GREDI
Groupe de Recherche en Économie
et Développement International

Cahier de recherche / Working Paper
18-02

Payment and Provision Consequentiality in Voluntary Contribution Mechanism: Single or Double "Knife-Edge" Evidence?

Jie HE
Jérôme DUPRAS
Thomas PODER

UNIVERSITÉ DE
SHERBROOKE

Payment and Provision Consequentiality in Voluntary Contribution Mechanism: Single or Double "Knife-Edge" Evidence?

Jie He, Département d'Économique, École de Gestion, Université de Sherbrooke, 2500 boulevard de l'Université, Sherbrooke, QC, J1K2R1, Canada. Email : jie.he@usherbrooke.ca
Jérôme Dupras, UQO, jerome.dupras@uqo.ca
Thomas Poder, CRCHUS, tpoder.chus@ssss.gouv.qc.ca

This version May 14th 2018

Abstract

We conducted a field stated preferences survey to understand the joint and separate effects of payment and provision consequences on hypothetical bias associated with voluntary contribution. Based on four treatment groups and a contingent-ranking willingness to pay (WTP) question, this paper provides some support for "single" knife-edge evidence, which suggests that a respondent facing positive provision consequences will report a significantly higher preference only if the payment consequence is co-presented. For the payment consequence, its negative impact on WTP was independent on the presence of provision consequence; we therefore reject the "double" knife-edge evidence.

Keywords: hypothetical bias, knife-edge evidence, payment consequence, provision consequence, contingent ranking, voluntary donation, ecosystem services valuation.

1. Introduction

Despite their popularity in eliciting willingness to pay (WTP) for environmental public goods and services, stated preference methods are known to be influenced by biases due to their hypothetical nature. Since the 1960s, substantial research has been carried out on the presence and magnitude of hypothetical bias in stated preference studies (e.g., Bohm, 1972; Seip and Strand 1992; Cummings et al., 1995, 1997; Neil et al., 1994, Champ et al. 1997; Foster et al. 1997; Byrnes et al. 1999; MacMillan et al. 1999; Brown and Taylor, 2000; Champ and Bishop 2001; Poe et al. 2002; Murphy et al. 2005; Christie 2007; Alpizar et al. 2008). The meta-analyses by List and Gallet (2001) and Little and Berrens (2004, with a larger database) reported average overestimation by a calibration factor of approximately 3 in hypothetical settings. The meta-analysis by Murphy et al. (2005) suggested a smaller median ratio of about 1.35 for hypothetical to actual value.

Carson and Groves (2007) advocated the stated preference methods by identifying consequential hypothetical scenarios. According to them, consequential surveys should satisfy two conditions. First, the respondent must view his responses as potentially influencing the agency's actions. Second, the respondent must care about the outcomes of those actions. In this setting, Carson and Groves (2007) discussed how a consequential survey question in a dichotomous, incentive-compatible choice referendum format can encourage respondents to answer truthfully, thereby avoiding hypothetical bias. Similarly, Vossler et al. (2013) confirmed that the two conditions of consequentiality also apply to a consequential discrete choice experiment study based on a game theory model. Most recent empirical studies, such as those by Landry and List (2007), Vossler and Evans (2009), Mitani and Flores (2009), and Vossler et al. (2013), provided supportive evidence in line with a proposition from Carson and Groves (2007). Based on field surveys or lab/field experiments, these studies generally concluded that "no significant hypothetical bias" occurred once the incentive-compatible referendum stated preference questions became consequential.

Herriges et al. (2010) explicitly interpreted the two consequentiality conditions of Carson and Groves (2007) into provision consequentiality, which suggests that respondents must believe that the results of the survey might influence an outcome they care about, as well as payment consequentiality, underlying that the respondent must perceive that there is some probability that they will have to pay. Several studies focused on "payment consequentiality", also known as "binding probability" (Cummings and Taylor, 1998; Landry and List, 2007; Mitani and Flores, 2009, etc.). These studies generally revealed a significant reduction in reported WTP for higher probability of real payment. Several other studies

examined the impact of "provision probability", also called "policy consequentiality", on hypothetical bias, where results were less consistent. Some studies (Nepal et al. 2009; Vossler and Waston 2013) seemed to confirm a positive effect of (perceived) provision probability, while others (Zawojska et al. 2015; Oehlmann and Meyerhoff 2017) did not find the WTP estimate varied with perceived provision consequentiality.

One possible explanation for such inconsistency in the studies on the impact of "provision probability" is associated tois the so-called "knife-edge evidence". First coined by Herriges et al. (2010), "knife-edge evidence" implies that "as long as the respondent believes that their answers will be consequential in *both* senses [payment and provision] with *any* probability, their dominant strategy will be to answer truthfully" (Herriges et al. 2010, p. 68). A series of studies, mostly based on lab experimental designs, have provided supportive elements for this evidence. Vossler and Evans (2009) found that under a financially binding scenario, once respondents' perceived consequentiality was controlled to be positive, their answers to the WTP question would be significantly different from those obtained under a purely hypothetical scenario, but similar between the provision probabilities of different levels above zero. Vossler et al. (2012) conducted a field experiment to compare three probabilistic provision rules. Their findings also showed that variations in provision rules did not play an important role in marginal utility when the experimental treatments were financially binding. Taylor et al. (2010) tested the influence of different provision rules and arrived at the same finding: what matters is not the type of provision rule, but the mere fact that there is one. This conclusion is further confirmed by Carson et al. (2014). Based on the results of the framed field experiments, they concluded that, once the experiment was set to be financially binding, the incentive structure of a referendum would not be altered "as long as the decision makers are likely to undertake the referendum proposed outcome if the specified plurality favor it", regardless of whether the provision consequence was deterministically ($p=1$) or stochastically binding ($0<p<1$).

Although the initial focus of "knife-edge evidence" is the potential equivalence between a partial provision consequence ($P>0$) with respect to the full provision consequence ($P=1$) in motivating true WTP revelation, one commonality among the abovementioned studies is their control in the experiments to ensure that respondents face some effective binding payment. One direct question that we can therefore ask will be this: is the presence of the payment consequence a *necessary* condition to ensuring the effectiveness of the provision consequence? On the other hand, we may also verify if the presence of the provision consequence is a necessary condition for the payment consequence to be effective. If one of the co-presence conditions holds, we will observe "single" knife-edge evidence. In the case where the co-presence conditions hold for

both consequentiality conditions, we can consider our results as supporting "double" knife-edge evidence. This will be the central research question of our paper.

This paper is based on a field stated preference survey with randomly invited respondents from the general population of Quebec, Canada. In this survey, people's preference for three real environmental projects in the Greater Montreal Blue Network was elicited by a contingent-ranking question based on voluntary contribution as payment vehicle. We explored the potential impacts of payment and provision consequences separately and jointly via four exogenous treatments. These arrangements allowed us to verify whether "knife-edge" evidence also applies to the voluntary contribution mechanism and provided an opportunity to see whether these two consequences are mutually pre-conditions for the effectiveness of the other dimension of consequentiality.

Most of the studies providing supportive "knife-edge" evidence have been based on an incentive compatible referendum WTP question format with a coercive payment vehicle. The issue of consequentiality for other WTP elicitation formats is rarely addressed in the literature. This is certainly related to the fact that many other elicitation formats suffer from incentive incompatibility issues (Carson and Groves, 2007). However, a general accepted assumption, applicable to both the incentive compatible binary referendum and all other elicitation formats, is that under an inconsequential scenario, as the probability of influencing the decision becomes zero, any response has the same influence on the respondent's utility (Carson et al. 2014). As truthful preference revelation is no longer a dominant strategy for both incentive-compatible or incompatible elicitation formats WTP under such contexts, it is a necessary condition for theory to be capable to making predictions about the direction and magnitude of the hypothetical bias in the other WTP elicitation formats, even if the consequentiality is not helpful for incentive issues (Interis and Petrolia, 2014). Our intention to test the hypothetical bias-related issues in incentive-incompatible WTP elicitation formats can also be motivated by the fact that some of these elicitation formats are still frequently employed in non-market valuation to provide policy guidance. We therefore believe a better understanding of the impact of consequentiality on these formats will be useful.

We found only two studies that treat the consequentiality of the other WTP elicitation format. Carlsson and Matinsson (2001) tested the external validity of choice experiments with donations for environmental projects and found no differences in preference between

hypothetical and actual scenarios. Admitting their test of external validity could not be seen as a test of truthful revelation of hypothetical bias due to incentive issues, they regarded their study as a test of external validity between hypothetical and actual experiments. Interis and Petrolia (2014) compared the potential difference in the effect of consequentiality between binary and multinomial choice surveys. They reported knife-edge evidence only for the multinomial choice data but not for the binary choice data, which revealed the potentially important role played by consequentiality on hypothetical bias in other WTP elicitation formats, incentive-compatible or not.

2. Survey and data

The real context played a major role in the execution of our study and in setting up the experiment. We used three carefully selected projects actually promoted by Nature Action Quebec (NAQ) in the Greater Montreal Blue Network in our stated preference survey. NAQ is a thirty-year-old non-profit organization that has made significant contributions to the implementation of innovative projects in several environment-related areas in partnership with municipalities, non-profit organizations, government agencies, communities, and private companies. Today, NAQ is one of Quebec's leading environmental project management organizations.

The three real projects were carefully selected with the help of five field experts (Poder et al. 2016). We used a contingent-ranking (CR) question to invite each respondent to rank the three real projects (opt-in) and an opt-out option (i.e., status quo). To make the comparison between projects easier, each restoration project was presented in terms of the potential improvements it could provide towards six ecosystem services: biodiversity, water quality, carbon sequestration, recreational activities, landscape aesthetics, and education. These ecosystem services were identified with the help of a focus group of experts (four biologists and one land-use planner of NAQ), based on a literature review of studies focusing on freshwater systems. The potential levels of improvement for each ecosystem service before and after implementing the restoration project were estimated according to the expected restoration efforts and related literature review. In addition to these ecosystem services, we considered two extra attributes in the survey: the size of the restoration area and the suggested average cost at household level that indicates the average amount that a donor should contribute to enable NAQ to raise the necessary funds and launch the corresponding restoration project. For the opt-out status quo

option, the variations of all attributes (i.e., ecosystem services and restoration area) were set to zero and associated with a zero-dollar contribution. Table A1 in the Annex of this paper provides the details on the attributes pertaining to each of the three projects.

We used the CR WTP question in our survey after considering several aspects. First, this choice was related to the fact that we used real projects that were under consideration by NAQ and already known by some citizens. Compared to a repeated experimental design, using a real project can make the provision consequence more credible and render our invitation for real donations at the end of the survey more natural. Second, compared to the standard multiple choice experiment that requires respondents to select the most preferred alternatives out of k>2 alternatives, the contingent-ranking question only asks individuals to compare and rank their preference, each describing a different tradeoff between the provision of an environmental good and its price (Lareau and Rae, 2001). We believe such a ranking procedure can mitigate the well-known strategic behavior observed in multi-candidate races with a simple plurality winner, in which the strategy of an elector is reduced to a binary choice between the two alternatives that they believe receive the most votes, independent of the agent's actual first choice (Carson and Groves 2007). Third, we also hope that CR with unique average cost information for each project can help to avoid the potential endogeneity of the provision consequence related to the biding process used in a CV design, which risks polluting the exogenous consequence treatments. Such endogeneity has been reported by Groothuis et al. (2017), who found that a higher bid price was interpreted by some respondents as a signal of higher provision probability. Certainly we recognize the potential complexity related to the CR WTP question. The ranking process may be difficult for respondents due to its information overload associated with the high number of alternatives and complex attributes (Lareau and Rea, 2001; Foster and Mourato 2002). Such difficulties may also cause unstable ranking results: Ben-Akiva et al. (1992) found that the stability of ranking information decreases with decreasing rank for a raking of 4 alternatives. Foster and Mourato (2002) indicated the necessity to incorporate procedures for screening out inconsistent responses in ranking data, but they also believe that if the primary focus is on estimating marginal rates of substitution between attributes, the ranking inconsistency may no longer represent a problem.

We adopted the strategy of split samples to form an exogenous experimental design and thereby controlled the two dimensions of consequentiality (i.e., provision and payment consequences). Using exogenous consequentiality treatments allowed us to control the possible endogeneity in

answers given by respondents about their subjective perception of consequentiality in the surveys (Herriges et al. 2010; Interis and Petrolia 2014). For the provision consequence, half of the questionnaires explicitly mentioned NAQ to highlight the credibility of provision: "The projects that you will evaluate are currently under the consideration of the non-profit organization Nature-Action Québec (NAQ)." A detailed introduction to NAQ was then given to the respondents. With respect to payment consequence, half the respondents received a questionnaire mentioning, at the beginning of the contingent-ranking question, that "at the end of the survey, you will be invited to make a donation to your most preferred project." In the treatment where both payment and provision consequences are presented, the donation is explicitly mentioned to be made to NAQ. Following ethical principles, at the end of the questionnaire (for all four treatments), we presented a debriefing with an alteration to consent requirement to explain the purpose behind our research:

> *Although the presented projects are real, the question regarding donation that we presented earlier was totally hypothetical. Under no circumstances will we contact you to make a donation. We only wanted to verify the difference between your willingness to pay for a hypothetical payment question and your willingness to pay for* ***a potentially real payment*** *question. […] Obviously, our intention was not to mislead you, but to gain a better understanding of the degree to which a hypothetical scenario could be credible.*

Following this explanation, we offered respondents the opportunity to decide whether or not they wished to continue participating in our survey by answering the following question:

Please confirm whether you do not wish your data to be used for our analysis:

- *Yes, I wish to remove my data from your study.*
- *No, I do not wish to remove my data from your study.*

Such a Partial Information Disclosure arrangement enabled us to create a real feeling of "payment threat" among respondents in the two treatments involving payment consequence, as they presumed that they need to make a real donation at the end of the questionnaire until they reached the debriefing session at the end of the survey. This two-dimension, four-treatment design is similar to that used by Mitani and Flores (2014) in their lab experiment, the difference being that they created a complete consequentiality in both dimensions, so P=1, while in our field

survey, we can only expect to create a partial consequentiality, i.e., 0<P<1. More information on our reasons for using the "Partial Information Disclosure" clause is in the Annex of this paper.

(Please insert Table 1 about here)

We can use the matrix in Table 1 to identify the four treatment groups in the survey. Between the four treatments, the only difference is the information given at the beginning of the CR WTP question. We consider such an arrangement sufficient, since our intention is to test how the hypothetical bias varies with the presence of the two dimensions of consequentiality, separately or jointly. This is different from previous studies which aimed to measure how the estimated WTP value via stated preference studies are compared to the "true" value (the criterion) of the good via criterion validity (Christie 2007) or external validity (Carlsson and Martinsson, 2001) tests. Certainly providing more powerful messages about the potential bias of stated preference methods, these criterion validity tests, requiring equivalent hypothetical and real scenarios are very rare in the literature, especially for case of public goods.[1] In addition, MacMillan et al. (1999) underlined the potential difference between the academic stated preference questionnaires that use a neutral-style text and the fund-raising mail shots that incorporate positive visual imagery and text in order to encourage donations. They concluded that an unrealistic charitable context may provide a different incentive for real and hypothetical donations, thus undermining any comparison between the two. To avoid such potential "contamination" in WTP estimation from wording style differences, we used the same wording style of the WTP question among the four exogenous treatments, all presented in a neutral academic style. Brown and Taylor (2000) separated their comparisons into two distinct steps. The first was between a hypothetical treatment, where no payments were expected from the subjects, and a *real* payment treatment, in which each subject was given a payment form and a pre-addressed, stamped envelope; the second was between a hypothetical treatment and the *actual* payments received by the fund-raising organization. The comparison between treatments conducted in our paper is more comparable to the first step comparisons in Brown and Taylor (2000).

[1] Some examples are the studies by Seip and Strand (1990), Navrud (1992), Brown et al. (1996), Foster et al. (1997), Champ et al. (1997), Vossler et al. (2003), Christie (2007), etc.

We used "suggested donation" as the payment vehicle in our survey, which was directly calculated from the implementation cost of each real project. It is a consensus that, due to the potential problem of free-riding related to donations, people tend to under-report their actual payments for public environmental goods (Carson and Groves, 2007). Poe et al. (2002) believed such incentive properties of donations in their turn contribute to an over-estimation of the hypothetical bias in stated preference studies. We support our choice to use donation as a payment vehicle with the following reasons. First, this decision depends on the context of the three real projects of NAQ. As we used the presence or absence of NAQ as the treatment for provision consequence, using donation as a payment vehicle provided the most neutral and realistic context, which is one of the criteria for the choice of payment vehicles emphasized in Mitchell and Carson (1989) and Carson et al. (2001). Second, Champ and Bishop (2001) indicated several benefits of the donation mechanism that may outweigh the drawbacks of potential free-riding. For the hypothetical bias-related validity test, a clear advantage of the donation mechanism is that validation against actual behavior appears to be relatively easy. This explains the existence of a relatively larger number of studies implementing actual and contingent donation comparisons (Duffield and Patterson, 1991; Seip and Strand 1992; Navrud 1993; Brown et al. 1996; Champ et al. 1997; Foster et al., 1997; Byrens et al. 1999; Murphy et al. 2002; Christie, 2007; Alpizar et al. 2008, etc.). Another benefit of the donation mechanism concerns the potential "vehicle bias" associated with a referendum. Numerous studies have documented the potential influence of public distrust on the efficiency of using universal tax as payment vehicles (Jorgenson et al. 1999; Jorgenson and Syme, 2000; Jones et al. 2008; Glenk and Fisher 2010; Habibov et al. 2017, etc.). Their general observation is that under the presence of institutional distrust among respondents, the use of governmental tax may lead to protest and underestimation of WTP, making the comparison between hypothetical and real WTP less predictable. Considering the current low level of public trust in municipal government in Quebec, particularly in Montreal, following the revelation of the numerous corruption cases in the distribution of municipal infrastructure construction contracts prior to 2010 (https://www.ceic.gouv.qc.ca/), we believe that using voluntary contributions to NAQ allows us to avoid such a bias. To mitigate respondents' free-riding incentive, we explicitly explained at the beginning of the WTP question that the suggested donation at the household level was calculated from 50% of the total cost estimation for each project. If NAQ collected 50% of the total cost for a specific project via donations, the Montreal government would match the other 50%, which would allow the project to be launched.

Our survey was conducted in March 2014. A total of 3046 Quebec residents 18 years and over were invited using the internet survey company Survey Sampling Inc. (SSI). Among the invitees, 2588 began the questionnaires, a response rate of 84.9%. A total of 362 respondents did not answer the contingent-ranking question, which reduced our database to 2226 respondents. From this sample, we further excluded 365 persons (12%) who indicated their unwillingness to be included in the database after the alteration to consent requirement was revealed.[2] We excluded another 161 participants (5.3%), as they provided protest answers explaining their choice of the status quo in a follow-up question.[3] We also identified 51 respondents who did not provide the most preferred alternatives in their contingent-ranking question. Consequently, we utilized data gathered from 1649 respondents in our analysis.[4]

Table 2 reports the frequency distributions for the different permutations of rankings, which are from the highest to lowest level. Table 2 indicates that, although the preference between the three opt-in real projects (B=Blainvill, E=Beloeil, V=Verdun) is variable among respondents, most of them present a clear preference for the opt-in alternatives rather than the opt-out one (S=status quo). Table 3 further aggregates the ranking results into four categories: status quo ranked last, Verdun ranked last, Beloeil ranked last and finally Blainville ranked last. For the total sample, 82.66% of respondents ranked the three opt-in alternatives higher, which reveals their general support for these new ecosystem improvement projects.

(Please insert Table 2 and Table 3 about here)

[2] A preliminary statistical check showed that the main socio-economic characteristics of the excluded respondents are not statistically different from the kept sample. The distribution of the number of excluded respondents was similar among the four treatments: treatments 1 (87 over 470, 21%), 2 (107 over 529, 20%), 3 (96 over 508, 18%) and 4 (75 over 503, 15%).

[3] We considered answers to be of the nature of protest when respondents explained their reason(s) for not choosing a project: “I should not have to pay for a problem created by others”, “I do not trust a non-governmental organization to appropriately manage the funds”, and “The projects do not make sense to me”.

[4] Among the 365 respondents who requested their data to be excluded from analysis, 4 did not give their most preferred alternative.

Another interesting finding in Table 3 is the ranking details summarized for each of the four treatment subsamples. Comparing the percentage points of the case where status quo is ranked last, we find that under the pure hypothetical treatment 1, a higher percentage (86.16%) of respondents ranked the status quo as a last option, while the same percentage point reduces to 80.33% those facing payment consequence in treatment 2. This may be regarded as preliminary evidence for payment consequence, since the only alternative that did not require a payment is the status quo. We also observe a lower percentage (83.25%) of respondents ranking status quo as the least preferred alternative while facing provision consequence. Finally, when both consequences were presented, 81.25% of respondents ranked the status quo as least preferred. Comparing this to the percentage points of the hypothetical and payment consequence treatments, we believe the co-presence of both consequences does affect people in a particular way.

3. Model and hypotheses

Considering ranked data provide more fully description of a respondent's preferences, we used both simple logit specification to study how an opt-in alternative was preferred to status quo and two generalization of the logit specifications to take advantage of the additional information revealed between most and least preferred alternatives. By using the three estimation models, we explored how an individual choice was influenced by both the proposed alternatives and by his/her personal tastes and checked whether the consequences exerts their influence in a heterogeneous ways through the both dimensions of determinants.

We first analyzed the information about people's preference for an opt-in alternative vs. status quo via a binary logit model. Lareau and Rae (2001) believed that before the ranking process, each respondent first resolved the tradeoff between the provision of an environmental good and its price in each opt-in alternative. The natural counterpart of such trade-off is the status quo. Following this logit, we can write the probability for an opt-in project *j* to be ranked higher order than status quo by a respondent *i* as follows:

$$Prob(respondent\ i\ perfers\ project\ j\ ) = f\left(V_{ji} > V_{ki}\right) = \frac{\exp(V_{ji})}{1+\exp(V_{ji})} \quad (1)$$

j=Verdun, Beloeil or Blainville and k=status quo. V_{ji} is the deterministic part of the utility of respondent *i* from the project *j,* U_{ji}.

To avoid the bias related to the potential imprecision in ranking data, especially for the alternatives that ranked in lower order, we also analyzed the most preferred alternatives via a multinomial logit model. The probability for an opt-in project *j* to be ranked highest by a respondent *i* among the four alternatives can be presented by equation (2), with j= Verdun, Beloeil, Blainville or Status quo, and k≠j.

$$Prob(respondent\ i\ perfers\ project\ j\ ) = f\left(V_{ji} > V_{ki}, j \neq k\right) = \frac{\exp(V_{ji})}{\sum_{j=1}^{J} \exp(V_{ji})} \quad (2)$$

Both the logit and multinomial logit models were estimated on project-level. The utility determination function of the respondent i for each of the project j can therefore be written as equation (3).

$$V_{ji} = \beta_{0,j} + \beta_j payment_i + \delta_j provision_i + z_i{}'\gamma_j \quad (3)$$

Here, *Payment* and *Provision* are two dummies that capture the influence of the two dimensions of consequentiality in the determination of utility, and z_i is vector of the socioeconomic characteristics of the respondent i.

The binary logit model and the multinomial logit allowed only the usage of sociodemographic characteristics of respondents to explain his/her choice. To explore the potential determinants from the project attributes and associated consequentiality impacts, we also employed the ranked-order logit model (Beggs et al., 1981). This model expresses the probability of obtaining a particular ranking order by respondent i as:

$$Prob(U_{r1\,i} > U_{r2i} > \cdots > U_{rHi}) = \prod_{h=1}^{H} \left[\frac{\exp(V_{hi})}{\sum_{j=h}^{H} \exp(V_{ji})}\right] \quad (4)$$

For our paper, j, h=Verdun, Beloeil, Blainville and Status quo and H=4. The rank-ordered logit model relies on the repeated application of the multinomial logit specification to the set of alternatives remaining after successive first choices have been eliminated from the available options (Foster and Murato, 2002).

For ranked-order logit model, we could assume the deterministic part of his/her utility to be dependent on the attributes describing alternatives and respondent fixed effect π_i as illustrated in equation (5).

$$V_{ji} = \pi_i + a_j'\rho + a_j'payment_i\beta + a_j'provision_i\gamma \quad (5)$$

As the payment and provision consequences were invariable for each respondent, to capture their impacts on one's decision, we needed to use the cross-term between the attributes and consequence dummies. Such arrangement also allowed us to check the potential heterogeneous impact of consequences between attributes.

In relating the estimation functions to our exogenous treatments, the first treatment group is our baseline, a hypothetical stated preference framework, so we had *reel=0* and *provision=0*. Comparison of answers from the first and second treatment groups, where *reel=1* and *provision=0*, indicated the impact of payment consequentiality, which was captured by the coefficient β_j (eq.3) or β (eq. 5).[5] We expected the probability for a same opt-in alternative to be ranked lower when the payment consequence was introduced, which signified $\beta_j < 0$ (eq. 3) and β <0 (eq. 5). Comparison of answers between the first and third treatment groups, in which *reel=0* and *provision=1*, revealed the impact of provision consequentiality, captured by the coefficient δ_j (eq. 3) or δ (eq. 5). We expected the probability for a same opt-in alternative to be ranked higher when the name NAQ was mentioned, since respondents may view the realization of the projects to be either more possible or of a better quality, so $\delta_j > 0$ or δ >0. This is similar to the findings in previous knife-edge evidence studies, such as Herriges et al. (2010), Vossler and Watson (2013), Interis and Petrolia (2014) and Groothuis et al. (2017).[6] We therefore formed the following two hypotheses:

[5] Such a strategy to give dichotomous 1/0 values to the treatments may overlook the potential heterogeneous impact of consequentiality among respondents, since respondents with different sociodemographic profiles may be affected by the presence of consequentiality to different extents. We did not take this direction in our analysis for this paper.

[6] Such a positive coefficient may seem to contradict Carson et al. (2014), Vossler and Evans (2009) or Vossler et al. (2012). This is because the comparisons conducted in these studies focused on the difference between the

Hypothesis I: $\text{Prob}_{\text{treatment1}} > \text{Prob}_{\text{treatment2}}$: $\beta_j < 0$ (eq. 3) or $\beta < 0$ (eq. 5).

Hypothesis II: $\text{Prob}_{\text{treatment1}} < \text{Prob}_{\text{treatment3}}$: $\delta_j > 0$ (eq. 3) or $\delta > 0$ (eq. 5).

Since the fourth treatment group includes both payment and provision consequentiality arrangements, the above two hypotheses, once held, should also allow us to test the hypothesis of Mitani and Flores (2014) that the final comparison between marginal WTP obtained from the first and fourth treatments depends on the counterbalancing of forces between the negative impact of payment consequence and the positive impact of provision consequence.

Simply adding up the impact of the two consequences, however, does not capture the underlying condition of "knife-edge" evidence, which requires the co-presence of both consequences as a necessary condition. To test whether such a co-presence condition is necessary for consequentiality, we further included a cross-term of the two consequence dummies in the utility determination function. We therefore obtained the equation (6) for binary and multinomial logit models and equation (7) for ranked-ordered logit model.

$$V_{ji} = \alpha_j + z_i'\gamma_j + \beta_j payment + \delta_j provision + \theta_j payment * provision + \varepsilon_{ji} \quad (6)$$

$$V_{ji} = \pi_i + a_j'\theta + a_j' payment_i\beta + a_j' provision_i\gamma + a_j'(payment_i \times provision_i)\theta \quad (7)$$

If the provision or payment consequence can effectively influence respondents' preference between projects only when it was co-presented with the other consequence, we could obtain the following Hypothesis III and Hypothesis IV; each discusses the conditions for the validation

conventional hypothetical scenario, where both provision and payment consequence were equal to 0 ($P_{payment}$=0 and $P_{provision}$=0), and the scenarios with various levels of provision probability ($P_{provision}$>0) co-presented with full payment consequence ($P_{payment}$=1). A more appropriate comparison of our study with Carson et al. (2014) will be with their "binding majority with free chance" scenarios (Table 2, p. 188). In these scenarios, as the provision of the good requires a different probability of payment, we can consider such variations as changes in provision quality. In that sense, we see that a better provision probability (quality) also leads to a higher percentage of "Yes" answers. A closer comparison of the logic of our scenario establishment would be Mitani and Flores (2014), where the payment and provision consequences are presented independently to the respondents.

of "single" knife-edge evidence for one of the two consequences, and both proceeded by following a logic similar to "difference-in-difference".

Hypothesis III: $Prob_{treatment1} = Prob_{treatment2}$ and $Prob_{treatment3} > Prob_{treatment4}$: $\beta_j = 0$ and $\theta_j < 0$ $(eq. 6)$, or $\beta = 0$ and $\theta < 0$ (eq.7).
Hypothesis IV: $Prob_{treatment1} = Prob_{treatment3}$ and $Prob_{treatment2} < Prob_{treatment4}$: $\delta_j = 0$ and $\theta_j > 0$ $(eq. 6)$, or $\delta = 0$ and $\theta > 0$ (eq. 7).

For the "knife-edge" evidence of the provision consequence, we compared the difference in the probabilities of a same opt-in alternative being ranked higher between Treatment 1 (provision absent, payment absent) and Treatment 3 (provision present, payment absent) with the difference between Treatment 2 (payment present, provision absent) and Treatment 4 (payment present, provision present). If the existence of the payment consequence is a *necessary condition* for "knife-edge" evidence of the provision consequence, we should expect no difference between Treatments 1 and 3; at the same time, the probability for the same project being preferred in Treatment 4 would be larger than that in Treatment 2. For the "knife-edge" evidence of the payment consequence, we should expect no difference between Treatments 1 and 2, but the probability for the same project to be ranked higher would be lower in Treatment 4 than in Treatment 3.

The simultaneous validation of Hypothesis III and Hypothesis IV formed the conditions to accept the idea of "double" knife-edge evidence. We summed up these conditions in Hypothesis V.

Hypothesis V: $Prob_{treatment1} = Prob_{treatment3}$, $Prob_{treatment1} = Prob_{treatment2}$, $Prob_{treatment2} \neq Prob_{treatment4}$ and $Prob_{treatment3} \neq Prob_{treatment4}$; $\beta_j = 0$, $\delta_j = 0$ and $\theta_j \neq 0$ or $\beta = 0$, $\delta = 0$ and $\theta \neq 0$.

As we only had a single cross-term to capture the necessity of the co-presence of one consequentiality for the other, under the validity of "double" knife-edge evidence, the sign of the θ_j (θ) would be under the influence of both provision and payment consequences. As we expected their effects to be in opposite directions, we could not predict the final total impact of such force-contrast.

4. Data analysis

Table 4 reported results based on the binary logit models which estimated the probability for each of the three opt-in alternatives to be ranked higher than status quo. Table 5 reported the results based on multinomial logit, which estimated the probability for an opt-in alternative to be ranked as the most preferred. Using the multinomial logit model requires the data to satisfy the assumption of independence from irrelevant alternatives (IIA). A standard Hausman test was therefore conducted and the coefficient estimates obtained before and after one of the alternatives (Beloeil) in the choice set was removed were compared. The values of the test statistics were reported at the bottom of Table 5. Since the differences in the coefficients were statistically insignificant according to the χ^2 distribution with 7 degrees of freedom, we could not reject the null hypothesis of IIA for both Model (1) without cross-terms between Payment and Provision and Model (2) with the cross-terms included.

In general, the results in Tables 4 and 5 revealed significant determination roles for several socio-demographic characteristics. The probability to prefer one of the three opt-in projects were found to be significantly higher for younger female respondents who made donation for the environmental causes in the past, have better understanding of current problems associated with aquatic environments and live currently in Grand Montreal regions.

In both Table 4 and 5, the first column of results for each project were based on Model (1), in which only separate dummies of the two consequences are included, they were related to Hypotheses I and II. The second columns of results for each project were based on Model (2), which further included the cross-term of the two consequence dummies, corresponding to the estimates that test Hypotheses III, IV and V. The stable negative and significant coefficients obtained in both tables for the dummy *Payment* provided clear confirmation for the Hypothesis I, which assumes that a same opt-in alternative were ranked lower when respondents were in face of the payment treat, everything else being equal. This results echoed well to most of existing literature studying the payment consequence (Cummings and Taylor, 1998; Landry and List, 2007; Mitani and Flores, 2009, etc), which in general believed the presence of payment treatment tends to reduce the hypothetical bias by adjusting people's stated preference downwards, with or without the presence of the provision consequence. The findings for the provision consequence were not conclusive. On one hand, the general insignificant coefficients of its single-term in Model (1) seemed to reveal the missing impact of the mention of NAQ on people's choice, which refused Hypothesis II. On the other hand, we did not obtain neither significant coefficients for its cross-term with Payment dummy (c.f. the statistical significance of the coefficients and that of the LR test comparing the models

w/o the cross-term), which means we could not confirm Hypotheses III and IV, although the coefficients for the cross-terms between two treatments were positive, confirming our expectation according to Hypothesis IV. The only exception was in the estimates of Model (2) for the project Blainville in Table (5) based on multinomial logit model, in which we got an insignificant coefficient for simple provision dummy and at the same time a significantly positive coefficient for the cross-term between two consequence dummies, suggesting that the provision consequence might have affected people's choice when it was co-presented with the payment consequence.

One possible explanation of such results was the potential heterogeneous impact of consequence between projects. This reminded us the necessity to verify whether the consequentiality affected people's choice according to the level of the attributes.[7] The ranked-order logit models were therefore estimated, in which the impacts of consequentiality were specified to each of the attributes via their cross-terms with consequences. Since people's preference for an opt-in alternative could be affected simultaneously by the environment-related attributes (Biodiversity, Water quality and Carbon sequestration) and the monetary attribute (cost), but in the opposite directions, the LR test revealed that there was no real gain of explanative power if we added the cross-terms with consequences to both the environmental and monetary attributes. We therefore gradually reduced the cross-terms of the attribute cost with consequence, which gave us the constraint model (2), where only environmental attributes were supposed to be influenced by consequences.

Generally speaking, the stable and significant positive coefficients reported for the three environment-related attributes (*Biodiversity, Carbon and Water quality*) in Table 6 confirmed that a higher level of these attributes can increase the probability for a project to be ranked high. For the attribute *cost*, however, its negative coefficients stayed only significant if we included its cross-term with both consequence dummies. Possible explanations for such results were the low variability of the costs among projects and their high correlation with improvements in the ecosystem services as promised by the projects.[8] The cross-terms of these environmental attributes

[7] Another aspect of the estimated results was that, due to the real nature of the projects studied in our survey, it was not possible to include all attributes in the same estimate since some attributes had perfect multi-collinearity among them. Bateman et al. (2006) faced the same correlation issue. Estimates with the other three attributes and cost, which are available upon request, gave very similar results to those in Table 6.

[8] The suggested contribution was actually directly calculated from the three projects' real cost. A range of higher and more variable costs would have probably helped us to obtain a more significant coefficient, while making the credibility of payment less obvious to respondents.

with payment consequence confirmed the general negative impacts of payment consequence on people's preference for a project, with or without the co-presence of the provision consequence (Hypothesis I). For the impact of the provision consequence, staying insignificant for Biodiversity and Water quality, our results did report that at least for the carbon sequestration, its co-presence with payment consequence ensure a positive and significant effect on people's preference, confirming therefore the Hypothesis IV. The WTP values reported in Table 7 (based on the constraint model 2.2) also reported a similar pattern: we only observed a relatively big increase in WTP for carbon sequestration (from 2.47$ to 2.96$) between WTP under payment consequence and that under both consequences, the corresponding variations in WTP for both biodiversity and water quality were found to be much smaller (biodiversity: 9.86$ to 9.90$) or in opposite direction (water quality: from 96.55$ to 86.77$). We could probably explain such heterogeneity in provision consequence by the relative large differences presented in carbon sequestration attributes between alternatives (c.f. Table A1) with respect to that for the other attributes.[9]

5. Discussion

The test of the "single" vs. "double" knife-edge evidence in this paper focused on the necessity of the co-presence of both payment and provision consequences. This was the first discussion of knife-edge evidence from such an angle. We would like to underline the difference of such a focus compared to previous "knife-edge" evidence studies, such as those by Carson et al. (2014) or Interis and Petrolia (2014). The purpose of those two papers was to verify whether people perceiving partial provision consequence ($0<P<1$) would provide the same WTP as people perceiving a full provision consequence ($P=1$). Their results can be regarded as a foundation for the test conducted in our paper, since the scenarios constructed in our survey did not allow us to realize a complete treatment of consequentiality; therefore, the validity of knife-edge evidence for the equivalence between complete and partial consequence is a necessary basis for our paper.

As mentioned above, our paper was based on contingent-ranking questions, with donation as the payment vehicle, the comparisons of people's answers between treatments did not allow us to discern the real difference between hypothetical and *real* WTP. This is certainly a limit of our paper, but it is also recognizes as a general difficulty faced by the literature in hypothetical bias of stated preference studies. MacMillan et al. (1999) pointed out the potential difference between the academic stated preference questionnaires that used a neutral-style text and the fund-raising mail shots that incorporated positive visual imagery and text in order to encourage donations. This may skew

[9] A similar pattern could also be observed in Table A2, where we did the estimation with the other four attributes.

direct comparison between the numbers from the true difference. For studies that followed cash validity strategies with lab experiments based on "homegrown value," selling a real deliverable market good for cash may result in an underestimation of its *real* WTP, thus enlarging the difference between the hypothetical and the real WTP. This is because the WTP for the deliverable good risks to be censored by the respondent's knowledge of the retail price for a substitute, which is often a part of the total utility, equal to the sum of consumer surplus plus the retail price. The same discussion also applies to studies on the external validity test, such as those by Vossler (2003) and Johnston (2006), which focused on the comparison between the pre-referendum survey organized by researchers and the subsequent real public policy referendum. At least two reasons led us to believe that a real policy referendum may not reflect people's real WTP. First, the public good nature of the policy discussed in the real referendum may have led some residents to under-report their preference in hopes of obtaining a free ride once the policy was finally implemented. Second, not all public referenda are decisive. A policy referendum may be simply regarded as advisory tool for policy makers, reducing voters' incentive to reveal their true preference.

The payment consequence treatment used in our study can be considered incomplete. Telling people that they would be asked to donate to their most preferred project at the end of the survey could not be considered as a full payment threat (P=1), since respondents still had the opportunity to modify the amount they wished to donate at the end of the survey; therefore, people might still don't feel the need to provide real WTP numbers. Thus, at most, we expected our treatment to form partial payment consequence, that means P>0. A similar consideration also applies to provision consequence treatment. In our survey, we mentioned that NAQ was the non-profit organization that would be in charge of project management. As the objective of the paper was to identify "single" vs. "double" knife-edge evidence, more precisely the necessity of having the co-presence of the two dimensions of consequentiality in a survey to reduce the potential hypothetical bias, we believe the identification of zero consequence (P=0) from positive ones (P>0) to be more central. In this sense, our assumption can be simplified, in that respondents in treatments 2, 3 and 4 at least perceived some significant difference in terms of consequence compared with the pure hypothetical treatment, but not a *full* consequence. If the confirmation of Hypothesis II could be used to argue that the respondents of our survey did perceive some payment consequence, the same argument for provision consequence perception were more difficult to defend. Since we did not explicitly ask respondents to auto-report their perceived consequence in the survey, as did Vossler and Watson (2013) and Interis and Petrolia

(2014), our study therefore had no precise control on the extent to which our treatments affected respondents' consequence perceptions. This may have affected our inferences as to how consequence indirectly affected the direction and magnitude of hypothetical bias, especially for provision consequence, whose impacts were only found to be partial or heterogenous between attributes in our results.

One indirect way of discerning such an influence of the treatments on respondents' consequence perception would be to use the information collected in our survey via the follow-up question asking respondents' certainty about their ranking answers. Champ et al. (1997) conducted a study to compare contingent donation with actual donation. They concluded that using a follow-up certainty question to recode uncertain Yes (certainty<10) as No answers had some potential for providing a robust lower-bound estimation of Hicksian surplus (real WTP).[10] An underlying finding from their conclusions is that more certain respondents may provide responses that are more consequential. Following their logic, we may expect consequentiality to lead respondents to adjust their certainty for contingent ranking response.

The follow-up certainty question that we asked in our survey was based on a 1-100 scale with a ladder width of 10, where 100 signified totally certain and 0 totally uncertain. Based on the responses to the contingent ranking question, we regrouped respondents into opt-in and opt-out categories. The opt-in category regrouped respondents choosing one of the three projects as their most preferred choice and the opt-out category included respondents who preferred the status quo to the three environmental improvement projects. Considering the opt-in respondents as providing positive answers with certainty varying between 0 and100 and those in the opt-out category as giving negatives answer with certainty varying between -100 and 0, we further ranked their certainty answers in the following order: certainly opt-in → uncertainly opt-in → uncertainly opt-out → certainly opt-out. Such transformation allowed us to obtain a certainty scale between -100 to 100, in which a higher number signified a more certain preference for opt-in projects. The pair-comparisons between treatments were conducted via Student's t-test. The statistics presented in Table 8 revealed that people under payment consequence had

[10] Champ et al. (1997) used a 1-10 scale to measure certainty, with 10 as totally certain and 1 as totally uncertain.

statistically lower certainty to opt-in than those under hypothetical treatment.[11] Everything else being equal, the respondents under payment consequence provided in average significantly more certain preference for the opt-out option (status-quo), with or without the presence of provision consequence. This echoed the conclusion of Champ et al. (1997) and provided another means of explaining the observed significantly negative coefficient associated with the payment consequence in our estimations. Simply presenting the provision consequence (mentioning NAQ) seems not to affect people's certainty, compared to that in the hypothetical treatment. However, once the payment and provision consequences were co-presented, respondents' average certainty level became significantly higher than that in simple payment treatment. We considered this statistically significant difference in certainty levels as an indirect evidence that mentioning NAQ created some provision consequentiality.

6. Conclusions

This paper provided new evidence on hypothetical bias and consequentiality via a field survey with a representative sample of the Quebec population. Based on a contingent-raking WTP question with donation as payment vehicle, we construct four exogenous treatments to separately and jointly identify the impacts of two dimensions of consequentiality, payment consequence and provision consequence, on people's preference for local environmental public goods. At the outset, based on a contingent-ranking survey with voluntary contributions, our findings provided some support to the so-called "knife-edge" evidence for the provision consequence, similar to that found in previous studies based on the incentive-compatible WTP elicitation format. Moreover, our results revealed "single knife-edge" evidence, according to which provision consequence exerts a positive impact on hypothetical bias mitigation only when payment consequence was co-presented. The same condition, however, does not hold for payment consequence, whose negative impact in hypothetical bias mitigation was found to be more independent. Our results also revealed that the influence of provision consequence via "knife-edge" evidence played in a heterogeneous way on people's choice, according to the level of the environmental attributes proposed in different projects.

[11] Based on our symmetrical order strategy, the significantly lower certainty for opt-in options can be interpreted as a significantly higher certainty for the opt-out option, which is the status quo.

Although we found some evidence supporting the necessity of the co-presence of payment consequence for effective provision consequence, such a joint impact of payment and provision consequences as observed in our data might simply be caused by the combination of the payment vehicles (donation) and our provision consequentiality design (mention of the non-profit organization NAQ). Since the co-presence of both aspects could considerably increase the credibility of the scenarios, it may in turn make the projects more preferable for respondents. It is necessary to re-test the "single" vs. "double" knife-edge evidences with more effective payment and provision consequence treatment designs and with surveys using more incentive-compatible payment vehicles and WTP elicitation formats..

In a general sense, research on hypothetical bias still requires expansion in both theoretical and empirical spheres. Although most researchers agree that this bias is the result of the hypothetical nature of stated preference valuation, there is still no widely accepted behavior-based theoretical framework to describe how people respond to questions about their WTP when the question is hypothetical (Murphy et al. 2005). The case illustrated in our paper showed two other related theoretical topics that may be worthy of study: the first is the theoretical foundation for the necessity of the presence of payment consequence for effective provision consequence, and the second is to understand the relationship between hypothetical bias and respondent uncertainty from a behavioral theory perspective. Such theoretical discussions will defectively provide effective guidance for future research along these lines.

References

Alpizar, F., F. Carlsson and O. Johansson-Stenman (2008). Does context matter more for hypothetical than for actual contributions? Evidence from a natural field experiment. Experimental Economics, 11: 299-314.

Bateman, I. J., Cole, M. A., Georgiou, S., & Hadley, D. J. (2006). Comparing contingent valuation and contingent ranking: A case study considering the benefits of urban river water quality improvements. *Journal of Environmental Management*, *79*(3), 221-231.

Beggi, S., S. Cardell, and J. A. Hausman (1981). Assessing the potential demand for electric cars. Journal of Econometrics 17: 1-19.

Beggs, S., S. Cardell and J. Hausman (1981). Assessing the potential demand for electric cars. Journal of Econometrics, 1981: 1-19.

Ben-Akiva, Moshe, T. Morikawa and F. Shiroishi (1991). Analysis of the reliability of preference ranking data. Journal of Business Research. 14: 149-164.

Blumenschein, K., Blomquist, G. C., Johannesson, M., Horn, N., & Freeman, P. (2008). Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment*. *The Economic Journal*, *118*(525), 114-137.

Bohm, P. (1972). Estimating demand for public goods: An experiment. *European Economic Review*, *3*(2), 111-130.

Brown, K. M. and L. O. Taylor (2000). Do as you say and say as you do: evidence on gender differences in actual and stated contribution to public goods. Journal of Economic Behavior & Organisation. 43: 127-139.

Brown, T. C., P. A. Champ, R. C. Bishop and D. W. McCollun (1996). Which response format reveals the trut about donations to a public good? Land Economics, 72(2): 152-166.

Bulte, E., S. Gerking, J. A. List and A. de Zeeuw (2005). The effect of varying the causes of environmental problems on stated WTP values: evidence from a field study. Journal of Environmental Economics and Management, 49(2005): 330-342.

Caplan, A. J., Grijalva, T. C., & Jakus, P. M. (2002). Waste not or want not? A contingent ranking analysis of curbside waste disposal options. *Ecological Economics*, *43*(2), 185-197.

Carlsson, F. and P. Martinsson (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiment? Journal of Environmental Economics and Management, 41: 179-192.

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and resource economics*, *37*(1), 181-210.

Carson, R. T., Groves, T. and List, J. A. (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, R. T., N. E. Flores and N. F. Meade (2001). Contingent valuation: Controversies and evidence. *Environmental and Resource Economics*. 37: 173-210.

Carson, R. T., T. Grove and J. 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, R., Groves, T., List, J., & Machina, M. (2004). Probabilistic influence and supplemental benefits: A field test of the two key assumptions underlying stated preferences. *Unpublished manuscript*.

Champ, P. A., Moore, R., & Bishop, R. C. (2009). A comparison of approaches to mitigate hypothetical bias. *Agricultural & Resource Economics Review*, *38*(2), 166.

Champ, P., R. Bishop, T.C. Brown and D. W. McCollum (1997). Using donation mechanisms to value nonuse benefits from public goods. Journal of Environmental Economics and Management, 33: 151-162.

Champ, P.A. and R. C. Bishop (2001). Donation payment mechanisms and contingent valuation: An empirical study of hypothetical bias. Environmental and Resource Econonics, 19: 383-402.

Christie, M. (2007). An examination of the disparity between hypothetical and actual willingness to pay using the contingent valuation method: the case of Red Kite Conservation in the United Kingdom. Canadian Journal of Agricultural Economics, 55: 159-169.

Cummings, R. G., & Taylor, L. O. (1998). Does realism matter in contingent valuation surveys? *Land Economics*, 203-215.

Cummings, R. G., Elliott, S., Harrison, G. W., & Murphy, J. (1997). Are hypothetical referenda incentive compatible? *Journal of Political Economy*, *105*(3), 609-621.

Cummings, R. G., Harrison, G. W., & Rutström, E. E. (1995). Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible? *The American Economic Review*, 260-266.

Drichoutis, A. C., A. Vassilopoulos, J. L. Lusk and R. M. Jr. Nayga (2015). Reference dependence, consequentiality and social desirability in value elicitation: A study of fair labor labeling. Paper presented at EAAE-AAEA Joint Seminar "Consumer Behavior in a Changing world: Food, Culture, and Society". March 25-27, 2015, Naples, Italy.

Foster, V. and S. Mourato (2000). Valuing the multiple impacts of pesticide use in the UK: A contingent ranking approach. Journal of Agricultural Economics, 51(1): 1-21.

Foster, V., & Mourato, S. (2002). Testing for consistency in contingent ranking experiments. *Journal of environmental Economics and Management*, *44*(2), 309-328.

Foster, V., I. J. Batement and D. Harley (1997). Real and hypothetical willingness to pay for environmental preservation: A non-experimental comparison. Journal of Agricultural Economics, 48(2): 123-138.

Garrod, G., & Willis, K. (1998). Using contingent ranking to estimate the loss of amenity value for inland waterways from public utility service structures. *Environmental and resource economics*, *12*(2), 241-247.

Glenk, K. and A. Fischer (2010). Insurance, prevention or just wait and see? Public preference for water management strategies in the context f climate change. *Ecological Economics*, 69: 2279-2291.

Groothuis, A. P., T. M. Mohr, J. C. Whitehead and K. Cockerill (2017). Endogenous consequentiality in stated preference referendum data: the influence of the randomly assigned tax amount. *Land Economics*, 93(2): 258-268.

Habibov, N., A. Cheung and A. Auchynnikava (2017). Does social trust increase willingness to pay taxes to improve public healthcare? Cross-sectional cross-country instrumental variable analysis. Social Science & Medicine, 189: 25-34.

Harrison, G. W. (2006). Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics*, *34*(1), 125-162.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature*, 1009-1055.

Harrison, G. W., & Rutström, E. E. (2008). Experimental evidence on the existence of hypothetical bias in value elicitation methods. *Handbook of experimental economics results*, *1*, 752-767.

Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological*, *44*(6), 735-752.

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

Interis, M. G. and D. R. Petrolia (2014). The effects of consequentiality in Binary- and Multinomial-Choice surveys. Journal of Agricultural and Resource Economics, 39(2): 201-216.

Jones, N., C. M. Sophoulis, C. Malesios (2008). Economic valuation of coastal water quality and protest responses : A case study in Mitilini, Greece. *The Journal of Socio-Economics*, 37: 2478-2491.

Jorgenson, B. S. and G. J. Syme (2000). Protest responses and willingness to pay: attitude toward paying for stormwater pollution abatement. *Ecological Economics*, 33: 251-265.

Jorgenson, B. S., G. J. Syme, B. J. Bishop and B. E. Nancarrow (1999). Protest responses in contingent valuation. *Environmental and Resource Economics*, 14: 131-150.

Kotchen, M. J. (2014). Reconsidering Donations for Nonmarket Valuation. *Environmental and Resource Economics*, 1-10.

Landry, C. E., & List, J. A. (2007). Using ex ante approaches to obtain credible signals for value in contingent markets: evidence from the field. *American Journal of Agricultural Economics*, *89*(2), 420-429.

Lareau, T. J. and D. A Rae (1989). Valuing WTP for diesel odor reductions: An application of contingent ranking technique. *Southern Economic Journal*, 55(3): 728-742.

List, J. A., & Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics*, *20*(3), 241-254.

Little, J., & Berrens, R. (2004). Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. *Economics Bulletin*, *3*(6), 1-13.

Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys*, *25*(2), 363-370.

MacMillan, D. C., T. S. Smart and A. P. Thorburn (1999). A field experiment involving cash and hypothetical charitable donations. Environmental and Resource Economics, 14: 399-412.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Millennium Ecosystem Assessment (MEA). (2005). *Ecosystems and Human Well-being: Synthesis*. Island Press, Washington.

Mitani, Y., & Flores, N. E. (2009). Demand revelation, hypothetical bias, and threshold public goods provision. *Environmental and Resource Economics*, *44*(2), 231-243.

Mitani, Y., & Flores, N. E. (2014). Hypothetical bias reconsidered: Payment and provision uncertainties in a threshold provision mechanism. *Environmental and Resource Economics*, *59*(3), 433-454.

Mitchell, R. C. and R. T. Carson (1989). *Using surveys to value public goods: The contingent valuation method*. Resource for the Future. Washington, DC. 1989.

Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, *30*(3), 313-325.

Murphy, J. J., T. Stenvens, D. I. Weatherhead (2002). An empirical study of hypothetical bias in voluntary contribution contingent valuation: Does cheap talk matter? Paper prepared for the World Congress of Environmental and Resource Economists. Monterey, CA. June 24-27, 2002. Paper number: 789000.

Narud, S. (1992). Willingness to pay for preservation of species: an experiment with actual payments. In Pricing the European Environment. New York: Oxford University Press.

Neill, H. R., Cummings, R. G., Ganderton, P. T., Harrison, G. W., & McGuckin, T. (1994). Hypothetical surveys and real economic commitments. *Land economics*, 145-154.

Oehlmann, M. and J. Meyerhoff (2017). Stated presences towards renewable energy alternatives in Germany: do the consequentiality of the survey and trust in intuitions matter? Journal of Environmental Economics and Policy, 6(1): 1-16.

Poder, T.G., & He, J. (2016). Willingness to pay and the sensitivity of willingness to pay for interdisciplinary musculoskeletal clinics: a contingent valuation study in Quebec, Canada. International Journal of Health Economics and Management, online first, doi: 10.1007/s10754-016-9193-5.

Poder, T.G., Dupras, J., Fetue Ndefo, F. & He, J. (2016). The Economic Value of the Greater Montreal Blue Network (Quebec, Canada): A Contingent Choice Study Using Real Projects to Estimate Non-Market Aquatic Ecosystem Services Benefits. Plos One, 11(8):e0158901. doi: 10.1371/journal.pone.0158901. eCollection 2016.

Poder, T.G., He, J., Simard, C., & Pasquier, J.C. (2014). Willingness to pay for ovulation induction treatment in case of WHO II anovulation: A study using the contingent valuation method. Patient Preference and Adherence. 8:1337-46.

Poe, G. L., & Vossler, C. A. (2011). Consequentiality and contingent values: an emerging paradigm. *The International Handbook on Non-Market Environmental Valuation. Cheltenham, UK: Edward Elgar Publishing*, 122-141.

Poe, G. L., J. E. Clark, D. Rondeau and W. Schulze (2002). Provision Point Mechanisms and Field Validity Tet of Contingent Valuation. Environmental and Resource Economics, 23: 105-131.

Poe, G.L. and C. A. Vossler (2009). Consequentiality and contingent valuation: an emerging paradigm. MPRA Munich Personal RePEc Archive. http://mpra.ub.uni-muenchen.de/38864/

Sander LD, Walsh RG, Loomis JB (1990) Toward empirical estimation of the total value of protecting rivers. Water Resource Research 26(7):1345–1357

Seip, K. and J. Strand (1992). Willingness to pay for environmental goods in Norway: A contingent valuation study with real payment. Environmental and Resource Economics, 2: 91-106.

Taylor, L. O., Morrison, M. D., & Boyle, K. J. (2010). Exchange rules and the incentive compatibility of choice experiments. *Environmental and Resource Economics*, *47*(2), 197-220.

The Economics of Ecosystems and Biodiversity (TEEB). (2010). *The Economics of Ecosystems and Biodiversity: Ecological and Economic Foundations.* P. Kumar (Ed.), Earthscan, London, Washington.

Vossler, C. A., & Evans, M. F. (2009). Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. *Journal of Environmental Economics and Management*, *58*(3), 338-345.

Vossler, C. A., & Watson, S. B. (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior & Organization*, *86*, 137-147.

Vossler, C. A., Doyon, M., & Rondeau, D. (2012). Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, *4*(4), 145-171.

Vossler, C. A., J. Kerkvliet, S. Polassky and O. Gainutdinova (2003). Externally validating contingent valuation: an open-space survey and referendum in Corvallis, Oregon. Journal of Economic Behavior & Organisation. 51: 261-277.

Wattage P, Mardle S (2008) Total economic value of wetland conservation in Sri Lanka identifying use and non-use values. Wetland Ecologic Management 16(5):359–369

Zawojska, E., M. Czajkowski, W. Budzinski, A. Wisniewska, C. A. Vossler (2015). The role of objective defined and subjectively perceived consequentiality in stated preference research. Paper presented at WONV 2015, 22-23 June 2015. Nancy, France.

Table 1. The Four Treatment Groups

		Payment Consequence	
		Hypothetical Payment	**Payment Threat**
Provision Consequence	**Hypothetical Project**	Purely hypothetical (Treatment 1, N=383)	Payment consequence only (Treatment 2, N=422)
	Project to be implemented by NAQ	Provision consequence only (Treatment 3, N=412)	Provision and payment consequences (Treatment 4, N=432)

Table 2. Frequency distribution for the rankings of the three projects and the status quo

Ranking	Nb of respondents	%
BEVS	315	19,10
BVES	171	10,37
VBES	235	14,25
EBVS	239	14,49
VEBS	242	14,68
EVBS	161	9,76
BESV	17	1,03
BSEV	6	0,36
SBEV	26	1,58
EBSV	12	0,73
SEBV	19	1,15
ESBV	6	0,36
BSVE	6	0,36
BVSE	11	0,67
SBVE	27	1,64
VBSE	17	1,03
SVBE	26	1,58
VSBE	6	0,36
SVEB	27	1,64
VSEB	12	0,73
SEVB	40	2,43
VESB	10	0,61
ESVB	10	0,61
EVSB	8	0,49
Total	1649	100,00

S: status quo, B: Blainville, E: Beloeil, V: Verdun.

Table 3. Frequency distribution for the rankings in different subsamples

	Total sample		Pure Hypothetical		Payment consequence		Provision consequence		Payment and provision consequence	
	N	%	N	%	N	%	N	%	N	%
Status quo ranked last	1363	82.66	330	86.16	339	80.33	343	83.25	351	81.25
Verdun ranked last	86	5.22	15	3.92	19	4.50	24	5.83	28	6.48
Beloeil ranked last	93	5.64	17	4.44	26	6.16	23	5.58	27	6.25
Blainville ranked last	107	6.49	21	5.48	38	9.00	22	5.34	26	6.02
Total	1649	100.00	383	100.00	422	100.00	412	100.00	432	100.00

Table 4. Probability that an opt-in project will be preferred over status quo (Binomial Logit) (Dependent variable: Preferred to status quo=1, else=0)

	Verdun		Beloeil		Blainville	
	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
Age	-0.0086*	-0.0086*	-0.0079*	-0.0079*	-0.0092*	-0.0092*
	(-1.830)	(-1.828)	(-1.675)	(-1.672)	(-1.940)	(-1.933)
Male	-0.4016***	-0.4013***	-0.2874**	-0.2866**	-0.2177	-0.2175
	(-2.871)	(-2.868)	(-2.057)	(-2.050)	(-1.559)	(-1.556)
Income>50000$/year	0.2490	0.2540	0.1914	0.2005	0.2013	0.2099
	(1.594)	(1.624)	(1.229)	(1.284)	(1.289)	(1.341)
Years of education	0.0388	0.0390	0.0369	0.0374	0.0290	0.0293
	(1.160)	(1.165)	(1.105)	(1.116)	(0.867)	(0.873)
Couple	-0.0709	-0.0727	-0.1086	-0.1112	-0.2286	-0.2319
	(-0.452)	(-0.463)	(-0.690)	(-0.706)	(-1.437)	(-1.456)
Children	0.0298	0.0308	0.0573	0.0594	0.1042	0.1062
	(0.402)	(0.415)	(0.772)	(0.799)	(1.378)	(1.402)
Urban	-0.0713	-0.0737	-0.0512	-0.0567	-0.1880	-0.1923
	(-0.464)	(-0.480)	(-0.334)	(-0.369)	(-1.203)	(-1.230)
Environmental donation	0.9487***	0.9497***	0.8569***	0.8589***	1.0065***	1.0086***
	(5.155)	(5.160)	(4.741)	(4.749)	(5.373)	(5.382)
Aqua situation unknown	-0.5910***	-0.5904***	-0.7864***	-0.7863***	-0.4858**	-0.4852**
	(-3.022)	(-3.018)	(-4.114)	(-4.110)	(-2.425)	(-2.420)
Resident of Grand Montreal	0.4024***	0.4024***	0.2368*	0.2364*	0.3326**	0.3322**
	(2.795)	(2.794)	(1.652)	(1.650)	(2.321)	(2.318)
Provision	-0.0450	-0.1558	0.0061	-0.2033	0.1029	-0.0951
	(-0.326)	(-0.742)	(0.044)	(-0.971)	(0.746)	(-0.443)
Payment	-0.2995**	-0.4002**	-0.2641*	-0.4546**	-0.4232***	-0.5911***
	(-2.151)	(-1.993)	(-1.898)	(-2.268)	(-3.019)	(-2.970)
Provision*Payment		0.1950		0.3719		0.3376
		(0.701)		(1.335)		(1.204)
Constant	1.5534***	1.6071***	1.5478***	1.6496***	1.6875***	1.7813***
	(2.876)	(2.942)	(2.865)	(3.014)	(3.099)	(3.229)
Observations	1649	1649	1649	1649	1649	1649
Log-Likelihood	-688.39	-688.14	-690.36	-689.46	-687.61	-686.88
Pseudo R2	0.0552	0.0555	0.0482	0.0494	0.0541	0.0551
LR test	0.49		1.79		1.49	
	(0.4833)		(0.1811)		(0.2280)	

* p<0.10, ** p<0.05 and ***p<0.01.

[1]. The Loglikelihood ratio test is to test the significance of inclusion of NAQ*real. H0: The model excluding Provision*Payment is not statistically different from the model including the cross-term.

Table 5. Probability that an opt-in project will be preferred the most (Multinomial Logit) (Base=Status Quo)

	Model (1)			Model (2)		
	Verdun	Beloeil	Blainville	Verdun	Beloeil	Blainville
Age	-0.0102*	-0.0119**	-0.0134**	-0.0101*	-0.0118**	-0.0135**
	(-1.755)	(-1.979)	(-2.281)	(-1.731)	(-1.963)	(-2.287)
Male	-0.3950**	-0.2700	-0.0860	-0.3953**	-0.2697	-0.0827
	(-2.293)	(-1.529)	(-0.495)	(-2.294)	(-1.527)	(-0.476)
Income>50000$/year	0.3310*	0.3936**	0.2965	0.3362*	0.4007**	0.3095
	(1.718)	(1.990)	(1.523)	(1.741)	(2.022)	(1.586)
Years of education	0.0271	0.0645	0.0099	0.0274	0.0649	0.0107
	(0.658)	(1.526)	(0.237)	(0.663)	(1.533)	(0.258)
Couple	-0.1687	-0.3107	-0.1337	-0.1707	-0.3128	-0.1349
	(-0.868)	(-1.560)	(-0.681)	(-0.878)	(-1.570)	(-0.686)
Children	0.0815	0.0608	0.0751	0.0820	0.0619	0.0781
	(0.885)	(0.645)	(0.808)	(0.888)	(0.655)	(0.839)
Urban	-0.0791	-0.2175	-0.4175**	-0.0785	-0.2195	-0.4287**
	(-0.407)	(-1.099)	(-2.168)	(-0.404)	(-1.109)	(-2.223)
Environmental donation	1.1040***	1.0029***	1.0244***	1.1039***	1.0032***	1.0279***
	(4.795)	(4.273)	(4.400)	(4.795)	(4.274)	(4.413)
Aqua situation unknown	-0.9721***	-1.0980***	-0.3355	-0.9699***	-1.0964***	-0.3326
	(-3.875)	(-4.096)	(-1.441)	(-3.866)	(-4.089)	(-1.427)
Resident of Grand Montreal	0.4499**	0.3714**	0.4144**	0.4520**	0.3720**	0.4128**
	(2.540)	(2.045)	(2.303)	(2.552)	(2.049)	(2.292)
Provision	0.1146	0.0360	-0.0427	0.0446	-0.0939	-0.3641
	(0.675)	(0.206)	(-0.249)	(0.170)	(-0.351)	(-1.378)
Payment	-0.4000**	-0.4253**	-0.3894**	-0.4416*	-0.5298**	-0.6807***
	(-2.322)	(-2.409)	(-2.238)	(-1.781)	(-2.097)	(-2.746)
Provision*Payment				0.0904	0.2072	0.5855*
				(0.263)	(0.587)	(1.682)
Constant	1.0072	0.6707	1.3997**	1.0282	0.7266	1.5460**
	(1.504)	(0.976)	(2.076)	(1.516)	(1.045)	(2.267)
Observations		1649			1649	
Log-Likelihood		-2141.93			-2139.50	
Pseudo R2		0.0266			0.0277	
IIA (Hausman test)[1]		9.99 (0.6165)			10.11 (0.6848)	
LR test[2]	4.86 (0.1822)					

* p<0.10, ** p<0.05 and ***p<0.01.

[1]. The hausman test used to verify whether our data respect the assumption of IIA (independence of irrelevant alternatives). We removed the alternative Beloeil to perform the reduced choice sets. H0: the data respect the assumption of IIA.

[2]. The Loglikelihood ratio test is to test the significance of inclusion of NAQ*real. H0: The model excluding Provision*Payment is not statistically different from the model including the cross-term.

Table 6. Ranked-order logit model

	Full model: attributes are affected by both consequences		Constraint model (1): Cost is only influenced by payment consequence		Constraint model (2): only non-monetary attributes are influenced by consequences	
	Model (1)	Model (2)	Model (1.1)	Model (1.2)	Model (2.1)	Model (2.2)
Biodiversity	0.0921***	0.1002***	0.0890***	0.0907***	0.0823***	0.0840***
	(10.141)	(9.255)	(11.472)	(10.848)	(13.384)	(12.205)
Biodiversity *Payment	-0.0221**	-0.0364**	-0.0220**	-0.0244**	-0.0092**	-0.0117*
	(-2.184)	(-2.492)	(-2.171)	(-2.168)	(-1.987)	(-1.731)
Biodiversity *Provision	-0.0073	-0.0225	-0.0015	-0.0044	-0.0014	-0.0043
	(-0.728)	(-1.523)	(-0.318)	(-0.626)	(-0.303)	(-0.612)
Biodiversity *Payment*Provision		0.0276		0.0046		0.0045
		(1.361)		(0.496)		(0.490)
Water Quality	0.8777***	0.9698***	0.8437***	0.8647***	0.7691***	0.7905***
	(8.576)	(8.036)	(9.562)	(9.170)	(10.824)	(10.077)
Water*Payment	-0.2354**	-0.4008**	-0.2347**	-0.2684**	-0.0918*	-0.1264
	(-2.053)	(-2.434)	(-2.046)	(-2.089)	(-1.665)	(-1.579)
Water *Provision	-0.1188	-0.2951*	-0.0531	-0.0916	-0.0540	-0.0926
	(-1.038)	(-1.761)	(-0.967)	(-1.121)	(-0.983)	(-1.133)
Water *Payment*Provision		0.3218		0.0646		0.0648
		(1.401)		(0.584)		(0.586)
Carbon	0.0312***	0.0390***	0.0283***	0.0301***	0.0220***	0.0239***
	(4.072)	(4.318)	(4.511)	(4.728)	(4.939)	(5.200)
Carbon *Payment	-0.0146*	-0.0288**	-0.0145*	-0.0176**	-0.0025	-0.0058**
	(-1.701)	(-2.339)	(-1.688)	(-2.017)	(-1.494)	(-2.347)
Carbon *Provision	-0.0048	-0.0196	0.0007	-0.0026	0.0007	-0.0026
	(-0.559)	(-1.576)	(0.421)	(-1.058)	(0.440)	(-1.045)
Carbon *Payment*Provision		0.0276		0.0062*		0.0062*
		(1.611)		(1.818)		(1.818)
Cost	-0.0239*	-0.0347**	-0.0187*	-0.0186*	-0.0073	-0.0073
	(-1.751)	(-2.167)	(-1.686)	(-1.683)	(-0.955)	(-0.958)
Cost *Payment	0.0219	0.0419*	0.0218	0.0217		
	(1.432)	(1.906)	(1.421)	(1.413)		
Cost * Provision	0.0100	0.0309				
	(0.655)	(1.392)				
Cost *Payment*Provision		-0.0391				
		(-1.275)				
Observations	6596	6596	6596	6596	6596	6596
Number of groups	1649	1649	1649	1649	1649	1649
Maximum Likelihood_initial	-4056.21	-4053.62	-4056.42	-4054.61	-4057.38	-4055.57
Maximum Likelihood_final	-3944.72	-3941.60	-3944.93	-3942.64	-3945.94	3943.64
Pseudo R2	0.0275	0.0276	0.0275	0.0276		
LR test[1]	6.24 (0.1821)		4.58 (0.2056)		4.60 (0.2037)	
LR test[2] (Full vs. constraint, Model (1)			0.43 (0.5123)		2.45 (0.2971)	
LR test[2] (Full vs. constraint, Model (2)				2.09 (0.3518)		4.09 (0.2522)

* p<0.10, ** p<0.05 and ***p<0.01.

1. The Loglikelihood ratio test is to test the significance of inclusion of NAQ*real. H0: The model excluding Provision*Payment is not statistically different from the model including the cross-term.

2. The Loglikelihood ratio test is to test the significance of inclusion of cost*provision and cost*Payment*Provision. H0: The model excluding the two terms is not statistically different from the model including the cross-term.

Table 7. WTP (based on constraint model 2.2, KR bootstrap of CI)

	Hypothetical	Payment consequence	Provision consequence	Payment+Provision consequence
Biodiversity				
WTP	11.46	9.86	10.87	9.90
2.5%	-10.81	-9.15	-10.21	-9.18
97.5%	33.72	28.88	31.96	28.98
Water quality				
WTP	107.80	90.55	95.17	86.77
2.5%	-99.89	-81.86	-86.49	-78.08
97.5%	315.48	262.97	276.84	251.62
Carbon sequestration				
WTP	3.26	2.47	2.90	2.96
2.5%	-2.30	-1.49	-1.94	-1.99
97.5%	8.82	6.42	7.74	7.90

Table 8. Variations in certainty of WTP responses according to the treatments

Certainty=[0, 100], 100 means totally certain and 0 means totally uncertain				
	Hypothetical	**Payment**	**Provision**	**Payment+Provision**
Hypothetical	--	--	--	--
Payment	[61.3 52.2] 2.78*** (0.00)	--	--	--
Provision	[61.3 62.4] -0.35 (0.72)	[52.2 62.4] 3.09** (0.00)	--	--
Payment+Provision	[61.3 60.6] 0.24 (0.80)	[52.2 60.6] 2.53** (0.01)	[62.4 60.6] 0.59 (0.55)	--

In the [], the first number is the mean value of the respondents' certainty in the treatment titled in column and the second number is the mean value of the respondent's certainy in the treatment titled in line. The number below the [] is the statistical value of student test, and the number in the paretheses is the probability for us to reject the H0 of the student test that the distributions of the certainty of the two treatments is the same.

Annex. Partial information disclosure clause

We initially planned to put up a real donation page directly linked to NAQ website to record respondents' real donation information. However, this arrangement was rejected by the ethical committee, in view of potential financial conflicts of interests (Article 11.10 of Ethical Conduct for Research Involving Humans of the Social Sciences and Humanities Research Council (SSHRC), Canada). Considering that the central focus of our study was to test the "single" vs. "double" knife-edge evidence associated to the hypothetical nature of the stated preference valuation methods, which should already be sufficient by creating an effective payment threat ($0<P<1$ but not $P=1$), we therefore proposed to follow the "Partial Information Disclosure" clause (Article 3.7A of Ethical Conduct for Research Involving Humans of SSHRC), which allowed us to make alterations to consent requirement after the WTP questions. Application of this clause strictly followed the conditions imposed by the Ethical Conduct for Research Involving Humans of SSHRC and was approved by the ethical committee after several rounds of discussion. More precisely, at the end of the questionnaire, we presented a debriefing with an alteration to consent requirement where we explained the purpose behind our research instead of directing respondents to the real donation webpage of NAQ:

> *Although the presented projects are real, the donation question that we presented earlier was totally hypothetical. Under no circumstances will NAQ contact you for a donation. We only wished to verify the difference between your willingness to pay for a hypothetical payment question and that for* ***a potentially real payment*** *question. […] Obviously, our intention was not to mislead you, but to gain a better understanding of the degree to which a hypothetical scenario could be credible.*

Following the ethical principles, we offered respondents the opportunity to decide whether or not they wanted to continue participating in our survey. Respondents were presented with the following question:

> *Please confirm whether you do not wish your data to be used for our analysis:*
> - o *Yes, I wish to remove my data from your study.*
> - o *No, I do not wish to remove my data from your study.*

Table A1. The Three Projects Used in our Survey

	Gohier's Brook – Blainville (QC)		**Dismantling of a dock – Verdun (QC)**		**des Trente's brook – Beloeil (QC)**	
	This project aims at restoring Gohier's brook and its riparian zones in the city of Blainville, and to educate neighboring owners. The intervention plan includes cleaning the brook, controlling invasive alien species, planting vegetation, and wildlife development.		The dismantling of a dock from the early 20th century is planned in the city of Verdun. This dock threatens both public safety and the natural environment of the St. Lawrence River. The project involves the natural restoration and development of the site by planting shrubs and aquatic and terrestrial grasses. Another objective is to control invasive alien species.		The des Trente's Brook in the city of Beloeil has long been neglected and a restoration must be performed. In addition to conducting various renaturalization activities and performing recurring maintenance, new trees will be planted plantations are planned on the banks of the brook.	
Attributes	**Level before**	**Level after**	**Level before**	**Level after**	**Level before**	**Level after**
Biodiversity	130 animal species	135 animal species	5 animal species	26 animal species	37 animal species	42 animal species
	75 plant species	80 plant species	28 plant species	36 plant species	121 plant species	121 plant species
Water quality	Medium	Very good	Good	Good	Very bad	Medium
Carbon sequestration	395 additional trees = 59 tons sequestered		6 additional trees = 1 tons sequestered		200 additional trees = 30 tons sequestered	
Recreational activities						
Landscape aesthetics	Medium improvement		High improvement		Medium improvement	
Restored area	1 hectare		0.01 hectare		3.5 hectares	
Education	Explanatory panels, work performed with the community		Explanatory panels		Explanatory panels, work performed with the community, permanent activities (community garden)	
Suggested donation	$40		$15		$35	

Table A2. Ranked-order logit model (with the other four attributes)

	Full model: attributes are affected by both consequences		Constraint model (1): Cost is only influenced by payment consequence		Constraint model (2): only non-monetary attributes are influenced by consequences	
	Model (1)	Model (2)	Model (1.1)	Model (1.2)	Model (2.1)	Model (2.2)
Education	1.1090***	1.3631***	1.0159***	1.0753***	0.8125***	0.8729***
	(4.474)	(4.671)	(5.007)	(5.236)	(5.656)	(5.934)
Education *Payment	-0.4882*	-0.9508**	-0.4849*	-0.5893**	-0.0963**	-0.2029***
	(-1.757)	(-2.386)	(-1.746)	(-2.086)	(-1.988)	(-2.868)
Education *Provision	-0.1835	-0.6677*	-0.0045	-0.1151	-0.0043	-0.1150
	(-0.661)	(-1.654)	(-0.094)	(-1.593)	(-0.089)	(-1.592)
Education *Payment*Provision		0.9017		0.2027**		0.2030**
		(1.621)		(2.085)		(2.088)
Activities Quality	-0.6487***	-0.8473***	-0.5756***	-0.6215***	-0.4165***	-0.4631***
	(-3.285)	(-3.638)	(-3.537)	(-3.740)	(-3.537)	(-3.784)
Activities*Payment	0.3529	0.7149**	0.3498	0.4317*	0.0463	0.1297
	(1.598)	(2.254)	(1.584)	(1.890)	(0.829)	(1.610)
Activities *Provision	0.0998	0.4765	-0.0400	0.0454	-0.0413	0.0440
	(0.452)	(1.488)	(-0.717)	(0.560)	(-0.741)	(0.542)
Activities *Payment*Provision		-0.7049		-0.1592		-0.1591
		(-1.594)		(-1.424)		(-1.422)
Landscape	2.4187***	2.9095***	2.2363***	2.3464***	1.8399***	1.9515***
	(4.888)	(4.976)	(5.473)	(5.592)	(6.183)	(6.251)
Landscape *Payment	-0.9657*	-1.8536**	-0.9576*	-1.1475**	-0.2018	-0.3949*
	(-1.747)	(-2.332)	(-1.733)	(-1.996)	(-1.347)	(-1.822)
Landscape *Provision	-0.2666	-1.1939	0.0815	-0.1207	0.0857	-0.1157
	(-0.483)	(-1.488)	(0.546)	(-0.547)	(0.574)	(-0.525)
Landscape *Payment*Provision		1.7251		0.3665		0.3650
		(1.558)		(1.220)		(1.216)
Cost	-0.0239*	-0.0347**	-0.0187*	-0.0186*	-0.0073	-0.0073
	(-1.751)	(-2.167)	(-1.686)	(-1.683)	(-0.955)	(-0.958)
Cost *Payment	0.0219	0.0419*	0.0218	0.0217		
	(1.432)	(1.906)	(1.421)	(1.413)		
Cost * Provision	0.0100	0.0309				
	(0.655)	(1.392)				
Cost *Payment*Provision		-0.0391				
		(-1.275)				
Observations	6596	6596	6596	6596	6596	6596
Number of groups	1649	1649	1649	1649	1649	1649
Maximum Likelihood_initial	-4056.21	-4053.62	-4056.42	-4054.61	-4057.38	-4055.57
Maximum Likelihood_final	-3944.72	-3941.60	-3944.93	-3942.64	-3945.94	3943.64
Pseudo R2	0.0275	0.0276	0.0275	0.0276		
LR test[1]	6.24 (0.1821)		4.58 (0.2056)		4.60 (0.2037)	
LR test[2] (Full vs. constraint, Model (1)			0.43 (0.5123)		2.45 (0.2971)	
LR test[2] (Full vs. constraint, Model (2)				2.09 (0.3518)		4.09 (0.2522)

* p<0.10, ** p<0.05 and ***p<0.01.

1. The Loglikelihood ratio test is to test the significance of inclusion of NAQ*real. H0: The model excluding Provision*Payment is not statistically different from the model including the cross-term.

2. The Loglikelihood ratio test is to test the significance of inclusion of cost*provision and cost*Payment*Provision. H0: The model excluding the two terms is not statistically different from the model including the cross-term.