

**Revisiting the gap between the willingness to pay and  
willingness to accept for public goods**

Christian A. Vossler\*  
Department of Economics  
University of Tennessee  
Knoxville, Tennessee, USA

Stéphane Bergeron and Maurice Doyon  
Département d'économie agroalimentaire et des sciences de la consommation  
Université Laval  
Québec City, Québec, Canada

Daniel Rondeau  
Department of Economics  
University of Victoria  
Victoria, British Columbia, Canada

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://doi.org/10.1086/721995>.

© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

\*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 are especially grateful to three anonymous referees who provided extensive and valuable comments. Special thanks to Monique Poulin for sharing her expertise on wetlands. We also benefited from feedback received at the annual meetings of the Association of Environmental and Resource Economists, European Association of Environmental and Resource Economics, and European Society for Cost-Benefit Analysis, as well as from seminar participants at University of Alabama and University of Maryland. 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.

## **Revisiting the gap between the willingness to pay and willingness to accept for public goods**

**Abstract:** Large differences between willingness to pay (WTP) and willingness to accept (WTA) compensation measures have raised concerns over the validity of stated preference methods for valuing public goods. These differences have also motivated the use of WTP scenarios, even where property rights imply a WTA framing. Extending the insights of Plott and Zeiler (2005), we argue that WTA-WTP gaps can be reduced by deploying incentive compatible surveys and controlling for game form misconceptions, including a failure of respondents to view the survey as consequential. In a study of large-scale wetland conservation, and using a scenario-specific consequentiality measure, we find that the WTA/WTP ratio is between 4.8 and 6.5 for respondents unlikely to have perceived the survey to be consequential. The ratio falls below 2 for respondents likely to hold consequentiality beliefs. Using incentive compatible mechanisms and controlling for game form misconceptions could be the path to ensuring valid welfare estimates.

**JEL Classifications:** C93; D6; Q51

**Keywords:** stated preferences; incentive compatibility; willingness to pay; willingness to accept; consequentiality; game form misconceptions

## 1. Introduction

Decades of stated preference research has produced a considerable body of evidence showing large differences between estimates of willingness to pay (WTP) for the provision of a good and measures of the willingness to accept (WTA) compensation to forego the same good. These WTA-WTP differences defy the intuition of many and have been interpreted as indications that “willingness to pay questions measure preferences but willingness to accept questions do not” (Horowitz and McConnell 2003, p. 544). They have also been viewed as more general evidence that “answers to contingent valuation questions do not actually reflect stable or well-defined preferences” (Hausman 2012, p. 47).

Conclusions drawn from prior research and the influential report of the NOAA Panel on Contingent Valuation (Arrow et al. 1993) have led researchers to rely heavily on the estimation of WTP. For instance, Lloyd-Smith and Adamowicz (2018) report that the number of WTP studies in the Environmental Valuation Reference Inventory database is fourteen times larger than the number of WTA studies. It is possible that a WTP frame (rather than WTA) is appropriate for the vast majority of these goods. However, the NOAA panel’s recommendation that a WTP frame be utilized even when existing property rights or entitlements call for WTA has almost certainly influenced the design of many studies.<sup>1</sup>

In this paper we argue that deploying an incentive compatible mechanism and controlling for game form misconceptions can significantly reduce the WTA-WTP disparity. In the empirical analysis of games (which a valuation survey is), a game form misconception arises whenever a player’s interpretation of the game deviates from the game she is asked to play (Chou et al. 2009). Game form misconceptions are not limited to errors and misunderstandings. They arise whenever a player holds inaccurate beliefs about *any* part of the game form. So, where researchers are deploying a theoretically incentive compatible mechanism, participants who believe the survey to be inconsequential suffer from game form misconceptions. In our analysis, we present empirical evidence supporting the conjecture that controlling for game form misconceptions can reduce the welfare gap.

---

<sup>1</sup> That the State of Alaska and U.S. Federal Government relied on a WTP study to assess the natural resource damages caused by the Exxon Valdez oil spill is one example, but WTP studies have also been conducted to measure compensable damages under different U.S. laws (e.g., Oil Pollution Act; Comprehensive Environmental Response, Compensation, and Liability Act; Clean Water Act).

While many estimated WTA-WTP gaps have been viewed as improbably large, there does not exist a precise theoretical prediction for the magnitude of the difference for public goods.<sup>2</sup> The widespread presumption that the welfare gap should be small emerged from the work by Henderson (1941) on what became known as the four Hicksian welfare measures (Hicks 1943). Of his findings, Henderson wrote “we shall normally expect the four results to lie so close together that it would not matter which we chose” (p.121). Willig (1976) later provided a plausible quantification of the welfare gap for a price change while Randall and Stoll (1980) analyzed the impact of a quantity variation. Findings from the two studies suggest that the difference between WTP and WTA measures should be small when considering marginal changes in prices or quantities for market goods with readily available substitutes.

Randall and Stoll (1980) note, however, that large differences in welfare measures might arise when the quantity change is large, the good is highly valued, or the price elasticity of income for the good is high and rises with income. Hanemann (1991) considers the class of goods for which the household cannot exercise a direct choice over the quantity it consumes. He shows that an exogenous change in the quantity of such a good can produce a significant welfare gap (potentially infinitely large) if the good does not have close market substitutes. This result implies that the large differences between WTP and WTA observed in stated preference studies are not sufficient to invalidate the methodology. Moreover, the practice of estimating WTP when the policy setting warrants a WTA approach could result in overly conservative welfare estimates and the under-provision of public goods.

In the absence of a test to determine whether the magnitude of the welfare gap is acceptable, it is imperative that researchers deploy theoretically sound value-elicitation methods when comparing WTA and WTP. Unfortunately, Johnston et al. (2017) document that much of the evidence on the welfare gap comes from older studies that do not reflect current best practices. This is in line with Tunçel and Hammitt (2014) who present evidence from a meta-analysis showing that the welfare gap may be sensitive to research methods. Of particular relevance is their finding that using an incentive-compatible mechanism decreases the WTA/WTP ratio.<sup>3</sup>

---

<sup>2</sup> Precise predictions can be derived if the researcher assumes a specific utility function and the value of key structural parameters such as the elasticity of income and substitution linked to the public good.

<sup>3</sup> 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 and methods.

The desirability of employing incentive compatible mechanisms in stated preference studies has long been recognized (e.g., Randall, Hoehn and Brookshire 1983). Yet, a granular understanding of the conditions ensuring that valuation surveys are incentive compatible remains relatively recent (Carson and Groves 2007; Vossler, Doyon and Rondeau 2012; Carson, Groves and List 2014). This is evident from our review of past studies of the WTA-WTP gap presented in the next section. We find that few of them employed a theoretically incentive compatible mechanism or controlled for relevant respondent beliefs. This casts doubts on the validity of the conclusions reached from this literature.

Empirical evidence from WTP studies supports the notion that using incentive compatible mechanisms and controlling for consequentiality beliefs enhances construct and criterion validity (Herriges et al., 2010; Vossler, Doyon, and Rondeau, 2012; Vossler and Watson, 2013; Vossler and Holladay, 2018). In contrast, the literature provides few links between WTA elicitation and the theoretical underpinnings of contingent valuation instrument design. A lab experiment by Lloyd-Smith and Adamowicz (2018) confirms that incentive compatibility and consequentiality beliefs affect the measurement of WTA. Yet, the relationship between the incentive compatibility of survey instruments and the WTA-WTP gap has not yet been explored, and this is a central objective of this paper.

Our research has parallels with arguments and insights obtained from laboratory experiments. The WTA-WTP gap was first identified in contingent valuation studies, but it was also reliably replicated in the laboratory using a range of different experimental procedures. This spawned a large literature and contributed to the growth of behavioral economics. Plott and Zeiler (2005; 2011) review and critically examine this research. They suggest that persistent and large WTA-WTP gaps in experiments may be due to poor experimental designs and “subject misconceptions” about the mechanisms used in these studies. They show that deploying an incentive compatible mechanism and controlling for subject misconceptions can significantly reduce the WTA-WTP gap for private goods.<sup>4</sup>

The two lines of research naturally intersect. To correctly interpret laboratory data on preferences, it is important for experimenters to use an incentive compatible mechanism and for

---

<sup>4</sup> Specifically, Plott and Zeiler (2005) find that by using an incentive compatible mechanism and controlling for misconceptions, the resulting welfare gap is not statistically different from zero. Subsequent research utilizing similar experimental procedures reveals that while the WTA-WTP disparity is sometimes reduced substantially, this is not always the case (e.g., Isoni, Loomes, and Sugden 2011; Fehr, Hakimov, and Kübler 2015).

participants to correctly understand the game form (i.e., hold beliefs about the entire set up and working of the mechanism that correspond to the actual game they are playing). This ensures that observed decisions are premised on an accurate understanding of the links between the participant's actions and her welfare. For stated preference methods to be valid, respondents must similarly hold corresponding beliefs about the scenario deployed, and perceive that their answers can influence whether the good will be provided (lost) and paid for (compensated for), as described in the survey.

To shed light on the importance of using incentive compatible mechanisms and accounting for misconceptions when measuring the WTA-WTP gap for public goods, we conduct a contingent valuation survey of the protection of vast areas of wetlands in northern Quebec. This region is important for mineral extraction and hydroelectricity production, and further development poses threats to water quality and biodiversity. Congruent with recent methodological guidelines (Johnston et al. 2017), we elicit values through a single binary choice (SBC) question framed as an advisory referendum with a coercive taxation-based payment vehicle. WTA and WTP are elicited and compared using a between-subjects design.

Our analysis focuses on the effects of controlling for respondents' perceived consequentiality of the survey instrument. The consequentiality measure employed is unique to the literature and closely tied to the necessary conditions for the incentive compatibility of the mechanism. We show that value estimates and the WTA/WTP ratio depend critically on consequentiality beliefs. For respondents who view the elicitation as very or somewhat unlikely to be consequential, the ratio ranges from 4.8 to 6.5. This aligns well with central estimates from the meta-analysis of environmental and other public/non-market goods by Tunçel and Hammitt (2014). When respondents believe that the survey is somewhat or very likely to be consequential, the estimated WTA/WTP ratio decreases to between 1.6 and 1.9 depending on the model specification. These ratios are considerably lower than those reported in prior contingent valuation studies of large-scale public goods.

## **2. Mechanism design theory and prior WTA-WTP comparisons**

The objective of this section is twofold. First, we provide a unified theoretical framework grounded in mechanism design theory and a theory of game form misconceptions. The second objective is to review and assess existing evidence on the WTA-WTP gap from the perspective

of this framework.

## **2.1. Valuation surveys as a mechanism design theory problem**

The construction of a valuation survey is a classical mechanism design and implementation theory problem. The survey designer wishes to determine the value of a public policy but cannot observe citizens' preferences directly. The objective is to design a mechanism (i.e., a game) in which participants provide private information that can be used to accurately infer preferences, and in turn, aid policy and regulatory analyses. From a theoretical perspective, one way of achieving this goal is to design a dominant-strategy incentive compatible mechanism. In such a game, a participant can do no better than to take actions that truthfully reflect her preferences.

For responses to valuation questions to have any economic meaning, it is necessary for respondents to believe that their actions (answers/choices) can affect outcomes and their utility (Carson and Groves 2007). Consequentiality is therefore a necessary pre-condition for the theoretical analysis of any game. Consequentiality ensures that players are incentivized, and that theory can be applied to analyze their behavior or the properties of mechanisms.

One of the great challenges faced by survey designers is that no assurance can be given that respondents perceive consequences. Carson and Groves (2007) discuss a set of additional conditions that are together sufficient for a valuation survey to be dominant-strategy incentive compatible. The key conditions are: (1) the survey contains a single binary choice question over a proposed policy and cost; (2) answers to the question can only influence the likelihood of implementing the two possible outcomes described in the survey (i.e., policy provision at the cost/compensation indicated, and the status quo); and (3) a monotonicity requirement whereby a positive (negative) vote weakly increases (decreases) the probability that the policy and its cost to the respondent will (not) be implemented. Their analysis is formalized under expected utility by Vossler, Doyon, and Rondeau (2012) and under the weaker assumption of mixture monotonicity by Carson, Groves, and List (2014).<sup>5</sup>

---

<sup>5</sup> These conditions can be modified for mechanisms other than an SBC. However, the Gibbard-Satterthwaite theorem (Gibbard 1973; Satterthwaite 1975) tells us that mechanisms that offer more than one binary discrete choice are unlikely to be incentive compatible without imposing additional restrictions on preferences. Vossler, Doyon, and Rondeau (2012) identify sufficient conditions for a series of binary choice questions (choice experiments) and Vossler and Holladay (2018) identify conditions for open-ended and payment card elicitation formats. Both studies restrict the analysis to expected utility preferences. Vossler and Holladay (2018) note that past studies using open-ended and payment card formats are unlikely to meet the sufficient conditions.

## 2.2. Valuation surveys and game form misconceptions

The description of a mechanism or game is called its “game form”. It includes a description of the players, the strategies available to them, relevant information conditions, the relationship between strategy choices and outcomes, and the links between outcomes and player payoffs (Chou et al. 2009). The conditions establishing the incentive compatibility of a mechanism are purely theoretical. Like those presented above for the SBC mechanism, they are derived under the standard assumption that players take the game form presented to them as given (Osborne and Rubenstein 1994).

In empirical applications of incentive compatible mechanisms, there is no guarantee that players adhere to the game form presented to them. Participants who do not take the game form presented to them as given are essentially playing a different game, raising doubt that their actions truthfully reveal preferences.

Plott and Zeiler (2005) argue specifically that WTA-WTP gaps observed in laboratory experiments might be driven by subject misconceptions about the valuation elicitation mechanism. Chou et al. (2009) emphasize that failures of game form recognition can arise when subjects fail to properly perceive any aspect of the game form. Cason and Plott (2014) further develop a theory of “game form misconceptions”. The theory applies equally well to valuation surveys, and we characterize respondents’ deviations from consequentiality and other elements of the survey scenarios as game form misconceptions.

Conceptually, a respondent to an SBC valuation question has formed beliefs about how her choice might affect the likelihood of various outcomes. Denote the possible outcomes in her belief set by a vector  $\mathbf{B}$ . Each element of  $\mathbf{B}$  is itself a vector of public goods and the optimal bundle of private goods she would consume given her income under this outcome (which accounts for possible payments/compensation at whatever level thought of by the respondent). For analytical purposes, we impose that the set  $\mathbf{B}$  includes the status quo outcome, as well as the precise scenario described in the valuation survey (and consequences on the respondent’s private consumption). The respondent is assumed to have complete and transitive preferences defined over the outcomes in  $\mathbf{B}$ .

Corresponding to  $\mathbf{B}$  is a vector of subjective probabilities,  $\mathbf{R}$ . Each element is the respondent’s subjective probability that the corresponding outcome in  $\mathbf{B}$  will emerge in the absence of the survey. Respondents assumed to take the game form as given not only believe that



their votes can affect some outcomes (and their utility), but also that their vote *only* affects the likelihood (in  $\mathbf{R}$ ) of the two outcomes described in the scenario. Respondents who fail to take or to accept any part of the game form as given, or who make systematic errors in their interpretation or understanding of the game form, are said to suffer from game form misconceptions.

In this framework, a valuation instrument is said to be consequential if a respondent believes that her answer to the valuation question results in a revised vector of probabilities  $\mathbf{R}' \neq \mathbf{R}$ . That is, voting yes or no affects *any* of the probabilities in  $\mathbf{R}$ . Consequentiality only requires that answering the survey affects the likelihood of some outcomes. It is not sufficient to ensure incentive compatibility.

It is now somewhat common for researchers (e.g., Vossler and Holladay 2018; Zawojka, Bartczak, and Czajkowski 2019) to ask separate belief questions about what has been labelled “policy consequentiality” (i.e., whether responses influence whether any new policy is implemented) and “payment consequentiality” (i.e., whether payment would be compelled if a new policy is implemented). A respondent who believes that her vote is consequential in either dimension meets the necessary condition of consequentiality since she believes that actions do map into outcomes and potential payoffs. However, different answers to payment and policy consequentiality questions likely indicate that the respondent is not taking the binary referendum game form as given. If this is the case, their vote is unlikely to reflect true preferences for the good described in the survey.

There is plenty of empirical evidence of other failures of game form recognition in the valuation literature (e.g., Carson et al. 1994; Cameron, DeShazo, and Johnson 2011). Some respondents appear to have their own beliefs about how their answer might influence various outcomes not mentioned in the survey. Signs that a respondent’s beliefs are not aligned with the desired valuation scenario can take many forms including protest votes. The valuation literature has discussed this phenomenon and how to treat the resulting data under various labels, including scenario adjustments and scenario rejection. The issue of game form recognition is closely related to what Carson and Groves (2007) label as the “face-value property”.

To the extent that the survey scenario can be misconceived by respondents, it is natural to ask whether WTA and WTP scenarios are subject to the same type or degree of misconceptions. It is plausible that part of the explanation for observed differences between WTA and WTP is

that they engender systematically different deviations. For instance, respondents in a WTA context might be more likely to think that a “no” response will lead to a higher compensation level compared to the number of WTP respondents who believe that a “no” response will lead to a lower tax/price. Both types of respondents who believe that their answers can affect the payment amount suffer from game form misconceptions and we should worry that their answers might be systematically biased. Similarly, the number of respondents who reject a WTP scenario in the ex-post assessment of natural resource damages might be greater than those who reject a WTA scenario if, for instance, individuals generally believe that polluters should pay. How the number of biased respondents vary across welfare frame, and how the magnitude of the resulting biases compare, are important empirical questions.

### **2.3. Prior evidence on WTA-WTP disparities**

We now use insights from the theoretical framework to evaluate prior comparisons of WTA and WTP. Table 1 provides a characterization of 45 stated preference studies involving public or non-market goods, identified from the meta-analyses of Tunçel and Hammitt (2014) and Koń and Jakubczyk (2019).<sup>6</sup> Unfortunately, our review of this literature suggests that these studies almost never deployed an incentive compatible mechanism. Further, only Petrolia and Kim (2011) controlled for consequentiality beliefs in the data analysis. While there are a variety of shortcomings in past studies, it is worth exploring how past designs might bias welfare gap estimates for non-market goods.<sup>7</sup>

Nearly all studies (91%) rely on elicitation formats other than SBC, using instead open-ended questions or closely related iterative bidding games, payment cards, or multiple-bounded

---

<sup>6</sup> 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 only elicits WTP; five studies do not use stated preference methods; and we could not gain access to one study. 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.

<sup>7</sup> Designers of many of the early valuation studies often went to extreme efforts to tell respondents that the valuation scenarios were purely hypothetical in the sense that their answers would not have any influence. Yet, many of the same studies encouraged people to respond as they would in a consequential setting. Purely hypothetical surveys raise concerns similar to those that try to motivate consequential beliefs: some respondents will see the survey as inconsequential and pay little or no attention to the scenario; and some respondents might still think that since a researcher is bound to report study results, their answers might still be influential.

discrete choice formats.<sup>8</sup> These approaches were intended to identify welfare distributions more efficiently. While these methods might have empirical advantages, they can motivate responses that do not reveal true preferences and thus bias the resulting estimates of the WTA-WTP gap.

As an example, consider a case where the participant perceives that her open-ended response primarily influences the price that might ultimately be paid if a policy is implemented. In a WTP setting, the respondent should under-reveal demand to lower the price. In contrast, a respondent with the same beliefs who is facing a WTA survey would inflate the minimum compensation demanded.

Next, consider a respondent who believes that her valuation response primarily influences the likelihood that the good will be provided. In a WTP frame, the respondent should bid zero when her valuation is less than the price she anticipates since stating this value is the strongest signal that can be sent against provision. In the WTA setting, this respondent would instead demand a very high compensation level that exceeds her valuation.

Finally, consider a respondent who believes that she mostly influences the likelihood of policy implementation, but whose valuation of the good is higher than the price she expects might be imposed on her. This respondent would enter a very high value in the WTP setting, but a zero in a WTA survey.

While the three cases above suggest that the effect of strategic behavior on the welfare gap is ambiguous, Carson and Groves (2007) argue that open-ended WTP questions (and related formats) lead to underestimates of demand. This view is supported by evidence that WTP elicited from an SBC question exceeds the measurement elicited from an open-ended question or a payment card (Champ and Bishop 2006). With an open-ended question, the strategic opportunity to influence the price is more salient, and thus may diminish beliefs that responses primarily influence provision. In turn, the above arguments suggest that the use of open-ended formats in past studies might have inflated the WTA-WTP difference.

About three quarters of the studies (73%) implement a private provision mechanism that can also provide incentives to misrepresent preferences. Examples of these mechanisms include

---

<sup>8</sup> 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 elicitation resemble incentive-compatible mechanisms used in laboratory experiments. In a lab setting, substantial training has been shown to be needed for people to understand these mechanisms (see Plott and Zeiler 2005).

voluntary contributions towards the provision of a public good or the purchase (or foregone consumption) of quasi-public goods such as hunting permits and medical treatments. As discussed by Carson and Groves (2007), a respondent may think that stating an intended donation that exceeds her WTP might increase the likelihood that a real fund-raising drive will take place, in which case the respondent would not be bound to make the same payment. In a parallel WTA setting, it is possible that people may inflate WTA to influence whether compensation should be paid (e.g., Bush et al. 2013). The private provision of quasi-public goods raises issues that parallel the use of open-ended elicitation formats, in that they can provide respondents with opportunities to influence the future price of the good.

Almost two-thirds of the studies (62%) only compare WTA and WTP based on within-subject designs, giving rise to an entirely different methodological criticism (see Charness, Gneezy, and Kuhn 2012).<sup>9</sup> When a survey or experiment contains two value elicitation questions, the first valuation question creates a reference point and unwanted psychological sources of variation, resulting in the non-independence of tasks. The studies discussed by Charness, Gneezy, and Kuhn (2012) provide evidence from multiple contexts that differences between two valuations tend to be larger when elicited using a within-subject design. Within our setting, this tendency would inflate the welfare gap. As a practical matter, one or both elicitation questions are likely to be viewed as unrealistic 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.

In summary, the vast majority of comparative WTA-WTP studies neither implemented incentive compatible elicitation mechanisms nor controlled for game form misconceptions. This limits our ability to learn much about the actual magnitude of the WTA-WTP difference from prior work and motivates our empirical application.

### **3. Survey description and design**

#### **3.1. Background**

The 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

---

<sup>9</sup> A related concern is that many studies, including several that use between-subject designs to compare WTA and WTP, ask each respondent to value multiple goods.

parallel. The region spans 1.2 million square kilometers or approximately 70% of the province. Despite its size, only 2% 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 large scale hydroelectric dams, logging operations and several active mines.

This territory is known for its high potential for additional 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 an economic stimulus package that would inject Can\$80 billion 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 people ever travel to the region studied. 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 in the region under study, and in the vicinity of a recently completed hydroelectric project and significant uranium deposits.

No respondent expressed concerns related to direct use activities (fishing, berry picking, duck hunting, etc.), and the vast quantity of undeveloped space provides ample substitutes. However, and in line with province-wide opinion polls, serious concerns were expressed over water quality and quantity issues. Based on this information, and with buy-in from other stakeholders (government, biologists), we wrote the survey with a focus on eliciting values associated with the role of wetlands in regulating and purifying water flows.

### **3.2. Survey description**

The development of the survey instrument 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% of northern wetlands. Both represent significant increases to the status quo of 14% mandated by existing provincial regulations. The 35% level was advocated by ecologists as the minimum threshold necessary for maintaining water quality. The 50% level was selected because it had been publicly suggested as a plausible conservation target by the Quebec Premier.

These targets give rise to four valuation scenarios: WTP to increase conservation from 14% of the territory to either 35% or 50%; and, WTA compensation to forego an increase in conservation from 14% to either a 35% or 50%. While our valuation scenarios were motivated by an important and publicly debated policy issue, the government was not, to our knowledge, committed to evaluating the specific proposals we devised for this study.<sup>10</sup>

We implemented a between-subjects design whereby each respondent was 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 format framed as an advisory referendum was used. The WTP scenario asked respondents to vote on a proposal to fund a conservation policy through an annual flat tax that would be invested in a northern Quebec conservation fund. The WTA proposal was to forego the protection of wetlands, allowing for additional government revenues from development, money that would be used to compensate households via a uniform refundable tax credit.<sup>11</sup>

Determining the range of annual prices (i.e., bids and offers) required to properly identify welfare 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.<sup>12</sup> The final survey implementation utilized prices from the set {10, 50, 100, 150, 200, 250, 300, 400, 500, 750} for WTP, and WTA compensation amounts from the set {100, 150, 200, 250, 300, 400, 500, 750, 1000, 1500}.

Participants were informed that their answers would remain anonymous. They were also told that the study was financed and approved by the Ministry of Development, Environment and

---

<sup>10</sup> The Government of Quebec ultimately adopted its Plan Nord with a commitment to protect 50% of the territory. As of late 2021, roughly 20% of the territory was protected, with a target to protect the remaining 30% by 2035.

<sup>11</sup> With a flat tax or refundable tax credit, all citizens, including those who do not pay income taxes, receive a payment. Flat taxes and refundable tax credits are familiar instruments in the Canadian taxation system, and they are used in Quebec for a variety of purposes.

<sup>12</sup> 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.

Parks.<sup>13</sup> The online survey included a narrative video and a written summary presenting the study area and briefly describing potential impacts of hydroelectric and mining activities on wetland ecosystems in a factual and neutral manner.<sup>14</sup> The video explained how increasing wetland conservation would benefit water quality, and the losses in economic activity and provincial revenue that would result from limiting highly profitable industrial development.

Following the video and summary, 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, it was increasingly common for studies to include a question that gauges whether respondents perceive that survey responses will influence an outcome they care about, which is often labelled as a “policy consequentiality” question. We refer to our formulation as a “scenario-specific consequentiality” question as it conceptually captures not only whether the respondent believes that her vote can affect her utility, but also whether her vote can specifically influence the likelihood the policy presented to them (including the cost/compensation amount) is implemented.<sup>15</sup>

For those who responded very or somewhat unlikely, we followed up with a question on 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 about environmental protection and economic development.

---

<sup>13</sup> The agency is currently known as the Ministry of Sustainable Development, Environment, and Fight Against Climate Change.

<sup>14</sup> An English translation of the survey instrument, including a link to the video, is included in the appendix.

<sup>15</sup> More recent studies ask both a “policy consequentiality” and “payment consequentiality” question. Conceptually, those indicating agreement with our scenario-specific consequentiality question should answer in the affirmative to both policy and payment questions, but as the latter questions are not framed as being specific to the policy described the converse is not necessarily true.

The online survey was conducted in February and March 2014. It was administered by a firm who sent email invitations to a sample of the Quebec population taken from their private panel. Potential respondents were targeted to generate a sample that matched the Quebec population according to age, education, income, gender, and location (metropolitan Montreal, metropolitan Quebec, and the remainder of Quebec). A total of 1048 surveys were completed, and all are included in the analysis.<sup>16</sup>

## 4. Results

### 4.1. Data

Table 2 compares the respondent sample and the adult population of the province of Quebec as drawn from 2016 census data. Our sample and the population match up reasonably well, although the survey sample is somewhat younger and better educated. We cannot rule out possible differences in unobservable characteristics between our sample and the general population, which may impact welfare estimates. The work reported here nevertheless relies on random treatment assignment, and the data remain useful for testing whether employing incentive compatible mechanisms and controlling for game form misconceptions can alter the WTA-WTP gap.

Table 3 describes key variables collected in the survey and used in the econometric analysis.<sup>17</sup> Consequentiality beliefs, which are summarized in Table 4, are central to the analysis. Using a Pearson's chi-square test, we fail to reject the hypothesis that the response distributions to the consequentiality question are equal across WTA and WTP samples ( $p$ -value=0.14). About 24% of participants responded "very unlikely" to the consequentiality question, indicating a lack of consequentiality based on our strict definition. Around 52% of respondents selected the "somewhat unlikely" option, with the remainder (23%) either somewhat or very likely to hold consequentiality beliefs. As very few respondents selected "very likely", in the analysis that follows we combine these respondents with the "somewhat likely" group.

Most respondents who answered very or somewhat unlikely to the consequentiality question indicated that they did not think that the survey would influence the "level of territory

---

<sup>16</sup> We were unable to obtain data on response rates.

<sup>17</sup> Some of the demographic variables collected in the survey, such as household income, are excluded from the regressions utilizing control variables due to a high (item) non-response rate.



protected” (73% for WTA and 69% for WTP). Fewer than 10% indicated that they would not have to pay the stated cost (or receive the stated compensation) if a policy was implemented, and even fewer suggested that the conservation target being considered differed from the one described in the survey. Approximately 15% believed that the government was not actually considering any conservation policy.

Table 5 provides summary statistics for key variables conditional on responses to the consequentiality question. Most of these variables are indicators, and inspection of the data reveals that most variables only differ across consequentiality-based subsamples by a few percentage points at most. Using Person’s chi-square tests, we fail to reject that the response distributions are equal across the three samples for any variable, even at the 10% significance level. This suggests that observable characteristics are poor predictors of consequentiality, which is a common finding in the literature (Börger et al. 2021). Moreover, it suggests that – at least based on observables – sample representativeness of the greater population is not necessarily lost if we condition welfare estimates on consequentiality level. Additional tests, reported in the appendix (available online), further indicate that the WTA and WTP samples are comparable.

#### 4.2. Welfare estimation

For the econometric analysis of the WTA-WTP relationship, we consider several model specifications, each applied to the WTP and WTA data separately. In all cases, we interpret the yes/no vote as an interval-censored signal of value, following the insights of Cameron and James (1987). We assume that one’s valuation  $V_i^*$  (WTP or WTA) is a linear function  $V_i^* = \mathbf{x}_i\boldsymbol{\beta} + u_i$ , where  $\mathbf{x}_i$  is a vector of covariates,  $\boldsymbol{\beta}$  is a vector of unknown parameters, and  $u_i$  is a mean-zero error term.

In the WTP frame, a “yes” vote at cost  $c_i$  implies a lower bound,  $V_i^* \geq c_i$ . A “no” vote instead establishes an upper bound,  $V_i^* < c_i$ . In the WTA frame, a “yes” vote implies an upper bound, whereas a “no” vote defines a lower bound. For “no” WTP votes and “yes” WTA votes, we further assume that the WTP and WTA intervals are bounded from below at zero, rather than unbounded, which gives rise to the interval  $0 \leq V_i^* < c_i$  in these cases. Last, we assume a normal distribution for the errors, with  $u_i \sim Normal(0, \sigma_i^2)$ .

Define *Vote* as an indicator that equals 1 for a “yes” WTP vote or a “no” WTA vote. The log-likelihood function is then:

$$[1] \quad \ln \mathcal{L} = \sum_{i=1}^N \left\{ (1 - Vote_i) \cdot \ln \left( \Phi \left( \frac{c_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma_i} \right) - \Phi \left( \frac{-\mathbf{x}_i \boldsymbol{\beta}}{\sigma_i} \right) \right) + Vote_i \cdot \ln \left( 1 - \Phi \left( \frac{c_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma_i} \right) \right) \right\},$$

where  $\Phi$  denotes the normal CDF.

We consider several specifications. Specification 1 excludes covariates, which are not needed to identify welfare distributions given the random variation in cost across respondents, and assumes a common standard deviation for the welfare distribution (i.e.,  $\sigma_i = \sigma$ ). The remaining specifications allow the welfare distributions to vary based on consequentiality beliefs.<sup>18</sup> Our scenario-specific consequentiality question conceptually captures the extent to which a respondent's beliefs diverge from the valuation scenario, and the theoretical framework predicts that those with divergent beliefs may no longer have incentives to truthfully reveal their preferences about the policy described in the survey. It stands to reason that the consequentiality response categories contain different fractions of people with beliefs that closely align with the scenario, and that we may expect to see differences in valuations across the three consequentiality groupings.<sup>19</sup>

Specification 2 excludes control variables, but both the means and the standard deviations of the welfare distributions are allowed to vary across the consequentiality categories. In Specification 3, we add control variables (defined in Table 3) to help adjust for differences in observed characteristics across consequentiality categories. 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 treats consequentiality as a potentially endogenous determinant of valuations. A common concern is whether estimated relationships between consequentiality beliefs and valuations can be interpreted causally. One threat to identification is that the same uncontrolled factors underlying beliefs (e.g., trust in institutions) may also be driving valuations. Another threat is that survey questions are unlikely to precisely quantify beliefs, giving rise to measurement error. Although the overall evidence is mixed, a handful of papers find that

---

<sup>18</sup> Figure A1 in the appendix shows how the percentage of “yes” votes varies with consequentiality level.

<sup>19</sup> Theory predicts that respondents who perceive their vote to be consequential, in the sense of the necessary pre-condition that actions affect outcomes and utility, should consider the scenario in the same way, regardless of the strength of this belief. This is known as the “knife-edge” result, and some tests of the effects of consequentiality beliefs support this prediction (Herriges et al. 2010). Our scenario-specific consequentiality question conceptually identifies consequentiality and adherence to game form. Scenario-specific consequentiality does not have a knife-edge property.

respondent belief measures should be treated as endogenous variables in WTP models (e.g., Herriges et al. 2010; Lloyd-Smith, Adamowicz, and Dupont 2019; Börger et al. 2021).

The political climate in Quebec was dynamic during our four-week period of data collection, characterized by waning public support for the incumbent political party.<sup>20</sup> Stated beliefs over consequentiality decreased notably over that time, which may reflect a sentiment that the proposed conservation policies would no longer be considered by a new government. Importantly, potential survey takers were invited incrementally. As such, we may consider the survey response date to be plausibly random, and construct an instrumental variable based on the response date to identify the effects of consequentiality on valuations. Ideally, we would have information on the date each participant was invited, as it is possible that those supportive of conservation policies were more inclined to complete the survey quickly. While we do not have the date each respondent was invited, 98% of respondents participated within 48 hours of receiving their invitation and respondents had a maximum of four days to complete the survey.

To facilitate estimation, we specify the following two-equation model:

$$\begin{aligned}
 [2] \quad V_i^* &= \alpha + C_{1,i}\theta_1 + C_{2,i}\theta_2 + C_{0,i}\mathbf{x}_i\boldsymbol{\beta}_0 + C_{1,i}\mathbf{x}_i\boldsymbol{\beta}_1 + C_{2,i}\mathbf{x}_i\boldsymbol{\beta}_2 + u_i \\
 C_i^* &= \mathbf{z}_i\boldsymbol{\gamma} + \varepsilon_i \\
 \begin{pmatrix} u_i \\ \varepsilon_i \end{pmatrix} &\sim \text{Normal} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right],
 \end{aligned}$$

where  $C_{0,i}$ ,  $C_{1,i}$ , and  $C_{2,i}$  denote the consequentiality indicators, and  $C_i^*$  is latent consequentiality. The outcome equation is an interval regression and, consistent with Specification 4, allows for the effects of control variables to vary across the three consequentiality levels. The control variables are demeaned so that the model intercept can be interpreted as mean WTP (or WTA) for respondents that selected “very unlikely”, and the coefficients on the consequentiality indicators reflect differences in valuations relative to the “very unlikely” group. The second equation is an ordered probit that defines latent consequentiality as a function of control variables and the instrumental variable (survey return date; Date IV). The errors of the two equations are assumed to follow a bivariate normal distribution. The correlation coefficient,  $\rho$ , is

---

<sup>20</sup> We collected surveys over the period 21 February to 17 March. According to polls, support for the incumbent party began falling in late February, in part because the government failed to pass a new budget and over speculation that an election would take place. On March 5<sup>th</sup>, the National Assembly of Quebec was dissolved, and an election took place on April 7. The incumbent party lost the election.

the covariance between the two error terms. A covariance of zero suggests that the two equations can be treated as independent.<sup>21</sup>

Table 6 presents welfare estimates based on the five econometric specifications.<sup>22</sup> The reported estimates reflect a weighted average of values across the 35% and 50% conservation scenarios. Additional estimation results are presented in the appendix. Using Specification 1, which ignores consequentiality beliefs, the WTA/WTP ratio is 3.55. Both WTA (\$1191.31) and WTP (\$335.17) are statistically significant, and the ratio is statistically different from one.

Specifications 2 through 4 indicate that conditioning on stated consequentiality can alter welfare estimates and the WTA/WTP ratio significantly. For each specification, the point estimate of the ratio is highest for those responding “very unlikely” to the scenario-specific consequentiality question and the lowest for those indicating somewhat or very likely beliefs. The latter ratio is statistically different from the other ratios for all specifications. Moreover, each ratio is statistically different from one. Further, WTA (weakly) decreases and WTP increases with consequentiality.<sup>23</sup> Although this finding is not universal, the latter result is consistent with most WTP studies that report measures based on consequentiality (e.g., Herriges et al. 2010).

Instrumental variables estimation (Specification 5) produces welfare estimates that are similar to those from Specification 4.<sup>24</sup> The instrumental variable, Date IV, is statistically significant ( $p < 0.01$ ) in the ordered probit equation for both the WTA and WTP models, implying that the rank condition for instrumental variables estimation is satisfied. The estimated correlation coefficients are small in magnitude (0.116 for WTA and -0.002 for WTP models) and statistically insignificant. This evidence suggests that the valuation equation is independent from the consequentiality equation, and likewise that stated consequentiality beliefs may be considered exogenous in this analysis.

Across Specifications 2 to 5, for “somewhat/very likely” respondents, WTP ranges from \$427.61 to \$495.18 whereas WTA ranges from \$755.57 to \$818.57. The welfare ratio ranges

---

<sup>21</sup> This estimator is implemented using the “eintreg” command in Stata (version 17).

<sup>22</sup> Additional estimation results are provided in the appendix. To compute delta-method standard errors for the welfare ratios, we make use of the “suest” command in Stata, which is designed to test cross-model hypotheses.

<sup>23</sup> For most specifications, WTA estimates are statistically equal for “very unlikely” and “somewhat unlikely”.

<sup>24</sup> Some studies find that consequentiality beliefs vary with the cost of the policy (e.g., Groothuis et al. 2017). Including the bid/offer amount in the model (i.e., in  $\mathbf{z}_i$ ) leads to convergence problems, which may be driven by the fact that the bid/offer appears as a covariate in the model and is used to bound the respondent’s WTP (or WTA). Nevertheless, if we separately estimate an ordered probit model for  $C_i^*$ , we find that the bid/offer is an insignificant determinant of consequentiality for WTA ( $p=0.59$ ) and WTP ( $p=0.24$ ). Figures A2 and A3 in the appendix further illustrate that consequentiality responses are largely invariant to the stated bid/offer amounts.

from 1.57 to 1.91. The ratios conditional on “very unlikely” and “somewhat unlikely” are in the 3.99 to 6.48 range, coinciding well with central estimates (geometric mean of 6.23) reported in the meta-analysis by Tunçel and Hammit (2014) for environmental goods.

As a possible explanation for patterns in the data, many WTP respondents answering very or somewhat unlikely to the scenario-specific consequentiality question may believe government officials have already decided to increase conservation efforts, but that their “no” vote may influence the decision of whether or how much to tax citizens. Such behavior would lead to WTP increasing with consequentiality beliefs. A parallel phenomenon may also explain why WTA decreases with consequentiality. If those without scenario-specific consequentiality beliefs think that the government has committed to foregoing conservation and that it is using the survey only to determine whether or how much compensation should be paid, the optimal response is to vote “no” in order to influence the compensation amount upward.<sup>25</sup> An alternative or complementary explanation is that answers to the consequentiality question may be correlated with unobserved factors such as distrust of the government or other drivers of scenario rejection.

WTP estimates reported in Table 6 range from \$237.86 to \$495.18 per person per year. To put these numbers in perspective, Pattison, Boxall, and Adamowicz (2011, p. 223) report that households in the Canadian province of Manitoba are willing to pay \$296 to \$326 per household per year over a five-year period for wetland program improvements. He, Dupras, and Poder (2017, p. 17) report wetland conservation values for Quebec that vary from \$447 to \$465 per household per year. While our range of estimates overlap those from closely related studies, we note that the scope of the conservation program we examine is much larger.<sup>26</sup> Our estimates translate into \$62 to \$129 per hectare per year, which suggests our figures are conservative when compared with those from the broader wetland valuation literature (Brander et al. 2006; He et al. 2015).<sup>27</sup> Many sources of variation may explain differences in values across studies (e.g., survey design, scope of the change, assumed extent of the market, etc.). At a minimum, our estimated values do not appear singularly out of line from previous findings.

---

<sup>25</sup> These arguments rest on the assumption that respondents view the survey as consequential in the general sense that responses may influence an outcome they care about. Responses to the scenario-specific consequentiality question cannot be used to rule out such a belief.

<sup>26</sup> While we elicited individual rather than household WTP, evidence from the literature suggests that differences between the two measures are likely to be small (Delaney and O’Toole 2006; Lindhjem and Navrud 2009).

<sup>27</sup> We arrived at this figure by first multiplying the mean WTP estimate by the number of adults living in the province (6,580,865 based on the 2016 Census), and then dividing by the number of hectares conserved under the 35% scenario relative to the baseline conservation of 14%.

### 4.3. Scope effects

For both WTP and WTA, the welfare estimates across the two conservation scenarios (35% and 50%) 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. However, as discussed by Lopes and Kipperberg (2020), the current view is that there are several valid economic and statistical explanations for scope insensitivity. One or more of them may apply to our study.

First, standard consumer choice theory only assumes that the utility function is weakly increasing in the quantity of a good. Since both conservation scenarios reflect very large-scale policy changes, it is conceivable that the marginal utility from additional conservation beyond 35% of the area is near or exactly zero. Evidence from focus groups suggests that those who use the area for recreation did not perceive any additional benefit from increasing the conservation target. A similar finding of scope insensitivity was found by Bateman et al. (2005), where respondents were evaluating a return in water quality from ecologically sound levels to natural pre-industrial levels. Pattison, Boxall, and Adamowicz (2011), in a different study of wetland conservation in Canada, also find that estimates are insensitive to scope when based on a between-subjects comparison.

Second, the 35% target was suggested by ecologists as the necessary level to maintain water quality and quantity. In contrast, the 50% target arose from political discourse that took place prior to the survey. In all valuation scenarios, respondents were given information about the ecologists’ assessment of the 35% target. This could have led to perceptions that the benefits from 50% conservation were not significantly larger than those obtained from preserving 35% of the area.

Third, the instructional video and survey instrument did not precisely use outcome metrics (e.g., water quality measures, etc.) to distinguish the two conservation scenarios. Instead, the tradeoffs between conservation and development were only expressed qualitatively.

A final explanation comes from the possibility that the identification of a scope effect is confounded by those who did not perceive the conservation target to be attainable. When we restrict the sample to include only respondents who indicated a positive probability that the conservation target would be achieved, we find weak evidence that WTP increases with scope. Specifically, using Specification 4, the mean estimate of the scope effect is \$93.21 [std. err. = 50.36], \$47.84 [40.02], and \$135.81 [82.25] conditional on “very unlikely”, “somewhat

unlikely”, and “somewhat/very likely” scenario-specific consequentiality beliefs. The first and third estimates are statistically significant at the 10% level. WTA estimates continue to be insensitive to scope based on this exclusion rule.

#### **4.4. Robustness checks**

Using Specification 4 as a basis of comparison, we carry out several robustness checks. First, the WTP version of the survey states that people with very low incomes may be exempt from paying the tax. As such, respondents from low-income households may have voted as if their tax liability would be zero or otherwise lower than the stated amount. As a robustness check, we dropped respondents from the WTP sample that reported household incomes below cutoffs used by the government to define low-income households (based on household size). Dropping these respondents leads to welfare ratios that are very similar to those reported in Table 6: “very unlikely” (ratio = 6.22 [std. err. = 1.07]), “somewhat unlikely” (4.69 [0.54]), and “somewhat/very likely” (1.74 [0.29]).

Second, pretests indicated that it would be more efficient to use different bid ranges for the WTP and WTA surveys. Welfare estimates can be sensitive to the distribution of bids, and it is an open question whether our bid design choices may have unintentionally influenced results. Restricting the sample to include only the common set of bid amounts (i.e., excluding \$10 and \$50 from WTP and \$1000 and \$1500 from WTA) yields some notable differences in welfare ratios conditional on “very unlikely” (2.55 [0.38]) and “somewhat unlikely” (1.96 [0.21]) beliefs. However, the restriction has little effect on the “somewhat/very likely” ratio (1.36 [0.21]), and this ratio is statistically different from those based on the other consequentiality levels. While the directional relationship between consequentiality and the ratio persists, we stress caution in interpreting results from this restricted sample. The WTA estimates based on the full sample all lie above the highest bid included in this restricted analysis (\$750), and so the WTA estimates in this robustness check may not be well identified.

Third, as documented previously, only some respondents indicated a belief that the conservation target presented to them was attainable. Restricting the sample to include only these individuals does not significantly alter the welfare ratios across the consequentiality groups: 7.21 [1.42], 4.45 [0.53], and 1.76 [0.32]. Fourth, while we combined the “very likely” and “somewhat likely” respondents due to the small number of individuals in the former group, it is natural to

ask what happens when we allow for welfare differences across all four consequentiality levels. When we do so, consistent with expectations, the WTP point estimate is higher for the “very likely” group relative to “somewhat likely”, and the opposite is true for WTA.

Finally, when constructing welfare intervals based on individual votes, we assumed a lower bound of \$0.<sup>28</sup> It is plausible that some respondents in the WTP frame favor development over conservation and may need compensation to agree to the conservation program. Negative WTA is less likely, since respondents in this scenario hold property rights and are potentially entitled to compensation; whether they value conservation is in a sense secondary. Dropping the lower bound restriction of \$0 for WTP leads to an estimate that is not statistically different from zero for the “very unlikely” group. Welfare ratios conditional on “somewhat unlikely” and “somewhat/very likely” become 5.88 [2.18] and 0.92 [0.25], respectively.<sup>29</sup>

Across these robustness checks, the “somewhat/very likely” WTA/WTP ratio remains less than 2. Furthermore, while the welfare ratios for the other consequentiality levels are more variable, the robustness checks clearly support the notion that controlling for scenario-specific consequentiality matters to the estimation of WTA/WTP ratios.

## 5. Discussion

The motivation to explore the behavioral underpinnings of stated preferences and subjects’ beliefs is now firmly grounded in a theory of incentive compatibility for advisory surveys. A consequential stated preference survey employing a single binary choice elicitation mechanism is known to be incentive compatible. Within this framework, we have shown empirically that controlling for the degree of consequentiality in a manner tied to the specific game form constructed in the survey can greatly and significantly reduce the estimated WTA/WTP ratio.

The key insights from our application to wetland conservation in Quebec can be summarized as follows. First, for respondents who most likely viewed the value elicitation scenario as consequential, the WTA estimate for our wetland policy is less than double the

---

<sup>28</sup> We nevertheless assume welfare measures follow normal distributions, and do not restrict estimated welfare distributions to be non-negative.

<sup>29</sup> In an earlier working paper version of this article (Vossler et al. 2020), we reported estimates from econometric models that do not impose a lower bound of \$0 for both WTA and WTP. This results in insignificant welfare ratios for those with “very unlikely” and “somewhat unlikely” beliefs, and WTA point estimates that are implausible in magnitude. WTA (and WTP) estimates conditional on “somewhat/very likely” remain statistically significant, and the welfare ratio for this consequentiality level is approximately 1.2 across a range of specifications.



comparable WTP estimate. This welfare ratio is substantially smaller than those reported in previous studies of large-scale environmental public goods. Second, respondents answering “very unlikely” or “somewhat unlikely” to our scenario-specific consequentiality question produce WTA estimates many times larger than WTP. The corresponding WTA/WTP ratio is in line with most studies in the stated preference literature.

Third, past studies comparing WTA and WTP generally do not produce data that should be relied upon to assess the validity of stated preference methods. The use of non-incentive compatible mechanisms and the failure to control for consequentiality and other game form beliefs provide plausible explanations for the suspiciously large WTA/WTP ratios previously observed. To the extent that evidence from prior WTA-WTP comparisons has discouraged researchers from employing WTA questions, our results suggest that WTA surveys might be credibly considered when property rights make it the natural choice.

Fourth, our empirical results complement a growing literature on the relationship between consequentiality and other factors that influence WTP estimates. This literature suggests that WTP respondents who do not view the survey as consequential hold, on average, lower values than those who do. By analogue, the forces driving this empirical finding suggests that WTA estimates should display the opposite tendency. Our empirical results support this conjecture, as we find WTA decreases with scenario-specific consequentiality.

Fifth, the theoretical framework recasts a lack of consequentiality as one of many possible expressions of game form misconceptions. Strictly defined in a game-theoretic framework, consequentiality only ensures that respondents believe that their answer has the potential to affect outcomes and their welfare without restrictions on how they think it might do so. Translating the information obtained in the survey into valid value estimates also requires (a) an incentive compatible mechanism, (b) respondents to understand the incentive structure associated with the elicitation format, and (c) respondents to have beliefs about the game form – both consequentiality and key aspects of the valuation scenario - that correspond to those intended by the researcher.

In our opinion, further progress can be made by better influencing, identifying, and controlling for respondents’ beliefs about the game they are playing, in a parallel approach with the experimental work of Plott and Zeiler (2005; 2011), Cason and Plott (2014), and others (e.g., Isoni, Loomes, and Sugden 2011; Bartling, Engl, and Weber 2015; Fehr, Hakimov, and Kübler

2015; Bull et al. 2019). For both stated preference surveys and experiments, there is a growing recognition that good mechanism design is necessary but insufficient to ensure data quality and reliability. Mechanism design theory assumes that respondents recognize and accept the game form presented to them. When people hold different beliefs about the form of the game being played, it can no longer be guaranteed that their decisions reveal their true preferences about the good of interest to the researcher.

It has been a long-standing practice in the stated preference literature to use questions to better understand respondent motivations, in the form of “scenario adjustment”, “scenario rejection”, “protest responses”, “consequentiality”, “payment credibility”, “trust in institutions”, and so on. Investigations often consider just one or a few of these potential measures of attitudes/beliefs. If the goal is to obtain an estimate of the value the good as described in the survey, it seems imperative to take a more holistic view and look into these related issues through the lens of game form misconceptions. It stands to reason that the various motivations that result in a mismatch between survey scenario and respondent beliefs about the game they are playing are connected in important ways. For instance, those who distrust the government may be less likely to believe that the survey is consequential, or view the survey as a way to influence payment but not provision.

The extent to which researchers should minimize the effects of game form misconceptions depends on the goal of the research. One possible objective is to obtain theoretically sound and empirically accurate estimates of Hicksian welfare measures. Under this high standard, a lack of consequentiality and other significant deviations from the game form that cause systematic biases in stated preferences must be minimized or controlled for. On the other hand, if the objective is to replicate how people would vote in an analogous but binding public referendum (i.e., a revealed preference criterion), this imposes weaker requirements. In both binding and advisory referenda, voters are free to believe that the objectives of the proposal may not be met or vote “no” if they object to the payment vehicle.

In our survey, we used a single belief question that encompasses consequentiality and broader adherence to the game form. A “very unlikely” response to our question could indicate either inconsequentiality or disbelief that the decision making process will be confined to the two options presented in the scenario (or both). Different responses to our question may reflect different degrees and types of deviations from the survey game form. Under this interpretation,

our belief question does not have the knife-edge property of theory whereby respondents with any positive levels of perceived consequentiality should consider the scenario in the same way.

One might argue that conditioning the consequentiality question directly on the valuation scenario, as we have done, is a simple approach to controlling for consequentiality and adherence to the game form simultaneously. However, the counterargument is that separate questions about consequentiality and other game form misconceptions are more likely to allow researchers to identify distinct sources of variations in valuation responses across respondents.

A way forward may be to use multiple questions but ensure that they can identify inconsequentiality as well as other deviations from game form. The current practice of using separate “policy consequentiality” and “payment consequentiality” questions is incomplete in this regard. For instance, a respondent may perceive the survey to be influential in both dimensions and thus answer in the affirmative to both consequentiality questions. Yet, this does not rule out the possibility that a respondent supports the presented policy, but believes that voting “no” could lead to a different level of provision that she prefers, or might lower the cost that she ultimately must pay (or compensation received). Current consequentiality questions do not capture such deviations from game form, and respondents holding such beliefs may not truthfully reveal their preferences for the policy described in the survey. Developing a set of questions that more precisely identify deviations from game is an important topic for future research.

Even with important advancements in survey methodology, it is unlikely that all or nearly all survey respondents will hold consequential beliefs and be free from game form misconceptions. This makes it paramount for researchers to design surveys that not only objectively motivate these beliefs but provide the means to causally identify the effect of beliefs on valuations. Acquiring information about relevant attitudes and ideologies that extend beyond the survey application is likely to supply useful data from which to explain beliefs. A complementary approach could be to randomly vary the information content of the survey across respondents, as has been done in prior applications (e.g., Herriges et al. 2010), in ways that enhance consequentiality and other beliefs. Indicators related to information treatments may then be used as instrumental variables for belief measures.

As emphasized by a careful reviewer of this research, asking questions related to consequentiality or other beliefs about the game form may violate conversational norms. This

may occur, for instance, when the survey instrument stresses consequentiality, and then asks respondents if they believe the survey to be consequential. The common practice of asking participants about their beliefs after the fact may introduce considerable errors in the measurement of those beliefs. Eliciting beliefs prior to the valuation scenario (as proposed by Lloyd-Smith, Adamowicz, and Dupont 2019) may help induce adherence to game form, and in turn enhance the validity of stated preferences. Yet, probing beliefs before the valuation question could also introduce doubts and raise new uncertainties in the minds of some respondents. Understanding how to best gauge respondent beliefs to obtain the most accurate valuation answers and to simultaneously favor the best possible empirical controls for those beliefs remain important areas for future research.

A considerable fraction of our sample thought it “very unlikely” or “somewhat unlikely” that the survey would inform whether the policy as described would be implemented. This may limit the generalizability of our findings, as well as the usefulness of our welfare estimates for policy and regulatory analyses. Furthermore, our consequentiality question provides an incomplete understanding of how and why respondents’ beliefs deviate from game form. It is possible that some sources of deviations have little impact on valuations, and that some who answered “very unlikely” or “somewhat unlikely” may have responded truthfully to the voting scenario. Feedback from focus groups suggests that our WTA payment vehicle was credible, and the distribution of responses to our consequentiality question is similar across WTP and WTA frames. Nevertheless, the relatively small fraction of WTA studies in the literature suggests that important obstacles to their implementation remain.

In their review of the literature and in their own CV study of environmental damage to a Spanish wetland of international significance, Brouwer and Martin-Ortega (2012) observe that up to 20% of respondents refused to pay any amount on the basis that polluters should be liable for the pollution they cause.<sup>30</sup> Using a WTA framing in circumstances when it is the appropriate measure may increase accuracy, reliability and validity by increasing consequentiality and greater acceptance of the game form.

---

<sup>30</sup> Since our study involved a good over which there was less controversy, and where either the WTP or WTA framing could realistically be justified, we have no reason to believe that the particular type of protest votes identified by Brouwer and Martin-Ortega (2012) affected our study.

Yet, a common concern is that a WTA scenario is more difficult to design in a credible way (Carson 2000). One possible explanation for this is that most people have experienced paying the government to provide public goods while fewer have been compensated for a lack of provision. The open-ended and related elicitation formats used in many past WTA studies naturally pose a threat to credibility, e.g., by suggesting that people can bargain over compensation with the government. To the extent that perceptions over WTA credibility have been confounded by the choice of elicitation format, the use of incentive compatible methods may dampen this concern.

On a final note, we acknowledge that it may be impractical to deploy incentive compatible elicitation mechanisms in some settings. This includes situations where surveys are used to estimate the WTA of landowners to undertake conservation efforts, to estimate values in settings where payment cannot be mandated, and when valuing goods and services prior to making them available in markets. In a study involving the valuation of a quasi-public good using a non-coercive payment vehicle, Frondel, Sommer, and Tomberg (2021) find that the WTA/WTP ratio decreases with beliefs associated with what the authors label as “perceived realism”. This provides suggestive evidence that it is important to understand the incentive properties of surveys and to control for game form misconceptions, even in contexts where incentive compatible elicitation is impractical.

## 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.
- Bartling, Björn, Florian Engl, and Roberto A. Weber. 2015. Game form misconceptions are not necessary for a willingness-to-pay vs. willingness-to-accept gap. *Journal of the Economic Science Association* 1: 72-85.
- 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 Zawojka. 2021. Payment and policy consequentiality in dichotomous choice contingent valuation: Experimental design effects on self-reported perceptions. *Land Economics* 97(2): 407-424.
- 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 and 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.
- Brouwer, Roy, and Julia Martín-Ortega. 2012. Modeling self-censoring of polluter pays protest votes in stated preference research to support resource damage estimations in environmental liability. *Resource and Energy Economics* 34(1): 151-166.
- Bull, Charles, Pascal Courty, Maurice Doyon, and Daniel Rondeau. 2019. Failure of the Becker-DeGroot-Marschak mechanism in inexperienced subjects: New tests of the game form misconception hypothesis. *Journal of Economic Behavior & Organization* 159: 235-253.
- Bush, Glenn, Nick Hanley, Mirko Moro, and Daniel Rondeau. 2013. Measuring the local costs of conservation: A provision point mechanism for eliciting willingness-to-accept compensation. *Land Economics* 89(3):490-513.
- Cameron, Trudy A., J.R. DeShazo, Erica H. Johnson. 2011. Scenario adjustment in stated preference research. *Journal of Choice Modelling* 4(1): 9-43.
- 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. 2000. Contingent valuation: A user’s guide. *Environmental Science and Technology* 34: 1413-1418.

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., W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert Cameron Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith. 1994. Prospective interim lost use value due to PCB and DDT contamination in the Southern California Bight. La Jolla, CA: US Department of Commerce (NOAA).

Carson, Richard T., Robert C. Mitchell, Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 2003. Contingent valuation and lost passive use: Damages from the Exxon Valdez oil spill. *Environmental and Resource Economics* 25(3): 257-286.

Cason, Timothy N., and Charles R. Plott. 2014. Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy* 122(6): 1235-1270.

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.

Chou, Eileen, Margaret McConnell, Rosemarie Nagel, and Charles R. Plott. 2009. The control of game form recognition in experiments: Understanding dominant strategy failures in a simple two person “guessing” game. *Experimental Economics* 12(2): 159-179.

Delaney, Liam, and Francis O’Toole. 2006. Willingness to pay: Individual or household? *Journal of Cultural Economics* 30(4): 305-309.

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.

- Gibbard, Allan. 1973. Manipulation of voting schemes: A general result. *Econometrica* 41(4): 587-601.
- Groothuis, Peter A., Tanga M. Mohr, John C. Whitehead, and Kristan Cockerill. 2017. Endogenous consequentiality in stated preference referendum data: The influence of the randomly assigned tax amount. *Land Economics* 93(2): 258-268.
- 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.
- He, Jie, Jérôme Dupras, and Thomas G. Poder. 2017. The value of wetlands in Quebec: A comparison between contingent valuation and choice experiment. *Journal of Environmental Economics and Policy* 6(1): 51-78.
- He, Jie, Fanny Moffette, Richard Fournier, Jean-Pierre Revéret, Jérôme Théau, Jérôme Dupras, Jean-Philippe Boyer, and Mathieu Varin. 2015. Meta-analysis for the transfer of economic benefits of ecosystem services provided by wetlands within two watersheds in Quebec, Canada. *Wetland Ecology and Management* 23(4): 707-725.
- Henderson, A. 1941. Consumer's surplus and the compensating variation. *The Review of Economic Studies* 8(2): 117-121.
- 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.
- Hicks, John R. 1943. The four consumer's surpluses. *The Review of Economic Studies* 11(1): 31-41.
- 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. *The 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.

Lindhjem, Henrik, and Ståle Navrud. 2009. Asking for individual or household willingness to pay for environmental goods? *Environmental and Resource Economics* 43: 11-29.

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.

Lloyd-Smith, Patrick, Wiktor Adamowicz, and Diane Dupont. 2019. Incorporating stated consequentiality questions in stated preference research. *Land Economics* 95(3): 293-306.

Lopes, Ana Faria, and Gorm Kipperberg. 2020. Diagnosing insensitivity to scope in contingent valuation. *Environmental and Resource Economics* 77(1): 191-216.

Osborne, Martin J., and Ariel Rubinstein. 1994. *A Course in Game Theory*. Cambridge, MA: The MIT Press.

Pattison, John, Peter C. Boxall, and Wiktor L. Adamowicz. 2011. The economic benefits of wetland retention and restoration in Manitoba. *Canadian Journal of Agricultural Economics* 59: 223-244.

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.

Plott, Charles R., and Kathryn Zeiler. 2011. The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations: Reply. *American Economic Review* 101(2): 1012-1028.

Randall, Alan, John P. Hoehn, and David S. Brookshire. 1983. Contingent valuation surveys for evaluating environmental assets. *Natural Resources Journal* 23(3): 635-648.

Randall, Alan, and John R. Stoll. 1980. Consumer's surplus in commodity space. *The American Economic Review* 70(3): 449-455.

Satterthwaite, Mark. 1975. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10(2): 187-217.

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., Stéphane Bergeron, Maurice Doyon, and Daniel Rondeau. 2020. Revisiting the gap between the willingness-to-pay and willingness-to-accept for public goods. CIRANO Working Papers, 2020s-48.

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.

Willig, Robert D. 1976. Consumer's surplus without apology. *The American Economic Review* 66(4): 589-597.

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

Zawojcka, 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.

**Table 1**  
Characterization of prior literature

Study	Type of good	Survey mode	Elicitation format	Provision mechanism	Comparison type	Controls for beliefs?
Aabø 2005	Other public or non-market	Field	Format 1: DM + OE Format 2: MBDC + OE	Public	Between	No
Amigues et al. 2002	Environmental	Field	WTA: OE WTP: SBC or OE	Private	Between	No
Bishop, Heberlein, and Kealy 1983	Other public or non-market	Field	SBC + OE	Private	Within	No
Bowker and MacDonalød 1993	Environmental	Field	PC	Public	Between	No
Brookshire, Randall, and Stoll 1980	Other public or non-market	Field	Iterative bidding	Private	Within	No
Brookshire and Coursey 1987	Environmental	Field	Survey 1: PC Survey 2: Smith Auction	Public	Between	No
Chiwaula et al. 2016	Health and safety	Field	OE	Private	Within	No
Carthy et al. 1999	Health and safety	Field	Chained approach	Private	Within	No
Chapman and Johnson 1995	Environmental; Health and safety; Other public or non-market	Lab	OE	Private	Within	No
Chilton et al. 2012	Health and safety	Lab	OE*	Private	Within	No
Del Saz-Salazar et al. 2009	Environmental	Field	WTA: OE WTP: SBC + OE	Public	Within	No
DuBourg, Jones-Lee, and Loomes 1994	Health and safety	Field	Stage 1: Iterative bidding Stage 2: PC	Private	Within	No
Flachaire, Holland, and Shogren 2013	Other public or non-market	Lab	OE*	Public	Between	No
Garbacz and Thayer 1983	Other public or non-market	Field	OE	Private	Between	No
Gerking, De Haan, and Schulze 1988	Health and safety	Field	PC	Private	Between	No
Gleason-Comstock et al. 2017	Health and safety	Field	OE	Private	Within	No
Griffin and Mjelde 2000	Other public or non-market	Field	OE	Public	Between	No
Guria et al. 2005	Health and safety	Field	Iterative bidding	Private	Within	No

Hajek and Stejskal 2015	Other public or non-market	Field	OE	Private	Within	No
Hartman et al. 1991	Other public or non-market	Field	OE	Private	Within	No
Huang et al. 2013	Environmental	Field	PC	Private	Within	No
Kufeoglu and Lehtonen 2015	Other public or non-market	Field	OE	Private	Within	No
Martin-Fernandez et al. 2017	Health and safety	Field	Double PC	Private	Within	No
Martin-Fernandez et al. 2013	Health and safety	Field	Double PC	Private	Within	No
McDaniels 1992	Health and safety	Field	Exp 1: OE Exp 2: SBC	Private	Exp 1: Within Exp 2: Between	No
Merkle et al. 2017	Other public or non-market	Field	OE	Private	Between	No
Nataf and Wallsten 2013	Other public or non-market	Field	OE	Private	Between	No
Navrud and Mungatana 1994	Environmental	Field	WTA: OE WTP: OE and PC	Private	Within	No
Nepal et al. 2018	Environmental	Field	SBC	Public	Between	No
O'Brien et al. 1998	Health and safety	Field	Iterative bidding	Private	Within	No
Painter et al. 2002	Other public or non-market	Field	OE	Public	Within	No
Petrolia and Kim 2011	Environmental	Field	SBC	Public	Between	Yes
Rowe et al. 1980	Environmental	Field	Iterative bidding	Public	Within	No
Schulze et al 1986	Environmental	Field	OE	Private	Between	No
von Selasinsky et al. 2017	Other public or non-market	Field	OE	Private	Within	No
Shefrin and Caldwell 2001	Other public or non-market	Lab	OE	Private	Study 1: Between Study 2: Within	No
Sun et al. 2005	Other public or non-market	Field	SBC	Private	Within	No
Tanrivermis 1998	Environmental	Field	OE	Public	Within	No
Tomohara 2005	Other public or non-market	Field	PC	Private	Within	No
Van de Berg et al. 2005	Health and safety	Field	Sample 1: SBC + OE Sample 2: OE	Private	Within	No
Van Kooten and Schmitz 1992	Other public or non-market	Field	OE	Private	Within	No
Venkatachalam and Narayanamoorthy 2012	Other public or non-market	Field	OE	Private	Between	No
Viscusi, Magat, and Huber 1987	Health and safety	Field	OE	Private	Within	No
Viscusi and Huber 2012	Health and safety	Field	Iterative bidding	Public	Between	No

Whynes and Sach 2007	Health and safety	Field	Iterative bidding	Private	Within	No
----------------------	-------------------	-------	-------------------	---------	--------	----

*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. *Controls for beliefs:* “Yes” denotes studies that control for consequentiality beliefs. 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).

**Table 2**

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

Socioeconomic variables	Sample (%)	Population (%)
<b>Gender</b>		
Male	55	49
Female	45	51
<b>Age</b>		
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
<b>Education</b>		
No diploma	2	17
High school or trade school	33	40
College or CEGEP	27	18
University	38	25
<b>Location</b>		
Quebec City	14	10
Montreal	51	50
Other	35	40
<b>Household income (Can\$)</b>		
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**  
Data description

<b>Variable</b>	<b>Description</b>	<b>Mean</b>	<b>Std. Dev.</b>
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
Very unlikely	= 1 if respondent selected “very unlikely” to the scenario-specific consequentiality question	0.24	0.43
Somewhat unlikely	= 1 if respondent selected “somewhat unlikely” to the scenario-specific consequentiality question	0.52	0.50
Somewhat/very likely	= 1 if respondent selected “somewhat likely” or “very unlikely” to the scenario-specific consequentiality question	0.23	0.42
Date IV	Day respondent completed survey, 1 to 25 (=1 for first day of survey returns)	8.24	7.39

**Table 4**  
 Scenario-specific consequentiality beliefs by welfare frame

	<b>WTA</b>	<b>WTP</b>
“very likely”	4.5%	3.6%
“somewhat likely”	16.7%	22.1%
“somewhat unlikely”	54.5%	49.8%
“very unlikely”	24.3%	24.5%

*Notes:* The table reports the percentage of respondents selecting the indicated option to the question “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?”



**Table 5**  
 Sample means by scenario-specific consequentiality beliefs

<b>Variable</b>	<b>Very unlikely</b>	<b>Somewhat unlikely</b>	<b>Somewhat/very likely</b>	<b><i>p</i>-value</b>
Scope	0.52	0.49	0.50	0.799
Charity	0.66	0.71	0.70	0.398
Env Org	0.04	0.05	0.05	0.625
Male	0.56	0.52	0.59	0.137
College	0.65	0.63	0.64	0.866
Homeowner	0.76	0.71	0.69	0.220
HH Size	2.86	2.80	2.76	0.712
Retired	0.07	0.10	0.08	0.292
Student	0.16	0.18	0.22	0.245
Quebec	0.18	0.12	0.15	0.114
Montreal	0.46	0.53	0.50	0.208

*Notes:* The last column presents the *p*-value associated with a Pearson's chi-square test of the hypothesis that the distribution of the variable is independent of consequentiality beliefs.

**Table 6**  
Welfare estimates (Can\$), by scenario-specific consequentiality beliefs

	WTA	WTP	WTA/WTP
<i>Specification 1: unconditional</i>			
Unconditional	1191.31*** (74.90)	335.17*** (18.47)	3.55*** (0.30)
<i>Specification 2: conditional</i>			
Very unlikely	1160.13*** (157.35)	238.28*** (28.26)	4.87*** (0.88)
Somewhat unlikely	1451.92*** (126.11)	336.08*** (25.11)	4.32*** (0.49)
Somewhat/very likely	755.57*** (96.73)	475.38*** (55.37)	1.59*** (0.28)
<i>Specification 3: control variables</i>			
Very unlikely	1194.18*** (140.74)	248.80*** (27.23)	4.80*** (0.77)
Somewhat unlikely	1503.98*** (137.20)	328.88*** (23.28)	4.57*** (0.53)
Somewhat/very likely	779.42*** (90.54)	495.18*** (56.55)	1.57*** (0.26)
<i>Specification 4: regression adjustment</i>			
Very unlikely	1498.59*** (176.45)	237.86*** (27.39)	6.30*** (1.04)
Somewhat unlikely	1461.38*** (127.05)	328.47*** (23.40)	4.45*** (0.50)
Somewhat/very likely	765.33*** (87.07)	483.37*** (57.29)	1.58*** (0.26)
<i>Specification 5: instrumental variables</i>			
Very unlikely	1629.44*** (187.39)	251.44*** (76.97)	6.48*** (2.12)
Somewhat unlikely	1321.02*** (89.93)	331.08*** (21.23)	3.99*** (0.37)
Somewhat/very likely	818.57*** (172.00)	427.61*** (77.33)	1.91*** (0.53)

Notes: \*, \*\*, and \*\*\* denote estimates that are statistically different from zero at the 10%, 5%, and 1% significance levels, respectively. Standard errors in parentheses. The text provides details on these model specifications and estimation methods.