American Economic Journal: Microeconomics 2012, 4(4): 145–171
http://dx.doi.org/10.1257/mic.4.4.145

Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments†

By Christian A. Vossler, Maurice Doyon, and Daniel Rondeau*

This paper explores methodological issues surrounding the use of discrete choice experiments to elicit values for public goods. We develop an explicit game theoretic model of individual decisions, providing conditions under which surveys with a single binary choice question, or sequence of binary choice questions, are incentive-compatible. We complement the theory with a framed field experiment, with treatments that span the spectrum from incentive-compatible, financially binding decisions to decisions with no direct financial consequences. The results suggest truthful preference revelation is possible, provided that participants view their decisions as having more than a weak chance of influencing policy. (*JEL* C83, C93, H41, Q23)

Survey-based methods for eliciting preferences have been a mainstay in various strands of research, including the valuation of public goods, the study of transportation mode choice, the measurement of the value of a statistical life, and the marketing of consumer products. More recently, surveys have been used to measure economic variables such as income expectations, life expectancy, and the level of happiness. Stated preference surveys are often the only practical approach for gauging preferences for non-market goods and for evaluating public policies prior to their implementation. In particular, these methods are widely recognized as the only approach for measuring passive-use values such as the existence value of endangered species protection. Estimates of non-market values are important components of many government-mandated benefit-cost analyses, and are admitted as evidence in legal proceedings over environmental damages (e.g., following an oil spill). Despite the widespread use of these methods, the role of economic incentives in shaping participant responses remains poorly understood.

* Vossler: Department of Economics and Howard H. Baker Jr. Center for Public Policy, University of Tennessee, 527C Stokely Management Center, Knoxville, TN 37996 (e-mail: cvossler@utk.edu); Doyon: Département d'économie agroalimentaire et des sciences de la consommation, Université Laval, 4310-F Pavillion Paul Comtois, Quebec, QC, G1K 7P4, Canada (e-mail: maurice.doyon@eac.ulaval.ca); Rondeau: Department of Economics, University of Victoria, 342 Business and Economics Building, Victoria, BC, V8W 2Y2, Canada (e-mail: rondeau@uvic.ca). We extend appreciation to Fréderic Roy-Vigneault for his excellent research assistance, and to Laure Saulais for her help with running the experiments. We thank two anonymous referees, Michael Price, and seminar participants at the Fourth World Congress of Environmental and Resource Economists, 2009 Southern Economic Association Meetings, Canadian Resource and Environmental Economics Working Group, and University of Tennessee for their helpful comments and suggestions. This research was funded by CIRANO and Agriculture and Agrifood Canada.

† To comment on this article in the online discussion forum, or to view additional materials, visit the article page at http://dx.doi.org/10.1257/mic.4.4.145.

In interpreting the results of stated preference surveys, polar views are often held among both researchers and policymakers. Some treat survey responses as motivated by economic incentives. They may go as far as viewing responses as *truthful* indications of preferences, regardless of *how* preferences are elicited. Others consider survey responses as answers to hypothetical (i.e., inconsequential) questions. They may dismiss the methodology entirely, advocate that estimated values be scaled (down), or support "cheap talk" and related entreaty methods designed to convince participants to behave as if incentives existed.[1]

Rather than adopt a particular perspective, we develop a game theoretic framework to analyze the incentive properties of discrete choice experiment (DCE) elicitation mechanisms.[2] We complement the theory with a field experiment, and find the empirical evidence speaks both favorably and unfavorably on the ability of DCE surveys to measure the demand for public goods with passive-use value.

In a DCE, participants are presented with a series of "choice sets." Each choice set is made up of two or more comparable goods defined by their respective levels of a common list of attributes. Participants are asked to indicate which good they prefer. DCEs are often thought to be superior to alternative elicitation approaches, such as a single binary choice (SBC) question (i.e., a single yes or no vote on a project), because of (perceived) gains in statistical efficiency; the ability to estimate the value of attributes at the margin (providing a richer depiction of preferences and facilitating inferences to other situations in benefits transfer studies); and the possibility of testing for internal consistency (Adamowicz 1998; Alpízar, Carlsson, and Martinsson 2003; Hanley, Mourato, and Wright 2001; Hanley, Wright, and Adamowicz 1998; Holmes and Adamowicz 2003).

The origins of DCEs lie in Lancaster's (1966) theory of value and McFadden's (1974) random utility theory. Closely related conjoint analysis applications in marketing and transportation date back to at least the 1970s (e.g., Green and Rao 1971; Norman and Louviere 1974; Beggs, Cardell, and Hausman 1981). DCEs gained prominence in the 1980s (e.g., Louviere and Hensher 1982) and researchers began applying the methods to issues in health and environmental economics in the 1990s (e.g., Carson, Hanemann, and Steinberg 1990; Ryan and Hughes 1997). Recent studies have expanded the use of DCEs to compare the risk and time preferences of smokers and nonsmokers (Ida and Goto 2009), estimate preferences for electricity reliability (Blass, Lach, and Manski 2010), and elicit discount rates for water quality policies (Viscusi, Huber, and Bell 2008).

The debate over the use of stated preference methods reveals both positive and negative results on their validity. On the one hand, there is strong evidence of "construct validity" and "convergent validity."[3] For example, stated preference surveys produce value estimates for quasi-public goods that are generally comparable to

[1] We note that the degree of opposition to survey methods appears to differ across disciplines and applications. Most of the criticism regarding the possibly "hypothetical" nature of surveys has been from economists (e.g., Diamond and Hausman 1994) who generally work outside the area of non-market valuation.

[2] DCEs are also commonly referred to as "choice modeling," "conjoint-based choice experiments" or simply "choice experiments."

[3] Construct validity refers to whether the measure of interest (e.g., willingness to pay) varies in ways suggested by economic theory. Convergent validity refers to whether the measure of interest is similar to other measures that it should be theoretically similar to.

those obtained from revealed preference methods, such as hedonic pricing and travel cost (Carson 1996). Similar supportive findings also appear in the transportation and marketing literatures (Alpízar, Carlsson, and Martinsson 2003; Carson, Flores, and Meade 2001). On the other hand, accumulated evidence, largely from laboratory experiments, shows systematic deviations between stated and revealed preferences (List and Gallet 2001; Little and Berrens 2004; Murphy et al. 2005). This is generally referred to as "hypothetical bias." This lack of correspondence has raised questions about the "criterion" (i.e., external) validity of stated preferences.

Two stylized facts appear to emerge from the limited number of criterion validity studies that specifically involve DCEs or other attribute-based methods. First, stated preference methods tend to provide biased estimates of "levels" such as total willingness to pay (WTP) or market shares (Louviere, Hensher, and Swait 2000; Lusk and Shroeder 2004; Taylor, Morrison, and Boyle 2010). Second, stated and revealed preference approaches produce similar estimates of "marginals," such as marginal WTP or marginal utility of attributes (Carlsson and Martinsson 2001; Lusk and Schroeder 2004; List, Singha, and Taylor 2006; Taylor, Morrison, and Boyle 2010). As is the case of most studies in the stated preference literature, these criterion validity studies involve a comparison between what is clearly an inconsequential decision setting and a revealed-preference measure. To the extent that perceived consequences exist in field surveys, evidence from these studies may be limiting.

In this study, we extend the existing theoretical and empirical literature related to stated preference surveys in general, and DCEs in particular. Building upon the insights of Carson and Groves (2007), we develop an explicit game theoretic model of individual choice for settings where participants face decisions on one or more choice sets that are disclosed in advance. The mechanism by which individual decisions translate into the implementation of a public good or policy can be explicit or unknown. The theory provides general conditions under which binary DCEs, including the special case of an SBC question, are incentive compatible. We show that for DCEs, incentive compatibility (IC) holds for a class of mechanisms that maintain specific links between the choice sets and potential public policies, maintain independence between choice sets, and meet specific monotonicity requirements (related to consequentiality).

In the empirical component of the research, we conduct a framed field experiment that makes use of an opportunity to elicit values from local citizens for tree row plantations in agricultural areas of the province of Quebec (Canada). The field experiment represents the first criterion validity study where estimates of WTP obtained from an incentive-compatible DCE and parallel stated preference survey are compared for a public good with passive-use value.

The baseline treatment is an incentive-compatible, financially binding DCE with a provision rule that makes it clear that choice sets are independent. The next treatment involves a theoretically equivalent mechanism, but the provision rule makes it less obvious that the independence condition holds. The third treatment involves a binding elicitation, but with a provision rule that is unknown to participants. The last treatment is a stated preference DCE designed to capture the incentives that may exist in the surveys used in many field contexts. In this treatment, participants' decisions have no immediate financial consequences. They are informed that responses will be shared with policymakers but no information is given on how responses will be used. The experimental treatments allow us to explore the incentive properties of DCEs.

We find that all three treatments with real payments lead to statistically equivalent WTP functions. This provides evidence that an explicit provision rule may not be essential for DCEs to reveal demand. The WTP function estimated from stated preference responses is statistically different from those of the other treatments unless we restrict the analysis to only those participants who believe that their choices have more than a "weak" chance of influencing policy. The results suggests that truthful preference revelation is perhaps more likely than theory alone would suggest. However, convincing participants that their choices are consequential appears to be critical to ensuring valid results.

I. A Theoretical Model of Discrete Choice Experiments

In this section, we develop a game theoretic representation of the incentives faced by participants in DCEs, including the special case of an SBC elicitation. Our objective and approach is to elaborate sets of sufficient conditions that ensure that DCEs are incentive compatible (i.e., participants cannot do better than voting truthfully). We focus our attention primarily on situations where each choice question pits a single project against the status quo (no project); and where, if there is more than one choice question, all of them are disclosed in advance to participants. Although it does not encompass all forms of DCEs used in practice, this setup fits many applications in the environmental and health policy literatures, as well as recent field experiments in marketing (e.g., Lusk and Schroeder 2006). Individual conditions are argued or shown to be necessary for the sufficient conditions to hold as a group. We cannot formally prove that they are universally necessary or rule out that other sets of sufficient conditions may exist.

A. General Setup

Consider a DCE where M participants are presented with $K \geq 1$ "choice sets." Each choice set presents a "project," indexed by k, that is defined by a vector of "attributes" ($\mathbf{a}_k$), and a cost (c_k). Participants are asked to make a choice between the project (a yes vote) and the status quo (a no vote) for each choice set. The votes of individual m are assembled into a vector $\mathbf{V}_m = (V_{m1}, V_{m2}, \ldots, V_{mK})$. In the discussion that follows, we write a subscript on a vector of votes (e.g., $\mathbf{V}_m$) to identify the individual casting them. A subscript on a single vote (e.g., V_k) denotes the project associated with the vote, and it is assumed that the vote is cast by respondent m.

The $M \times K$ votes are received by a policymaker who selects an outcome from the set of N possible "policies," each defined by a vector of attributes ($\mathbf{A}_n$), and a cost (C_n). From the respondent's perspective, the votes potentially influence the probabilities, P_n, that a policy will be implemented.[4] Formally, the probabilities are

[4] As an example, Vossler and Evans (2009) consider a situation where the project in an SBC referendum corresponds exactly with a policy, but where the policymaker considers his own preferences for the proposed project

determined by the "policy function" F:$(\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_M; G) \rightarrow P = (P_1, P_2, \ldots, P_N)$, where G represents the policymaker's preferences, constraints, and other considerations that may enter the decision-making process. In what follows, we adopt the shorthand $F(\mathbf{V}_m, \mathbf{V}_{-m}; G) = \mathbf{P}(\mathbf{V}_m, \mathbf{V}_{-m}; G) = (P_1, P_2, \ldots, P_N)$, where $\mathbf{V}_{-m}$ is the matrix of all votes by respondents other than m. By construction, $0 \leq P_n \leq 1 \forall n$, and $\sum_{n=1}^{N} P_n < 1$. The probability that the status quo is maintained is then given by $P_0 = 1 - \sum_{n=1}^{N} P_n$.

The probability-based policy function is important in two ways. First, it acknowledges that there is uncertainty surrounding how one's responses affect the policy outcome. Second, it makes it explicit that one's votes interact with the choices of other participants. From our perspective, these are two critical aspects of DCEs that have often been overlooked.[5]

For a representative participant, the utility of the status quo is denoted by U_0 and the utility of policy n is given by $U_n = u(Y - C_n; \mathbf{A}_n)$ where Y is the individual's income. It will be convenient to define $\mathbf{U}$ as the column transposition of the vector of utilities $(U_1, U_2, \ldots, U_N)$ to represent the utility levels that an individual would get from the implementation of each possible policy.

We adopt an expected utility framework whereby the agent wishes to maximize

$$(1) \qquad EU_m(\mathbf{V}_m, \mathbf{V}_{-\mathbf{m}}; G) \;=\; \mathbf{P}(\mathbf{V}_m, \mathbf{V}_{-m}; G) \;\cdot\; \mathbf{U} \;+\; P_0 U_0.$$

This model of choice can only be appropriate if participants meet the basic rationality requirements of expected utility theory. In particular, agents must be able to assign a utility level to each possible policy and to the status quo.

B. Additional Concepts and Notational Conventions

Marginal effect of a yes vote: For any project k and policy n, we adopt the notation that $\Delta P_n / \Delta V_k$ measures the change in the probability that policy n is implemented when a vote on project k is changed from a no to a yes.

Independence between choice sets: We say that a policy function F maintains independence between choice sets if and only if $\Delta P_n / \Delta V_k \neq 0$ for at most a single policy n and at most a single project k. This means that a vote on project k has a marginal impact on the probability of implementing at most one policy (and the status quo), and that the probability of implementing policy n can only be affected by the votes on no more than one project. This will turn out to be a key requirement for DCE surveys to be incentive compatible. Yet, this is not a condition that can easily be assumed to hold in an advisory field survey.

One-to-one correspondence: A DCE is said to maintain the "one-to-one" property if and only if $N = K$, $\mathbf{a}_k = \mathbf{A}_n \ \forall (k = n)$, and $c_k = C_n \ \forall (k = n)$. This implies

along with the result of the referendum.

[5] At this stage, we take the function F as given. Uncertainty and participant beliefs about the function are addressed later.

that each project corresponds exactly to one possible policy, and that both are identical in attributes and costs. When a specific project is identical to a policy, we denote it with the subscript s.

Consequentiality: Carson and Groves (2007) correctly argue that the common perception that a "hypothetical" survey can only yield hypothetical answers is misguided. They point out that the relevant question is whether or not the survey is perceived by participants to be "consequential." There are two assumptions underlying consequentiality. The first is that individuals care about the potential benefits or costs of at least some of the N public policies contemplated. In other words, at least some of the policies enter the participant's utility function. The second assumption is that "the agent answering a preference survey question must view their responses as potentially influencing the agency's actions" (p. 183). This can be broadly expressed as requiring that a vote on project k has some expected influence on the choice of policies: $\Delta P_n / \Delta V_k \neq 0$ some of the time for at least one of the N possible policies (not necessarily identical to project k). Consequentiality provides incentives for participants to partake in the DCE, makes standard economic analysis applicable to participant choices, and makes the development of DCEs a legitimate application of mechanism design theory.

Weak positive monotonicity of the policy function: The following condition, $\Delta P_s / \Delta V_s \geq 0$, with $\Delta P_s / \Delta V_s > 0$ at least some of the time, states that a yes vote for project s must result in a weak increase in the probability that the corresponding policy will be implemented. The monotonicity requirement only needs to be weak with the caveat of a strict inequality "at least some of the time" to guarantee that the participant perceives a yes vote as increasing the probability that policy s will be implemented. Several commonly used provision rules meet this criterion, including the standard majority rule.

C. Incentive Compatibility of Single Binary Choice Questions

A careful reading of Carson and Groves (2007) allows one to identify four conditions which together are sufficient to ensure that participants will truthfully vote according to their preference between a single project and the status quo:

(i) the participants care about the outcome;

(ii) the authority can enforce payments by voters;

(iii) the elicitation involves a yes or no vote on a single project; and

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

We have defined most of these conditions in the previous section. Condition (i) and (ii) trivially require that the attributes and costs of policies enter the participant's utility function. Condition (iii) imposes that the votes cast by participants relate to the implementation of a single possible policy. It is implicit in Carson and Groves' (2007) discussion that this single policy meets the one-to-one matching condition such that the policy is identical to the proposed project. Formally, this restricts the policy function to map a participants' single vote on project s to the unique public policy, $(\mathbf{A}_s, C_s) = (\mathbf{a}_s, c_s)$. It follows that each participant's vote is a singleton $\mathbf{V}_m = (V_s)$ and the policy function aggregating responses returns a single value for P_s. Without this restriction, voting for s could result in a change in the probability that some other policy will be implemented. A utility-maximizing voter facing such a problem would take the impact of his vote on the likelihood of those other policies into consideration, potentially resulting in a vote on the project that reflects at least partially the individual's preferences for other policies. This situation almost certainly leads to a loss of IC.

We now provide a succinct proof of Carson and Groves' (2007) argument that an SBC elicitation meeting conditions (i–iv) is incentive compatible. It formalizes their claim and it will be a useful stepping stone for analyzing the DCE mechanisms of primary interest to this paper.

PROPOSITION 1: *An SBC question is incentive compatible if conditions* (*i–iv*) *hold.*

PROOF:

Consider the vote of an individual in the group of M participants. Let the individual's true preference ranking between project s and the status quo be represented by the singleton T_s (i.e., $T_s =$ yes if $U_s \geq U_0$, and $T_s =$ no otherwise).[6] The expected utility of the participant, given T_s and arbitrary votes by all other participants $(\mathbf{V}_{-m})$, is denoted by

$$(2)\qquad EU_m(T_s, \mathbf{V}_{-m}; G) = P_s(T_s, \mathbf{V}_{-m}; G)U_s + \big(1 - P_s(T_s, \mathbf{V}_{-m}; G)\big)U_0.$$

The proof proceeds by establishing that voting truthfully is a dominant strategy. Keeping $\mathbf{V}_{-m}$ constant, consider a vote $V_s \neq T_s$ ($V_s =$ yes if $T_s =$ no, or $V_s =$ no if $T_s =$ yes). With conditions (i–iv) holding, deviating from truthful voting results in a change in expected utility equal to

$$(3)\qquad \begin{aligned} \Delta EU_m &= EU(V_s, \mathbf{V}_{-m}; G) - EU(T_s, \mathbf{V}_{-m}; G) \\ &= \big[P_s(V_s, \mathbf{V}_{-m}; G) - P_s(T_s, \mathbf{V}_{-m}; G)\big](U_s - U_0). \end{aligned}$$

If $U_s - U_0 \geq 0$ ($T_s =$ yes), it follows that $P_s(V_s, \mathbf{V}_{-m}; G) - P_s(T_s, \mathbf{V}_{-m}; G) \leq 0$, and therefore that $\Delta EU \leq 0$ by virtue of the monotonicity condition. Conversely, if the truthful vote is a no (i.e., $U_s \leq U_0$) the first term in the second line of equation (3) is non-negative, while the second is non-positive. Once again, deviating from true

[6] The weak inequality could be reversed (i.e., $T_s =$ yes if $U_s > U_0$, and $T_s =$ no if $U_s \leq U_0$) without affecting the proof.

preferences cannot increase the participant's expected utility. This conclusion holds regardless of the votes of others. We therefore conclude that truthful revelation is a dominant strategy.

Truthful revelation is only a weakly dominant strategy unless the mechanism is strictly monotone $(\Delta P_s/\Delta V_s > 0)$ everywhere and preferences result in a strict ranking between the project and the status quo $(U_s \neq U_o)$. Where the mechanism is strictly monotone and the strict inequality in utilities holds for all players, the demand-revealing strategy is strictly dominant for all and constitutes the unique Nash Equilibrium of the group decision game.

It can be shown that removing any of the four conditions leads to a loss of IC. Consequentiality (restricted to the weakly positively monotonic subset of consequential functions), in particular, is essential to the decision problem faced. It also provides a theoretical prediction and a diagnosis of the mechanism's incentive design. It should be clear that without weak positive monotonicity, deviating from the truth can increase utility and result in a loss of IC.

D. Incentive Compatibility of Discrete Choice Experiments

Extending a DCE to include two or more choice sets can easily eliminate a participant's incentives to vote truthfully. The possibility that more than one policy could be implemented; that a single vote could impact the implementation probability of multiple policies; or that votes on different projects could affect the probability of a single policy can all be shown to potentially create incentives to deviate from truthful voting.

It is conceivable that a DCE could be designed to result in the implementation of more than one policy. In principle, this could be done while maintaining IC. However, this proves possible only if preferences are additively separable across projects. IC of a DCE is lost when various projects are imperfect substitutes or complements, or when they are tied together by the participant's budget constraint.[7] These situations make the marginal value of one project (and therefore each vote) contingent upon the bundle of projects actually implemented. When multiple projects can be implemented, votes would differ from those when only one can, and uncertainty regarding implementation in the multiple policy case will generally lead to a loss of IC. In any event, it would be very difficult if not impossible to recover preferences from such an exercise. Thus, sufficient conditions for IC maintain that only one policy can be implemented.

Ensuring IC of the binary DCE will generally require modifying condition (iii) above, and adding two additional conditions:

(iii′) at most one policy can be implemented;

(v) the policy function F maintains independence between choice sets; and

[7] This is a well-known problem in auction theory, where combinatorial auctions have been devised to attenuate the complementarity and substitution problems associated with simultaneously selling multiple units. Cramton, Shoham, and Steinberg (2006) provide a good introduction to combinatorial auctions.

(vi) there is a one-to-one correspondence between projects and policies.

PROPOSITION 2: *A DCE consisting of a sequence of binary choice questions, each involving a choice between a project and the status quo, is incentive compatible if conditions* (*i*), (*ii*), (*iii′*), (*iv*), (*v*), *and* (*vi*) *hold.*

PROOF:

Consider the votes of an individual and, without loss of generality, order the different projects according to the level of utility they confer to this individual. The participant's complete preference ordering (including the status quo) will therefore take the form: $U_1 \geq U_2 \geq \cdots \geq U_0 \geq \cdots \geq U_K$ and the column vector of utilities, $\mathbf{U}$, will be ordered accordingly (except for the absence of U_0). The truthful demand revealing vector of votes, $\mathbf{V}_m = \mathbf{T}_m$, now contains K elements, where project k receives a yes vote if $U_k \geq U_0$ and a no vote otherwise. Expected utility given $\mathbf{T}_m$ and arbitrary votes by all other participants $(\mathbf{V}_{-m})$ is denoted by

$$(4)\quad EU_m(\mathbf{T}_m, \mathbf{V}_{-m}; G) = \mathbf{P}(\mathbf{T}_m, \mathbf{V}_{-m}; G) \cdot \mathbf{U} + P_0\, U_0$$
$$= \sum_{k=1}^{K} P_k(\mathbf{T}_m, \mathbf{V}_{-m}; G) U_k + \left(1 - \sum_{k=1}^{K} P_k(\mathbf{T}_m, \mathbf{V}_{-m}; G)\right) U_0.$$

The rest of the proof proceeds as before. An arbitrary variation away from $\mathbf{T}_m$ results in a change of expected utility

$$(5)\quad \Delta EU_m = EU(\mathbf{V}_m, \mathbf{V}_{-m}; G) - EU(\mathbf{T}_m, \mathbf{V}_{-m}; G)$$
$$= \sum_{k=1}^{K} \Big\{ \big[P_k(\mathbf{V}_m, \mathbf{V}_{-m}; G) - P_k(\mathbf{T}_m, \mathbf{V}_{-m}; G)\big] U_k \Big\}$$
$$+ \sum_{k=1}^{K} \Big\{ \big[P_k(\mathbf{T}_m, \mathbf{V}_{-m}; G) - P_k(\mathbf{V}_m, \mathbf{V}_{-m}; G)\big] U_0 \Big\}.$$

If F maintains independence between choice sets and the one-to-one correspondence condition holds, the effect of changing the vote on project k is strictly limited to modifying P_k and P_0. This reduces equation (5) to

$$(6)\quad \Delta EU_m = \big(P_k(\mathbf{V}_m, \mathbf{V}_{-m}; G) - P_k(\mathbf{T}_m, \mathbf{V}_{-m}; G)\big)(U_k - U_0).$$

Weak positive monotonicity guarantees once again that $\Delta EU_m \leq 0$ and that deviating from $\mathbf{T}$ cannot increase the participant's expected utility.

Independence between choice sets and one-to-one correspondence also guarantees that the impact of changing several votes on expected utility can be reduced to the sum of each specific change. These individual marginal effects are always non-positive regardless of $\mathbf{V}_{-m}$. We conclude that truthful revelation in a DCE is therefore a weakly dominant strategy. Truthful revelation is strictly dominant and the resulting Nash Equilibrium is unique if the mechanism maintains strict monotonicity $\left(\Delta P_k/\Delta V_k > 0\right)$ everywhere and if $U_k \neq U_0$ for all projects.

E. Incomplete Information and Uncertainty

The IC of DCEs can continue to hold if participants are uncertain about the exact shape of F, or if they form their own beliefs about how votes will be used by the authority. There is no updating process possible, and so the participants' initial priors about F form the basis of the optimization problem they must solve. As long as beliefs are consistent with the conditions laid out, the elicitation will remain incentive-compatible.

Incomplete information about the preferences of others is one source of uncertainty for participants. Participants in a collective decision setting might be expected to form beliefs about the preferences and beliefs of others in order to infer how they might vote. From a theoretical perspective, these beliefs are largely immaterial in an SBC elicitation. Only in some extreme cases would beliefs about others result in a loss of IC. One example is a majority-rule referendum where a participant believes with absolute certainty that others form a majority. This participant puts a zero probability on the possibility that his vote could be pivotal (i.e., consequential). Whether this belief is correct or not is irrelevant if they form the basis of decisions. The violation of consequentiality (in the belief space) eliminates the incentives of the voter to truthfully reveal his preferences.

Any other type of beliefs about other voters leaves intact the set of conditions required for IC. In particular, an agent who has uninformative/flat priors about other voters (i.e., he believes that all other voters have equal probabilities of voting yes or no), must still consider the properties of the policy function in order to ascertain whether or not it is in his best interest to vote truthfully. Hence, flat priors about others do not negate the need for a well-behaved policy function (or voters' beliefs about it).[8]

Incomplete information about the policy function is another important source of uncertainty. In most policy-related DCEs, participants are not given precise information about how their choices translate into policy or agency action. It is therefore useful to consider how participant uncertainty about the nature of the policy function F might affect incentives.

One way of modeling this uncertainty is to postulate that participants form beliefs about a number of possible policy functions and their respective implementation probabilities. In this context, the DCE can still be incentive compatible if each conjectured policy function respects the previously elaborated conditions; and each policy function has $\Delta P_k/\Delta V_k \geq 0 \ \forall k$ with at least one function where $\Delta P_k/\Delta V_k > 0$ some of the time for all k.

[8] This discussion extends to a single choice question, where a participant is asked to choose among more than two options. In such a setting, flat priors about other voters can negate the classic Condorcet problem whereby one might vote for his second choice instead of his first if it sufficiently increases the likelihood that this second choice will prevail over the third-ranked option. However, such analysis still requires that the choice mechanism meets certain properties (e.g., monotonicity).

In essence, previous conditions are extended to the beliefs underlying the participant's decision. They form a strong enough set that even when the participant considers the decisions of other participants, there is no strategic incentive to deviate from truthful preference revelation. A participant could even conjecture that others have beliefs that are inconsistent with the sufficient conditions and it would still be optimal for him to vote according to his preferences. The converse is not true, however. There are no beliefs about other players, their preferences and their own beliefs that absolve the mechanism from the need to meet the stated conditions.

In a setting characterized by two or more choice questions, participants are arguably more likely to conjecture that the policy process involves some aggregation of all participants' responses to the different choice questions. Participants are then left to form their own beliefs about how the projects relate to policies that may actually be devised; what the true cost of such policies might be; how their choices affect the policy design; and how their choices impact the likelihood that any particular policy will be implemented.

It is, therefore, difficult to be confident that the beliefs of all participants will satisfy the requirements for IC. In particular, one-to-one correspondence seems unlikely to hold which raises doubts that the independence condition will be met. For individual votes to truthfully reflect preferences over the projects presented, the impact of each vote must still affect only one distinct possible policy outcome. If, in the participant's beliefs, one vote can affect more than one policy; or, if more than one vote can affect the likelihood of a single policy, votes on individual choice questions will once again reflect potential trade-offs between costs, benefits, and probabilities of implementation across different policies and IC will be lost.

F. Combining Binding Project Implementation and Policy Input

We now briefly discuss the conditions for IC as they directly relate to our field experiment. In all experiment treatments, participants are told that their choices would inform the development of policies by government agencies. In addition, in three treatments, projects (including the cost component) can actually be implemented as a result of participant choices. These three treatments thus involve both a direct project and a broader policy component, and the resulting incentives must be considered jointly. Here, a participant's votes must maximize

$$(7)\quad EU_m(\mathbf{V}_m, \mathbf{V}_{-m}; G) = \sum_{k=1}^{K}\sum_{b=1}^{B} P_{k,b}(\mathbf{V}_m, \mathbf{V}_{-m}; G)u(Y - c_k - C_b; a_k, \mathbf{A}_b) + \sum_{k=1}^{K} P_{k,0}(\mathbf{V}_m, \mathbf{V}_{-m}; G)u(Y - c_k; a_k) + \sum_{b=1}^{B} P_{0,b}(\mathbf{V}_m, \mathbf{V}_{-m}; G)u(Y - C_b; \mathbf{A}_b) + P_{0,0}U_0,$$

where the subscript b denotes the participant's beliefs about the broader policies that might be devised by authorities, and their attributes and costs $(\mathbf{A}_b, C_b)$. The possible states of the world resulting from the experiment are characterized by joint probability/priors, $P_{k,b}$, that a given direct project and a given policy will be implemented. $P_{k,0}$ is the probability that project k is implemented while no policy is put in place, and $P_{0,b}$ is the probability that policy b is put in place while no direct project k is implemented. Finally, $P_{0,0}$ is the probability that the status quo remains.

For these three treatments to be incentive compatible, one of following three conditions must hold in addition to the conditions discussed previously:

(1) the projects maintain the one-to-one matching properties with *both* the actual projects that can be directly implemented, and with the broader policies or the beliefs $(\mathbf{A}_b, C_b)$ about them;

(2) the policy component is inconsequential (or believed to be): $\Delta P_{k,b}/\Delta V_k = 0 \forall k$; $\Delta P_{0,b}/\Delta V_k = 0 \forall b$; or

(3) the direct project component is inconsequential (or believed to be): $\Delta P_{k,b}/\Delta V_k = 0 \forall k$; $\Delta P_{k,0}/\Delta V_k = 0 \forall k$.

It is not an oversight that conditions (2) and (3) call for the *in*consequentiality of one of the two roles played by the DCE. Each of these two conditions has the effect of reducing the DCE to being exclusively about the direct implementation of a project, or about a broader policy, but not both. If either condition applies, the DCE is returned to a single purpose and the relevant analysis of incentives performed previously applies.

The joint situation is more complex since each vote has the potential to affect both the probabilities of implementing a direct project, and the probabilities of implementing more general public policies. The absence of a one-to-one correspondence between the choice sets, direct projects and the public policies (or beliefs) almost certainly leads to a failure of IC. Without the one-to-one correspondence, a vote has a marginal effect on the project and one (or more) public policy with different attributes. It follows directly that a participant's votes on the projects cannot be guaranteed to truthfully represent his preferences over the projects themselves.

Once again, this result suggests that a field experiment that combines both direct incentives, and the potential for broader public impacts, is not likely to be incentive-compatible. Behaviorally and empirically, however, it might be sufficient for "good measurement" that the direct incentives have a much greater marginal impact on the participant's utility function than the uncertain and more distant policy ramifications. We address this issue in our interpretation of experiment results.

II. Field Experiment Design

A. *Description of Projects and Choice Sets*

The field experiment component of this research is part of a broader study on the value of goods and services linked to agroforestry. Our experimental design allows the exploration of methodological issues surrounding the use of DCEs to elicit the

Table 1—Discrete Choice Experiment Attributes and Levels

Attribute	Description	Levels
Length	Length of tree planting, in meters	300, 600, 1000
Width	Number of rows in tree planting	1, 3
Location	Location of tree planting	Riparian (along stream or river) Windbreak (along roadside)
Cost	Cost of tree planting, in C$	10, 25, 50, 75

value of public goods. More specifically, experiment treatments that span the spectrum from incentive-compatible, financially binding decisions to decisions with no direct financial consequences are used to elicit values for riparian and windbreak tree planting projects on agricultural land in the province of Quebec (Canada).

Salient project attributes and costs were identified through multiple pretests involving 140 individuals. Three attributes describe the tree planting projects: *Location*, whether riparian (e.g., along a stream) or as a windbreak alongside a road; *Length*, that is, the linear meters of streamside or roadside covered by the planting; and *Width*, the number of rows of trees to be planted. Associated with each project is a *Cost*, which is the amount that the participant would have to pay if the project was implemented.

Table 1 presents the project attributes, costs and their levels. The *Location* and *Width* attributes each have two levels, the *Length* attribute has three levels, and *Cost* has four levels. A full factorial design would therefore require $2 \times 2 \times 3 \times 4 = 48$ different choice sets. To reduce the total number of choice sets, we generated 24 unique choice sets using the SAS macro *%mktex* with the objective of maximizing *D*-efficiency, while maintaining the ability to statistically identify all main effects and two-way interactions (excluding *Cost* from the interactions).[9] The resulting *D*-efficiency is 99 percent, and the design has perfect balance with respect to the three project attributes, i.e., each of the possible $2 \times 2 \times 3$ "goods" appears twice in the design. These 24 choice sets were split into two blocks, with each block containing all 12 goods (once) such that the lone difference across blocks was the project costs. Participants were randomly assigned one of the two blocks in the experiment. To remain consistent with field survey settings, the fact that different participants saw different costs was not disclosed. Each choice set was presented as a separate binary choice referendum whereby a "yes" vote is a choice in favor of the tree planting project, and a "no" vote favored the status quo.

B. Experiment Treatments

The study design includes four treatments. We label the first three as real payment (*RP*) treatments, as participants have the direct opportunity to fund an actual tree planting project through their choices. Two of these three treatments are designed

[9] It is a common practice to assume that the variance of WTP does not depend on the levels of attributes and, since the variance is identified through variation in *Cost*, this eliminates the need for interactions involving *Cost*.

to be theoretically incentive compatible if the potential influence of the participant's choices on broader policy decisions is ignored or negligible. We label the fourth as the stated preference (*SP*) treatment, since no project could be implemented as a direct result of participant choices. In all treatments, it is common knowledge that information will be shared with policymakers.

One aspect of our design bears noting since it has potential importance when interpreting empirical results. In particular, there was no explicit payment vehicle associated with broader policies that might be implemented. As a result, there was no indication to *SP* respondents whether payment—in the event a broader policy is implemented—would be collected through donations or coercively, such as through a tax. Although this was a deliberate design choice to maintain parallelism across treatments, this is one aspect of our design that differs from some field survey applications.[10]

From an incentive standpoint, this design feature is likely to weaken the perceived consequences of the policy component in all four treatments. The potential problem comes from whether participants, especially those in the *SP* treatment, believe that the costs stated in the survey reflect the potential costs to them upon implementation. If anything, the weak payment mechanism in the *SP* treatment should be expected to make it easier to statistically reject that the *SP*-based WTP function is equal to those estimated from the incentive-compatible *RP* treatments. This will make one of our main results—that *SP* and *RP* treatments produce similar WTP functions but only when participants believe the survey to be sufficiently consequential—a conservative statement about the ability of DCEs to produce valid value estimates.

Below, we provide details on the four treatments in the experimental design, which were implemented using split samples. The order in which we present treatments reflects our expectation regarding the likelihood of IC.

Real Payment, *Independent Lottery Policy Function*.—The first treatment, our baseline, is labeled *Real payment with Independent Lottery Policy Function* (*RP-IND*). Participants' votes probabilistically lead to the implementation of one of the 12 projects or the status quo. Participants are instructed that one of the 12 choice sets will be randomly chosen at the end of the experiment, each with equal probability.[11] This random selection procedure was chosen to make it clear that choice sets are independent. For the selected choice set, the proportion of yes votes among all participants is computed. Then, the project identified in the choice set is implemented with a probability equal to the proportion of yes votes it received. Under this probabilistic provision rule (and unlike the common majority-vote rule), the voter has a strictly positive probability of being pivotal.

[10] We note, however, that the exclusion of a coercive payment vehicle is a fairly common practice for field survey DCEs.

[11] This type of random selection procedure is commonly used in experiments as a means of allowing multiple decisions while keeping the incentives presented in individual situations undisturbed. If participants are expected utility maximizers, this maintains the basic incentives to seriously consider choices individually, uncontaminated by the risk of having to pay for multiple projects. While 1-in-12 confers only a small probability that any given choice set will be selected, the mechanism ensures with probability 1 that one of the choice sets will be selected, with a possible individual cost of up to C$75. Viewed from this perspective, each voting decision is now potentially a costly one if care is not taken to consider its implications.

In execution, one of the 12 choice sets is selected by rolling a 12-sided die. Then two 10-sided dice are used to obtain a number between 0 and 99, which becomes the "acceptance level." If the percentage of yes vote equals or exceeds the acceptance level, the project is accepted and each participant pays his individual cost amount. The selected project, as described, would then be undertaken. Otherwise, no money is collected and no project is carried out.

If we record a yes vote on project k by participant i as $V_{ik} = 1$, a no vote as $V_{ik} = 0$ and if we define $W_k = \sum_{m=1}^{M} V_{mk}$, the policy function (as it pertains to direct project selection only) takes the form $F(\mathbf{V}, \mathbf{V}_{-m}; G) = \left(\frac{W_1}{KM}, \frac{W_2}{KM}, \ldots, \frac{W_k}{KM}, \ldots, \frac{W_K}{KM}\right)$, where KM is the total number of votes cast. If beliefs about broader possible government policies are consistent with the choice sets (condition (1)) or if the broader policy implications are inconsequential (condition (2)) or perhaps if they are considered much less important than the real and immediately costly project implementations, then $\Delta P_k / \Delta V_k > 0 \; \forall k$, at least as a first-order effect. By Proposition 2, this mechanism is incentive compatible and truthful preference revelation is a dominant strategy.

Real Payment, *Aggregate Lottery Policy Function*.—In the second treatment, *Real Payment with Aggregate Lottery Policy Function* (*RP-AGG*), each vote for a particular project, or the status quo, directly maps into the probability that this project or status quo is implemented. This is made operational by assigning a separate color to each of the 12 projects and the status quo. Each participant's yes vote adds a poker chip of a distinct color to a bag. A no vote, favoring the status quo, adds a black colored chip. After all votes are cast, the bag is filled with $(W_1, W_2, \ldots, W_K)$ of the K different colored chips, and $\left(KM - \sum_{k=1}^{K} W_k\right)$ black chips (for a total of KM chips). The outcome of the experiment is determined by a single draw from the bag. If a colored chip is drawn, the corresponding project is implemented. If a black chip is drawn, the status quo prevails.

With this procedure, the probability that project k is implemented is once again given by W_k / KM. The policy function is in effect identical to that of the *RP-IND* treatment. It is similarly incentive compatible with truthful demand revelation as a dominant strategy if we abstract from broader policy implications (i.e., if condition (1) or (2) holds). However, as votes for all choice sets are considered simultaneously in the aggregation process, we posit that it could be less clear to participants that the independence property holds.

Real Payment, *Unknown Policy Function*.—In the third treatment, *Real Payment with Unknown Policy Function* (*RP-UNK*), the elicitation is again consequential and there are direct financial consequences. However, while participants are informed that their choices will be used to determine which project may be implemented, no description of the policy function is given to participants. Since each participant only participates in a single treatment, they have no indication of what the policy function might be. Participants are left to form beliefs about how a project (or status quo) might be selected, and therefore how votes could influence the outcome of the experiment. Without information about the policy function, there can be no guarantee that participants' beliefs maintain independence or that the mechanism is incentive compatible even when the direct project component is considered in isolation.

After all choices were made, participants learned the provision rule. One of the 12 choice sets was selected at random using a 12-sided die, and a simple majority-vote rule determined whether the selected project would be carried out.

Stated Preference.—The last treatment, *Stated Preference* (*SP*), is a DCE wherein participants vote on the 12 projects without direct financial consequences. As in other treatments, participants are told that the results of the study would be provided to a government agency. To the extent that they believe that their choices can influence actual policy decisions (including the cost assessed), this too is a consequential elicitation. However, the caveats expressed earlier regarding the payment vehicle and the likelihood that beliefs meet the conditions for IC apply.

C. Experiment Protocol and Survey Description

Aside from the changes in provision rule, the experimental protocol was identical across all treatments. Participants were given a show-up payment of 100 Canadian Dollars (C$) in the *RP* treatments and C$50 in the *SP* DCE.[12,13] The difference in the show-up payments was meant to equate expected earnings across treatments and their determination was based on our pilot data.

Experiment instructions (available as an online Appendix) were presented by the same moderator in all sessions, with the help of a PowerPoint presentation. The slides included a brief introduction of the study along with a description of the ecological role of trees and of the attributes focused on in the study (length, width, and location along a water body or as a windbreak along a road). Representative examples of tree-row plantations were presented using computer-edited photographs to illustrate plantations and the attribute levels that participants would vote on. The presentation for the first two *RP* treatments also included two examples of how the policy function works (e.g., how the project implementation would be decided after the votes). Participants were also given a paper copy of the photographs and a sheet summarizing the benefits of various riparian and windbreak tree plantings. Participants were informed that their surveys would be shared with policymakers, but that anonymity would be maintained, as per the approved human ethics research protocol.

For each treatment, six versions of the survey were randomly distributed to participants to be completed using a pencil. The versions differed in the order the choice sets were presented to minimize possible order effects and by the different cost vectors.

[12] Participants were recruited by announcing an expected gain of C$50 with the understanding that the amount might be greater or lower, depending on the outcome of the session.

[13] One concern is that experiment participants treat earnings as "house money" rather than as part of their income (Thaler and Johnson 1990). Although house money effects have been documented in other contexts, the study by Cherry, Kroll, and Shogren (2005) suggests that this effect might not be prevalent in a one-shot public goods experiment. To help mitigate possible house money effects, payment was given in cash upon arrival. Participants were told orally and in writing that this money was theirs to keep and informed that they were free to leave at any moment and still retain the show-up payment. Two participants chose to do so.

Aside from the discussion of the method used to select a project for direct implementation, the process closely paralleled a field survey. In the first section of the survey, participants were asked attitudinal questions on environmental issues. The second section contained the DCE. The third section included typical demographic questions, in addition to questions intended to gauge consequentiality and the extent to which participants had considered other participants' possible preferences and votes when they made their own decisions.

D. *Participants*

Participants were volunteers recruited through a mailing list of employees and friends of Laval University, and a mailing list of the Institut des Nutraceutiques et des Aliments Fonctionnels (INAF). These lists included roughly 5000 people who had expressed an interest in receiving news from Laval University or INAF. Participants needed to be at least 18 years old and could only participate once. Eight sessions, two for each treatment, took place between March and June 2009 in the INAF building located in Quebec City. Volunteers were randomly assigned to a treatment to facilitate the identification of treatment effects.

Two hundred and twenty participants completed the experiment, with roughly equal numbers of participants in the four treatments: 58, 55, 52, and 55 respectively.[14] Three projects were actually implemented as the result of votes in the six *RP* sessions: a 1 km by 3 rows windbreak, a 1 km by 1 row windbreak, and a riparian band of 1 km by 3 rows.

Table 2 presents some basic demographic information on our participants. We note that the demographics are similar across treatments (detailed information available upon request). Overall, the average participant has a higher household income and is better educated relative to the general population of Quebec. The age distribution (and average age), employment rate, and percentage of males are similar. As the primary goal of the study is to provide insight on important methodological issues, we make no claims about sample representativeness.

III. Results

A. *Analysis of Willingness to Pay*

We begin our analysis with the estimation of a WTP regression based on the maximum likelihood estimator of Cameron and James (1987). In particular, we treat $WTP_{t,ik}$ as a censored dependent variable for which we obtain the signal $WTP_{t,ik} \geq c_{t,ik}$ if, in treatment t, participant i votes "yes" to cost $c_{t,ik}$ associated with a project k, or the signal $WTP_{t,ik} < c_{t,ik}$ if participant i votes no.

Let $WTP_{t,ik}$ be a linear function of a row vector of covariates, $\mathbf{x}_{t,ik}$, such that $WTP_{t,ik} = \mathbf{x}_{t,ik}\boldsymbol{\beta}_t + \varepsilon_{t,ik}$, where $\boldsymbol{\beta}_t$ is a column vector of unknown parameters and

[14] Since treatments could be run with few constraints on the number of participants, no participant was refused for overbooking reasons. This explains the small difference in the number of participants across treatments. All participants who participated were retained in the sample.

Table 2—Descriptive Statistics

Variable name	Description	Sample mean (SD)
Graduate degree	= 1 if participant has a graduate degree	0.35 (0.48)
Income	Participant's household income, in C$1,000s; the midpoint of the category chosen by the participant is used	62.34 (40.58)
Household size	Number currently living in the participant's household	2.57 (1.45)
Male	= 1 if participant is Male	0.45 (0.50)
Age	Participant's age, in years	40.30 (13.80)
College degree	= 1 if participant has a college degree or higher	0.62 (0.49)
Employed	= 1 if participant is employed	0.77 (0.42)
Environmental	= 1 if participant is currently a member of an environmental organization	0.07 (0.25)
Charity	= 1 if participant donated to any charity in past 12 months	0.82 (0.38)
Consequential	Participant's indication of the indirect consequences of DCE choices; 1 ("not at all") to 6 ("very strong") likert scale question	3.54 (1.04)

$\varepsilon_{t,ik}$ is a normally distributed mean-zero error term with treatment-specific standard deviation σ_t.[15] With the linear conditional mean function, assuming the error term has a normal distribution is analogous to assuming a normal WTP distribution. Let $y_{t,ik} = 1$ denote a "yes" vote and $y_{t,ik} = 0$ indicate a "no" vote. Further, denote $I_{t,ik}$ as an indicator variable that equals 1 if participant i faces treatment t and equals 0 otherwise. Then, the log-likelihood function is

$$\text{(8)} \quad \ln L = \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{k=1}^{K}\left\{y_{t,ik}\ln\left[1-\Phi\left(\frac{c_{t,ik}-\mathbf{x}_{t,ik}\boldsymbol{\beta}_t}{\sigma_t}\right)\right]\right.$$
$$\left. + \ (1-y_{t,ik})\ln\left[\Phi\left(\frac{c_{t,ik}-\mathbf{x}_{t,ik}\boldsymbol{\beta}_t}{\sigma_t}\right)\right]\right\}.$$

Included as covariates are project attributes, as well as control variables for household size, income and attainment of a graduate degree.[16] To control for unmeasured factors specific to an individual we estimate cluster-robust standard errors. With

[15] The importance of allowing for different error variances across treatments has been highlighted in a number of recent papers in the stated preference literature (Cameron 2002; Carlsson and Johansson-Stenman 2010; Salisbury and Feinberg 2010).

[16] In our initial specification, we included all demographic variables defined in Table 2, as well as interactions between all non-cost project attributes. The more parsimonious specification we present is justified by statistical tests.

our functional form and error distribution assumption, interpretation of estimated parameters is analogous to that of a standard linear regression model that treats *WTP* as a directly observed (i.e., uncensored) dependent variable.[17] Estimation is carried out with user-defined maximum likelihood procedures programmed by the authors in Stata.

The estimation results appear as Model 1 in Table 3. For all treatments, the estimated marginal WTP for the *Length*, *Width*, and *Location* attributes are statistically significant (beyond the 1 percent level) and economically meaningful. For instance, *ceteris paribus*, the estimated WTP function for the *RP-IND* treatment suggests that participants are willing to pay about C$0.04 for every one meter increase in length (or C$4 for a 100m increase), C$12.48 for an additional row of trees and an additional C$9.75 if the tree planting is in a riparian area. This suggests a high WTP for many of the tree projects offered to participants. As further evidence of construct validity, those with higher income and education (the latter effect is only significant at the 7 percent significance level) are willing to pay more, whereas participants with larger households (which likely reflects higher demands on household income) are willing to pay less.

The standard deviation of *WTP* is lowest for the *RP-IND* treatment. Recall that this is the treatment hypothesized to be most transparent in terms of its direct financial incentives and the project selection rule. Relative to the *RP-IND* treatment, the standard deviation of the error term for the other three treatments is statistically different. In particular, it is about 50 percent higher in the *RP-UNK* and *SP* treatments, and about 70 percent higher in the *RP-AGG* treatment. The higher variances for these treatments may reflect greater uncertainty over the policy function, the role of preferences of other agents and/or the interaction of these two factors. A higher variance in the *SP* treatment may reflect decision error stemming from participants who do not perceive the elicitation as consequential.

Applying Wald tests, we test for equality in the estimated WTP functions across treatments (in particular, equal marginal WTP for all attributes as well as an identical intercept), and present the results in Table 4. The results can be summarized as follows: the WTP functions for the three *RP* treatments are statistically identical; and the *SP* function is statistically different from that of any other treatment. Based on additional attribute-specific tests, we fail to reject the null hypothesis of equal marginal WTP for any attribute (even at the 10 percent level), for any pairwise test involving two *RP* treatments. In contrast, the marginal WTP for *Length* and *Width* are statistically higher for the *SP* treatment. To put this into perspective, the estimated mean WTP for a moderate-level project ($Length = 600$; $Width = 3$; $Location = 0$) is C$45.05 in the *RP-IND* treatment compared to C$59.63 in the *SP* treatment (a 32 percent increase). These means are statistically different [$\chi^2(1) = 6.10, p = 0.01$].

[17] We also estimated models with logistic, log-logistic, log-normal, and Weibull distributions for the error term. The signs and significance of slope coefficients, as well as the main conclusions we reach regarding treatment effects, are consistent across the alternative specifications.

Table 3—Willingness to Pay Regressions

	Model 1	Model 2
Real payment, Independent Lottery Policy Function Rule (*RP-IND*)		
Length [meters]	0.039***	0.039***
	(0.004)	(0.004)
Width [rows of trees]	12.475***	12.493***
	(1.399)	(1.405)
Location [= 1 if Riparian; = 0 if Windbreak]	9.745***	9.773***
	(2.710)	(2.714)
Intercept	−15.036***	−15.185**
	(6.095)	(6.091)
Scale (σ)	21.637***	21.741***
	(2.191)	(2.216)
Real payment, Aggregate Lottery Policy Function Rule (*RP-AGG*)		
Length [meters]	0.035***	0.035***
	(0.007)	(0.007)
Width [rows of trees]	11.355***	11.345***
	(2.374)	(2.368)
Location [= 1 if Riparian; = 0 if Windbreak]	9.244***	9.210***
	(3.448)	(3.441)
Intercept	−3.674	−3.755
	(8.698)	(8.681)
Scale (σ)	37.099***	36.945***
	(5.012)	(4.966)
Real payment, Unknown Policy Function Rule (*RP-UNK*)		
Length [meters]	0.039***	0.039***
	(0.007)	(0.007)
Width [rows of trees]	14.853***	14.851***
	(1.819)	(1.816)
Location [= 1 if Riparian; = 0 if Windbreak]	17.584***	17.591***
	(4.603)	(4.602)
Intercept	−28.832***	−29.015***
	(9.935)	(9.912)
Scale (σ)	32.507***	32.492***
	(4.293)	(4.306)
Stated Preference (*SP*)		
Length [meters]	0.063***	0.139***
	(0.006)	(0.028)
Length × *Consequential*		−0.021***
		(0.007)
Width [rows of trees]	17.625***	17.867
	(2.579)	(11.985)
Width × *Consequential*		−0.080
		(3.06)
Location [= 1 if Riparian; = 0 if Windbreak]	15.930***	43.319***
	(4.178)	(17.888)
Location × *Consequential*		−7.578*
		(4.281)
Intercept	−30.333***	−47.051
	(7.978)	(31.672)
Consequential		4.426
		(7.885)
Scale (σ)	33.099***	34.389***
	(3.554)	(8.064)
Scale × *Consequential*		−0.959
		(2.096)
Log-likelihood	−1,307.844	−1,289.584
Observations	2,640	2,640

Notes: Cluster-robust standard errors are in parentheses. Both models also include the control variables *Graduate degree, Income,* and *Household size.*

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 4—Tests of Equal Willingness to Pay Functions

Hypothesis	χ^2	*p*-value
RP-IND = *RP-AGG*	2.38	0.6665
RP-IND = *RP-UNK*	4.19	0.3806
RP-AGG = *RP-UNK*	6.80	0.1466
RP-IND = *SP*	19.29	0.0007
RP-AGG = *SP*	12.39	0.0147
RP-UNK = *SP*	20.36	0.0004
RP-IND = *RP-AGG* = *RP-UNK*	7.82	0.4515
RP-IND = *RP-AGG* = *RP-UNK* = *SP*	30.72	0.0022

Notes: Tests are based on Model 1, allowing for unequal variances. Key to Abbreviations: *RP-IND* = *Real payment, Independent Lottery Policy Function*; *RP-AGG* = *Real payment, Aggregate Lottery Policy Function*; *RP-UNK* = *Real payment, Unknown Policy Function*; *SP* = *Stated Preference*.

B. *Additional Analysis Based on Subjective Measures*

Strategic Consideration of the Preferences of Other Participants.—In treatments *RP-IND* and *RP-AGG*, Proposition 2 holds when the direct project policy function is considered in isolation of broader public policy influences. In practice, unless condition (1) or (2) is satisfied, the elicitation mechanisms are not incentive compatible. Furthermore, and as discussed in the theory section, strong assumptions are needed to ensure IC of the *RP-UNK* and *SP* treatments.

IC can be lost when a participant perceives a lack of independence between voting choices. In such a situation, it can be optimal to vote strategically after considering how his votes and those of others influence the decision maker. Recall that our main theoretical result is that under the sufficient conditions for IC, it is a dominant strategy to vote truthfully on each project no matter how other participants are voting. Signs that a participant has considered the votes of others when choosing his own *could* therefore indicate strategic voting, and a loss of IC. In an effort to measure the strategic consideration of other participant preferences, we asked participants (after they had voted but before the project selection procedure in the *RP* treatments) whether they had considered how other participants might vote. When applicable, participants were asked to indicate how many of their own votes were affected by their consideration of others' votes. For our purposes, participants indicating such voting behavior are characterized as voting strategically.

Responses by treatment are presented in Table 5. The most striking finding is that just one *SP* participant indicated strategic voting. There is some evidence of strategic voting in the *RP-IND* and *RP-AGG* treatments, where roughly 10 percent indicated that strategic motivations affected two or more votes. The most support for strategic voting comes from the *RP-UNK* treatment, where 25 percent indicate having considered how other participants might vote.

The impact of strategic voting appears modest at best. For the *RP-UNK* treatment, less than 6 percent suggested that strategic considerations altered three or more votes and just one of these participants indicated it had affected more than five votes. In terms of the number of strategic voters, using pairwise Fisher Exact Tests,

Table 5—Evidence of Strategic Voting (*considerations of other participants' choices*)

Response	*RP-IND*	*RP-AGG*	*RP-UNK*	*SP*
None at all	51	48	39	54
Affected 1 vote	2	2	5	1
Affected 2 votes	4	3	5	0
Affected 3 to 5 votes	1	1	2	0
Affected >5 votes	0	1	1	0
Total	58	55	52	55

Note: Key to Abbreviations: *RP-IND* = *Real payment, Independent Lottery Policy Function*; *RP-AGG* = *Real payment*, *Aggregate Lottery Policy Function*; *RP-UNK* = *Real payment, Unknown Policy Function*; *SP* = *Stated Preference.*

we fail to reject equality among the *RP* treatments. However, we reject equality, at least marginally, when comparing the *SP* and *RP* treatments.[18]

As a robustness check, we re-estimated the censored regression model while excluding the handful of participants for which there is evidence of strategic voting. All of our main results are robust to this exclusion. The overall evidence suggests that while some strategic voting may be present in the RP treatments, its influence is negligible.[19]

Indirect Policy Consequences.—IC of the *SP* treatment requires that participants perceive their votes to have some influence on broader policy decisions. The question we asked all participants (after they had voted) roughly translates from French as "To what extent do you believe that your choices will be taken into account by public authorities?" The six response categories, and the number of participants that selected each, are presented in Table 6. Since the IC of the *SP* treatment requires that participants believe their answers can influence policy decisions, it is useful to explore the link between the answers to this question and elicited preferences.

First, we expanded the econometric model by adding interactions between all *SP* treatment variables and the variable *Consequential*, defined as the likert-scale response to the question regarding the consideration of votes by public authorities.[20] The estimated model is presented as Model 2 in Table 3. All coefficients on the attribute interaction variables are negative, suggesting that the marginal WTP for each project attribute is decreasing in the degree of consequentiality. The interaction with *Length* is significant at the 1 percent level whereas the interaction with *Location* is significant at the 10 percent level. The conditional WTP function for the *SP* treatment is statistically equal to the (unconditional) WTP function for the *RP-IND*,

[18] The *p*-values associated with the comparisons to *RP-IND*, *RP-AGG*, and *RP-UNK* are 0.079, 0.066, and 0.001, respectively.

[19] Affirmative responses to our strategic voting question could simply indicate other-regarding preferences such as altruism and equity (see for instance Messer et al. 2010). If so, these responses would not indicate potential violations of IC.

[20] In undertaking this analysis, we initially considered the *Consequential* variable (and corresponding interaction variables) as endogenous in the WTP regression. However, using the variables in Table 2 that met the exclusion restriction as well as occupation indicator variables as instruments, we fail to reject the hypothesis that the additional covariates included in Model 2 are jointly exogenous (this finding is based on a GMM over-identification test).

TABLE 6—BELIEFS ABOUT INDIRECT CONSEQUENCES (*degree/chance of policy influence*)

Response	*RP-IND*	*RP-AGG*	*RP-UNK*	*SP*
1 – Not at all	4	1	1	1
2 – Very weakly	2	11	7	5
3 – Weakly	13	20	16	19
4 – Moderately	25	17	22	25
5 – Strongly	11	5	4	4
6 – Very strongly	3	1	2	1
Total	58	55	52	55

Note: Key to Abbreviations: *RP-IND* = *Real payment, Independent Lottery Policy Function*; *RP-AGG* = *Real payment, Aggregate Lottery Policy Function; RP-UNK* = *Real payment, Unknown Policy Function*; *SP* = *Stated Preference*.

RP-AGG, and *RP-UNK* treatments when *Consequential* is evaluated at values of 4.1 and higher, 4.0 and higher, and 4.1 and higher, respectively. In other words, conditional upon participants perceiving their responses to have more than a weak level of policy influence, stated preferences are equal to revealed preferences.[21]

Second, to provide (further) empirical evidence that either condition (1) or (2) holds for the *RP* treatments, we estimated a model that allows all the treatment-specific coefficients to vary by reported perceptions of consequentiality. None of the added variables for the *RP* treatments have even marginally statistically significant coefficients. In fact, the coefficients are jointly equal to zero $[\chi^2(12) = 9.25, p = 0.68]$. This evidence strongly suggests that the direct financial consequences of *RP* choices have a much larger effect on marginal utility than the outside policy implications.

IV. Discussion

From a mechanism design perspective, eliciting truthful preferences for public goods using discrete choice experiments (DCEs) is a challenging task. As we have shown, incentive compatibility requires participants to believe not only that their decisions are consequential, but also that policymakers use the information in such a way that maintains choice set independence and one-to-one matching between the projects identified in choice sets and the policies that might be implemented. Furthermore, a number of alternative beliefs held by respondents are likely to fail the test of incentive compatibility and these beliefs cannot be directly observed or confidently controlled for in a field survey.

An important take-home message from our field experiment is that economic incentives do appear to be in play even in a stated preference survey where financial consequences are remote and the payment mechanism is vaguely articulated.

[21] To confirm this result, we re-estimated Model 1 while excluding the *SP* participants who perceived their influence to be weak at best (i.e., *Consequential* ≤ 3). Using Wald tests, we fail to reject equality between the *SP* treatment WTP function and any of the three *RP* functions: *RP-IND* = *SP* $[\chi^2(4) = 5.47, p = 0.2421]$; *RP-AGG* = *SP* $[\chi^2(4) = 4.89, p = 0.2992]$; *RP-UNK* = *SP* $[\chi^2(4) = 7.20, p = 0.1255]$.

Willingness-to-pay estimates exhibit only a modest (positive) bias—approximately 30 percent across the tree planting projects we investigate. Once we condition the analysis on the belief that responses had more than a "weak" impact on policy, however, stated and real payment willingness-to-pay functions are statistically identical. This finding contrasts with the sharper empirical results from field and laboratory studies that focus on consequentiality (Carson et al. 2004; Vossler and Evans 2009; Herriges et al. 2010), but nevertheless highlights the importance of consequentiality and the potential value of including pertinent survey questions to control for it. The evidence lends support to the view that the notion of consequentiality is far more important in assessing the criterion validity of surveys than the "real versus hypothetical" distinction.

We cannot definitively explain why participants in our study do not misrepresent their preferences in a stated-preference DCE. We speculate, however, that the complexity of the decision task and the opaqueness of the policy process are likely to limit the ability of participants to identify strategic links between survey choices. Our experimental evidence suggests that even with direct financial incentives, elicited willingness to pay does not necessarily depend on whether we explicitly specify an incentive-compatible implementation rule. However, we note that uncertainty over the decision rule did lead to a higher error variance. Participants who are uncertain about the policy function may develop beliefs that are consistent with our sufficient conditions, or, simply have too little information to stray away from their true preferences over each choice set.

Caution should be taken before broadly interpreting our evidence as it is likely that the degree of elicitation bias depends on survey design, as well as the goods being valued. In particular, we did not explicitly use a coercive payment mechanism in our stated preference treatment, and this may have weakened the incentives to answer truthfully. A coercive payment scenario might have brought the willingness to pay in the stated preference treatment even closer to that from the real payment treatments. Also, our results may not hold for stated preference DCEs conducted on a hotly debated topic where participants have entrenched political views or question the objectives of policymakers. In such cases, beliefs about the strategic value of non-truthful voting may be much stronger and introduce bias in the measurement of preferences.

Our analysis has been confined to binary DCEs with choice sets consisting of one public project and the status quo. The assumptions required for incentive compatibility are much more substantive when choice sets include three or more options (or involve private goods). Extending the theory to multiple options per choice set would likely require strong restrictions on the utility functions allowed and beliefs about the preferences of others. It remains unclear how valid these refinements might be, and how well they would inform empirical results. This remains an important area for future research.

Elements of our design—questions about indirect consequences and strategic voting in particular—proved very informative in the data analysis and interpretation of results. Yet, we have only taken a first step in this line of debriefing. Much insight could be gained by probing deeper into participants' beliefs when the consequentiality of the DCE is not clear. For instance, as the one-to-one correspondence and independence assumptions are crucial to incentive compatibility, participants could be asked questions about their perception of the relationship between choice sets. This could be combined with a split-sample comparison of surveys that include/do not include instructions for participants to treat questions as independent (such instructions are often included in practice).

Respondents in our real payment treatment with an undisclosed decision rule, and those in the stated preference treatment who see their responses as sufficiently consequential, did not behave significantly differently than those in real payment treatments with explicit decision rules. This perhaps bodes well for DCEs since field surveys do not explicitly characterize the mapping between responses and actual policy outcomes. Yet, given the central role played by respondents' beliefs about this mapping in both ensuring a theoretically incentive-compatible survey and obtaining reliable empirical estimations of preferences, additional efforts could be invested to understand the importance of incomplete information and uncertainty in DCEs.

REFERENCES

Adamowicz, Wiktor. 1998. "Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation." *American Journal of Agricultural Economics* 80 (1): 64–75.

Alpízar, Francisco, Fredrik Carlsson, and Peter Martinsson. 2003. "Using Choice Experiments for Non-Market Valuation." *Economic Issues* 8 (1): 83–110.

Beggs, Steven, Scott Cardell, and Jerry Hausman. 1981. "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics* 17 (1): 1–19.

Blass, Asher A., Saul Lach, and Charles F. Manski. 2010. "Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability." *International Economic Review* 51 (2): 421–40.

Cameron, Trudy Ann. 2002. "Alternative Non-Market Value-Elicitation Methods: Are the Underlying Preferences the Same?" *Journal of Environmental Economics and Management* 44 (3): 391–425.

Cameron, Trudy Ann, and Michelle D. James. 1987. "Efficient Estimation Methods for 'Closed-Ended' Contingent Valuation Surveys." *Review of Economics and Statistics* 69 (2): 269–76.

Carlsson, Fredrik, and Olof Johansson-Stenman. 2010. "Scale Factors and Hypothetical Referenda: A Clarifying Note." *Journal of Environmental Economics and Management* 59 (3): 286–92.

Carlsson, Fredrik, and Peter Martinsson. 2001. "Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?" *Journal of Environmental Economics and Management* 41 (2): 179–92.

Carson, Richard T. 1996. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics* 72 (1): 80–99.

Carson, Richard T., Nicholas Flores, and Norman Meade. 2001. "Contingent Valuation: Controversies and Evidence." *Environmental & Resource Economics* 19 (2): 173–210.

Carson, Richard T., and Theodore Groves. 2007. "Incentive and Information Properties of Preference Questions." *Environmental & Resource Economics* 37 (1): 181–210.

Carson, Richard T., Theodore Groves, John List, and Mark Machina. 2004. "Probabilistic Influence and Supplemental Benefits: A Field Test of the Two Key Assumptions Underlying Stated Preferences." Unpublished.

Carson, Richard T., Michael Hanemann, and Dan Steinberg. 1990. "A Discrete Choice Contingent Valuation Estimate of the Value of Kenai King Salmon." *Journal of Behavioral Economics* 19 (1): 53–68.

Cherry, Todd L., Stephan Kroll, and Jason F. Shogren. 2005. "The Impact of Endowment Heterogeneity and Origin on Public Good Contributions: Evidence from the Lab." *Journal of Economic Behavior and Organization* 57 (3): 357–65.

Cramton, Peter, Yoav Shoham, and Richard Steinberg. 2006. "Introduction to Combinatorial Auctions." In *Combinatorial Auctions*, edited by Peter Cramton, Yoav Shoham, and Richard Steinberg, 1–13. Cambridge, MA: MIT Press.

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* 8 (4): 45–64.

Green, Paul E., and Vithala R. Rao. 1971. "Conjoint Measurement for Quantifying Judgmental Data." *Journal of Marketing Research* 8 (3): 355–63.

Hanley, Nick, Susana Mourato, and Robert E. Wright. 2001. "Choice Modelling Approaches: A Superior Alternative for Environmental Valuation?" *Journal of Economic Surveys* 15 (3): 435–62.

Hanley, Nick, Robert E. Wright, and Wiktor Adamowicz. 1998. "Using Choice Experiments to Value the Environment: Design Issues, Current Experience and Future Prospects." *Environmental & Resource Economics* 11 (3–4): 413–28.

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.

Holmes, Thomas P., and Wiktor L. Adamowicz. 2003. "Attribute-Based Methods." In *A Primer on Non-Market Valuation*, edited by Kevin J. Boyle, Patricia A. Champ, and Thomas C. Brown, 171–219. Boston: Kluwer Academic Publishers.

Ida, Takanori, and Rei Goto. 2009. "Simultaneous Measurement of Time and Risk Preferences: Stated Preference Discrete Choice Modeling Analysis Depending on Smoking Behavior." *International Economic Review* 50 (4): 1169–82.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74 (1): 132–57.

List, John A., and Craig A. Gallet. 2001. "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis." *Environmental & Resource Economics* 20 (3): 241–54.

List, John A., Paramita Singha, and Michael H. Taylor. 2006. "Using Choice Experiments to Value Nonmarket Goods and Services: Evidence from Field Experiments." *B. E. Journal of Economic Analysis & Policy* 6 (2): Article 2.

Little, Joseph, and Robert Berrens. 2004. "Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis." *Economics Bulletin* 3 (6): 1–13.

Louviere, Jordan J., and Daniel A. Hensher. 1982. "On the Design and Analysis of Simulated Choice or Allocation Experiments in Travel Choice Modelling." *Transportation Research Record* 890 (1): 11–17.

Louviere, Jordan J., Daniel A. Hensher, and Joffre D. Swait. 2000. *Stated Choice Methods: Analysis and Applications*. New York: Cambridge University Press.

Lusk, Jayson L., and Ted C. Schroeder. 2004. "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks." *American Journal of Agricultural Economics* 86 (2): 467–82.

Lusk, Jayson L., and Ted C. Schroeder. 2006. "Auction Bids and Shopping Choices." *B. E. Journal of Economic Analysis & Policy* 6 (1): Article 4.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by P. Zarembka, 139–45. New York: Academic Press.

Messer, Kent, Gregory L. Poe, Daniel Rondeau, William D. Schulze, and Christian A. Vossler. 2010. "Social Preferences and Voting: An Exploration Using a Novel Preference Revealing Mechanism." *Journal of Public Economics* 94 (3–4): 308–17.

Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherhead. 2005. "A Meta-Analysis of Hypothetical Bias in Contingent Valuation." *Environmental & Resource Economics* 30 (3): 313–25.

Norman, K., and Jordan Louviere. 1974. "Integration of Attributes in Public Bus Transportation: Two Modeling Approaches." *Journal of Applied Psychology* 59 (6): 753–58.

Ryan, Mandy, and Jenny Hughes. 1997. "Using Conjoint Analysis to Assess Women's Preferences for Miscarriage Management." *Health Economics* 6 (3): 261–73.

Salisbury, Linda C., and Fred M. Feinberg. 2010. "Alleviating the Constant Stochastic Variance Assumption in Decision Research: Theory, Measurement, and Experimental Test." *Marketing Science* 29 (1): 1–17.

Taylor, Laura O., Mark D. Morrison, and Kevin J. Boyle. 2010. "Exchange Rules and the Incentive Compatibility of Choice Experiments." *Environmental & Resource Economics* 47 (2): 197–220.

Thaler, Richard H., and Eric J. Johnson. 1990. "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice." *Management Science* 36 (6): 646–60.

Viscusi, W. Kip, Joel Huber, and Jason Bell. 2008. "Estimating Discount Rates for Environmental Quality from Utility-Based Choice Experiments." *Journal of Risk and Uncertainty* 37 (2–3): 199–220.

Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. 2012. "Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments: Dataset." *American Economic Journal: Microeconomics*. http://dx.doi.org/10.1257/mic.4.4.145.

Vossler, Christian A., and Mary F. Evans. 2009. "Bridging the Gap between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality." *Journal of Environmental Economics and Management* 58 (3): 338–45.