, it may be fruitful to compare, within a single study, the ratios obtained using within- versus between-subject data. Third, it would be very informative to conduct studies where variations in policy scenarios can introduce variations in substitution and income effects.

References

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Shuman, 1993. Report of the NOAA Panel on Contingent Valuation. *Federal Register* 58: 4601-4614.

Bateman, Ian J., Philip Cooper, Stavros Georgiou, Stale Navrud, Gregory L. Poe, Richard C. Ready, Pere Reira, Mandy Ryan, and Christian A. Vossler. 2005. Economic Valuation of Policies for Reducing Acidity in Remote Mountain Lakes. *Aquatic Sciences* 67(3): 274-291.

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

Brander, Luke M., Raymond J. G. M. Florax, and Jan E. Vermaat. 2006. The empirics of wetland valuation: A comprehensive summary and a meta-analysis of the literature. *Environmental & Resource Economics* 33: 223-250.

Brookshire, David S. and Don L. Coursey. 1987. Measuring the value of a public good: An empirical comparison of elicitation procedures. *The American Economic Review* 77(4): 554-566.

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 & 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.

Charness, Gary, Uri Gneezy, and Michael A. Kuhn. 2012. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior and &Organization* 81: 1-8.

Chilton, Susan, Michael Jones-Lee, Rebecca McDonald, and Hugh Metcalf. 2012. Does the WTA/WTP ratio diminish as the severity of a health complaint is reduced? Testing for smoothness of the underlying utility of wealth function. *Journal of Risk and Uncertainty* 45: 1-24.

Fehr, Dietmar, Rustamdjan Hakimov, and Dorothea Kübler. 2015. The willingness to pay – willingness to accept gap: A failed replication of Plott and Zeiler. *European Economic Review* 78: 120-128.

Flachaire, Emmanuel, Guillaume Hollard, and Jason F. Shogren. 2013. On the origin of the WTA-WTP divergence in public good valuation. *Theory and Decision* 74: 431-437.

Hanemann, W. Michael. 1991. Willingness to Pay and Willingness to Accept: How Much Can They Differ? *The American Economic Review* 81(3): 635-647.

Hausman, Jerry. 2012. Contingent Valuation: From Dubious to Hopeless. *Journal of Economic Perspectives* 26(4): 43-56.

Herriges, J., Kling, C., Liu, C.-C., Tobias, J. 2010. What are the consequences of consequentiality? *Journal of Environmental Economics and Management* 59(1): 67-81.

Horowitz, John K. and K.E. McConnell. 2003. Willingness to accept, willingness to pay and the income effect. 2003. *Journal of Economic Behavior & Organization* 51: 537-545.

Isoni, Andrea, Graham Loomes, and Robert Sugden. 2011. The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Comment. *American Economic Review* 101: 991–1011.

Johnston, Robert J., Kevin J. Boyle, Wiktor Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Cameron, W. Michael Hanemann, Nick Hanley, Mandy Ryan, Riccardo Scarpa, Roger Tourangeau, and Christian A. Vossler. 2017. Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists* 4(2): 319-405.

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.

Koń, Beata and Michał Jakubczyk. 2019. Is the literature on the WTP-WTA disparity biased? *Journal of Behavioral and Experimental Economics* 82: 101460.

Lloyd-Smith, Patrick and Wiktor Adamowicz. 2018. Can stated measures of willingness-to-accept be valid? Evidence from laboratory experiments. *Journal of Environmental Economics and Management* 91: 133-149.

Petrolia, Daniel R. and Tae-Goun Kim. 2011. Preventing loss in coastal Louisiana: Estimates of WTP and WTA. *Journal of Environmental Management* 92: 859-865.

Plott, Charles R. and Kathryn Zeiler. 2005. The Willingness to Pay – Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *The American Economic Review* 95(3): 530-545.

Tunçel, Tuba and James K. Hammitt. 2014. A new meta-analysis on the WTP/WTA disparity. *Journal of Environmental Economics and Management* 68: 175-187.

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 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 & Organization* 86: 137-147.

Whitehead, John C. 2016. Plausible responsiveness to scope in contingent valuation. *Ecological Economics* 128: 17-22.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*, Second Edition. Cambridge, MA: The MIT Press.

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.

Appendix A. Econometric models

Table A1. Specification 1: fixed location/scale model

Dependent variable: = 1 if respondent voted "yes" (*Vote*)		
	WTA	WTP
Bid	0.0000127	–0.000883***
	(0.000143)	(0.000250)
Intercept	–0.709***	–0.279***
	(0.0957)	(0.0872)
Observations	514	534
Log-L	–283.987	–363.562
R^2	0.0000	0.0171

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2. Specification 2: varying location/scale model

Dependent variable: = 1 if respondent voted "yes" (*Vote*)				
	WTA[1]	WTA	WTP[2]	WTP
Bid	–0.000312	–0.000119	–0.000288	–0.000850***
	(0.000324)	(0.000240)	(0.000519)	(0.000256)
Bid × Weakly	0.000277	–0.0000839	–0.000799	
Consequential	(0.000381)	(0.000243)	(0.000629)	
Bid × Moderately	0.000837*	0.000811***	–0.000606	
Consequential	(0.000443)	(0.000283)	(0.000736)	
Weakly	–0.298		0.674***	0.456***
Consequential	(0.235)		(0.221)	(0.138)
Moderately	–0.0118		1.01***	0.938***
Consequential	(0.272)		(0.253)	(0.161)
Intercept	–0.564***	–0.714***	–0.345*	–0.194
	(0.1900)	(0.0969)	(0.181)	(0.132)
Observations	514	514	534	534
Log-L	–274.892	–276.022	–345.147	–345.953
R^2	0.0320	0.0281	0.0669	0.0647

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. [1] Test of $\beta_{\text{Weakly Consequential}} = \beta_{\text{Moderately Consequential}} = 0$: $\chi^2(2) = 2.25$, p=0.3240. [2] Test of $\beta_{\text{Bid} \times \text{Weakly Consequential}} = \beta_{\text{Bid} \times \text{Moderately Consequential}} = 0$: $\chi^2(2) = 1.62$, p=0.4440.

Table A3. Specification 3: varying location/scale model, with control variables

Dependent variable: = 1 if respondent voted "yes" (*Vote*)				
	WTA[1]	WTA	WTP[2]	WTP
Bid	–0.000317	–0.000179	–0.000450	–0.000896***
	(0.000343)	(0.000255)	(0.000535)	(0.000266)
Bid × Weakly	0.000252	5.07e–06	–0.000706	
Consequential	(0.000399)	(0.000255)	(0.000649)	
Bid × Moderately	0.00105**	0.00101***	–0.000333	
Consequential	(0.000472)	(0.000297)	(0.000757)	
Weakly	–0.207		0.568**	0.375***
Consequential	(0.250)		(0.228)	(0.143)
Moderately	–0.0255		1.03***	0.938***
Consequential	(0.292)		(0.261)	(0.166)
Scope	0.188	0.196	0.0349	0.0334
	(0.130)	(0.129)	(0.117)	(0.117)
Charity	–0.448***	–0.457***	0.287**	0.300**
	(0.140)	(0.139)	(0.136)	(0.135)
Env Org	–0.0237	–0.0363	0.312	0.302
	(0.340)	(0.339)	(0.294)	(0.293)
Male	0.118	0.120	–0.333***	–0.335***
	(0.130)	(0.130)	(0.118)	(0.117)
College	–0.282**	–0.278**	0.0471	0.0441
	(0.133)	(0.133)	(0.124)	(0.124)
Homeowner	0.146	0.149	–0.440***	–0.443***
	(0.157)	(0.157)	(0.141)	(0.140)
HH Size	0.127**	0.130**	0.101*	0.102**
	(0.0581)	(0.0580)	(0.0520)	(0.0519)
Retired	–0.732***	–0.737***	0.475**	0.460*
	(0.263)	(0.263)	(0.237)	(0.235)
Student	0.295*	0.294*	–0.286*	–0.292*
	(0.174)	(0.171)	(0.169)	(0.169)
Quebec	0.453**	0.458**	–0.0857	–0.0722
	(0.191)	(0.191)	(0.185)	(0.184)
Montreal	0.0658	0.0641	0.122	0.121
	(0.147)	(0.146)	(0.128)	(0.128)
Intercept	–0.690***	–0.801***	–0.262	–0.143
	(0.205)	(0.103)	(0.186)	(0.137)
Observations	514	514	534	534
Log-L	–256.077	–256.538	–322.268	–322.885
R^2	0.0983	0.0967	0.1288	0.1271

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Control variables are demeaned. [1] Test of $\beta_{\text{Weakly Consequential}}=\beta_{\text{Moderately Consequential}}=0$: $\chi^2(2)=0.92$, p=0.6306. [2] Test of $\beta_{\text{Bid} \times \text{Weakly Consequential}}=\beta_{\text{Bid} \times \text{Moderately Consequential}}=0$: $\chi^2(2)=1.24$, p=0.5387.

Table A4. Specification 4: varying location/scale model, with regression adjustment (coefficients on interactions with controls omitted for convenience)

Dependent variable: = 1 if respondent voted "yes" (*Vote*)				
	WTA[1]	WTA	WTP[2]	WTP
Bid	–0.000421	–0.000446	–0.000519	–0.000939***
	(0.000371)	(0.000334)	(0.000569)	(0.000272)
Bid × Weakly	0.000389	0.000363	–0.000733	
Consequential	(0.000427)	(0.000361)	(0.000683)	
Bid × Moderately	0.00116**	0.00136***	–0.000134	
Consequential	(0.000511)	(0.000399)	(0.000792)	
Weakly	0.173		0.575**	0.381**
Consequential	(17.3)		(0.234)	(0.150)
Moderately	0.358		0.966***	0.933***
Consequential	(17.3)		(0.268)	(0.175)
Intercept	–1.08	–0.865***	–0.254	–0.147
	(17.3)	(0.118)	(0.191)	(0.142)
Observations	514	514	534	534
Log-L	– 243.984	–244.227	–315.583	–316.332
R^2	0.1409	0.1400	0.1468	0.1448

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Control variables are demeaned. Model includes a full set of interactions between the three consequentiality levels and the same control variables included in Specification 3. [1] Test of $\beta_{\text{Weakly Consequential}}=\beta_{\text{Moderately Consequential}}=0$: $\chi^2(2)=0.46$, p=0.7963. [2] Test of $\beta_{\text{Bid} \times \text{Weakly Consequential}}=\beta_{\text{Bid} \times \text{Moderately Consequential}}=0$: $\chi^2(2)=1.50$, p=0.4729.

Table A5. Specification 5: IV probit

	WTA		WTP	
	Vote	Consequential	Vote	Consequential
Bid	–0.000171	0.000377	–0.000888***	–0.0000665
	(0.000317)	(0.000174)	(0.000311)	(0.000323)
Bid × Weakly	–3.36e-06			
Consequential	(0.000315)			
Bid × Moderately	0.000993**			
Consequential	(0.000468)			
Weakly			0.436	
Consequential			(1.03)	
Moderately			1.06	
Consequential			(2.04)	
Date IV		–0.0261***		–0.0232**
		(0.0100)		(0.00994)
Scope	0.195	0.0100	0.0356	0.0112
	(0.130)	(0.0998)	(0.122)	(0.121)
Charity	–0.457***	0.0983	0.298**	0.0881
	(0.139)	(0.110)	(0.142)	(0.142)
Env Org	–0.0361	0.0569	0.295	0.0917
	(0.339)	(0.244)	(0.318)	(0.283)
Male	0.120	0.0554	–0.337***	0.180
	(0.130)	(0.101)	(0.120)	(0.122)
College	–0.278**	–0.111	0.0405	0.147
	(0.134)	(0.107)	(0.138)	(0.130)
Homeowner	0.149	–0.0493	–0.433*	–0.277*
	(0.157)	(0.120)	(0.232)	(0.143)
HH Size	0.130**	–0.0113	0.102*	0.0498
	(0.0580)	(0.0447)	(0.0522)	(0.0530)
Retired	–0.737***	0.130	0.458*	–0.173
	(0.263)	(0.173)	(0.239)	(0.245)
Student	0.294*	0.234*	–0.295*	0.219
	(0.171)	(0.136)	(0.178)	(0.173)
Quebec	0.458**	–0.308**	–0.0752	0.236
	(0.191)	(0.156)	(0.191)	(0.187)
Montreal	0.0643	–0.0745	0.116	0.0472
	(0.146)	(0.113)	(0.146)	(0.134)
Intercept	–0.801***		–0.206	–0.649***
	(0.103)		(1.08)	(0.105)
Cut1		–0.719***		–0.832***
		(0.0870)		(0.0932)
Cut2		0.806***		0.536***
		(0.0880)		(0.0903)
ρ	0.00499		–0.0477	
	(0.110)		(0.809)	
Observations	514		534	
Log-L	–763.985		–871.352	

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Control variables are demeaned. Models estimated using the "eprobit" command in Stata v.15.