Incentive compatibility and respondent beliefs:
Consequentiality and game form

Daniel Rondeau
Department of Economics, University of Victoria, BC, V8W 2Y2

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

Abstract: Answers to valuation questions in stated preference surveys can be analyzed as economic decisions only if respondents believe their choice(s) are consequential (i.e., can affect their welfare). The empirical evidence we review indicates that the information content of surveys can significantly influence consequentiality beliefs, and controlling for beliefs can impact welfare estimates and improve validity. The review also uncovers several opportunities to improve upon current practices. First, most surveys do not deploy incentive compatible mechanisms that provide respondents with the correct incentives to truthfully reveal their preferences. Second, existing consequentiality measures do not fully capture consequentiality and are challenging to interpret. Finally, studies do not generally measure or control for other beliefs required to ensure that estimated value are consistent with economic theory. Hence, we provide a theoretical framework that links incentive compatibility conditions to a respondent's beliefs about these conditions. This motivates a theory-driven proposal to improve belief elicitation and foster greater validity of survey results.

1. Introduction

Much has been learned from the large number of stated preference (SP) survey studies motivated exclusively or primarily by behavioral observations and hypotheses. Understanding human behavior empirically or comparing results under different methodologies continues to produce useful results that can spur further refinements of the methods we deploy. Yet, the econometric methods deployed to analyze survey results and estimate demand curves are constructed directly from a theory of optimal choice (e.g., a Random Utility Maximization model or otherwise). Furthermore, when the purpose is to provide input into benefit-cost and regulatory analyses, measures of willingness to pay (WTP) or willingness to accept (WTA) rest on the broad axioms of consumer theory, and producing defensible Hicksian welfare measures requires a theoretically grounded survey instrument. Estimates of value can be promoted with greater confidence in policy and legal circles if a solid case can be made that they are consistent with economic theory.

The review of recent literature we present in this paper suggests that the use of theory-supported value elicitation procedures is the exception rather than the rule. Moreover, few studies attempt to measure and account for empirical deviations from theory. With these realities in mind, the main objective of this paper is to advance and substantiate the claim that using a theoretically valid elicitation mechanism and controlling for respondent beliefs about the mechanism can markedly improve on the current state of practice.

To achieve this objective, we discuss the theoretical foundations of SP surveys, provide empirical evidence that theory-driven designs and analysis improve welfare estimation, and provide recommendations intended to help move research forward. Throughout, our focus is on the valuation of public goods.

We distinguish between three distinct concepts related to the validity of SP value elicitation procedures: (1) consequentiality; (2) incentive compatibility; and (3) adherence of respondents to the game form presented in the survey. A reading of the valuation literature uncovers much confusion between consequentiality and incentive compatibility. This leads to potential misinterpretations of answers to consequentiality questions and leaves potentially fruitful research strategies untapped. Hence it is our assessment that consequentiality and its relationship to incentive compatibility and respondents' beliefs warrant further discussion.

Consequentiality is a foundational precept of economics. It is the simple requirement that an agent's utility can be affected by her decisions. Although the term consequentiality was coined in the context of valuation surveys, the idea is not specific to this sub-field of economics. For instance, Smith (1982) discusses the need for "saliency" as a necessary condition for laboratory experiments to be valid microeconomic systems.[1] Without consequentiality/saliency, economics does not provide a framework to analyze decisions. As Cason and Plot (2014, p. 1237) remark,

> *"The connection between the consequences of acts and the acts themselves is sufficiently obvious as part of the theory of choice in economics that it would seem to require no discussion. Together they form a "feasible set" of consequences from which the choice is selected".*

If responses to a valuation question are consequential, we can safely assume that choices can be analyzed using the tools of economics. The next challenge for the researcher is to deploy a value elicitation mechanism that is incentive compatible (IC), which means that a person's response to a valuation scenario, in theory, is a true reflection of her preferences for the good being valued. In the absence of incentive compatibility, respondents are presented with

---

[1] Following Smith (1982), Bergstrom and Stoll (1989) discuss the need for "saliency" as a condition for a contingent valuation experiment to be a valid microeconomic system.

incentives to misreport their preferences through strategic voting or otherwise, leaving researchers without a proper foundation to estimate valid welfare measures.

Consequentiality is of course necessary for a survey value elicitation to be IC, but it is not sufficient. Carson and Groves (2007) lay out a set of sufficiency conditions for SP surveys, while Vossler, Doyon, and Rondeau (2012) and Carson, Groves and List (2014) formalize them in expected utility and non-expected utility frameworks, respectively. Vossler, Doyon, and Rondeau (2012) also expand the set of conditions for a single valuation question to discrete choice experiments (DCEs) that deploy multiple binary choice questions.[2]

Recent parallel developments in experimental economics examine the possibility that even with consequential IC mechanisms, players may operate under a different set of rules than those presented to them. This has been referred to as failures of "game form recognition", or as "game form misconceptions" (Plott and Zeiler 2005; Chou et al. 2009; Cason and Plott 2014; and Bull et al. 2019). Cason and Plott (2014, p. 1237) define game form misconception as "a failure of the decision maker to recognize the proper connections between the acts available for choice and the consequences of choice...."

Their research investigates the long running puzzle posed by the results of financially incentivized experiments with the Becker-DeGroot-Marschak (BDM) mechanism (Becker, DeGroot, and Marschak 1964). In the BDM mechanism, a player with preferences over a good (or money) has a weakly dominant strategy to bid her true value. Yet, most people who are not trained to use the mechanism bid differently. In the selling experiments of Cason and Plott (2014) and Bull et al. (2019), fewer than 20% of subjects submit an asking price between [$1.95,

---

[2] Throughout the paper, we use the term "discrete choice experiment" to define a class of value elicitation formats that present respondents with multiple choice tasks, and for each task ask respondents to select the option (e.g., a particular policy or the status quo) they prefer.

$2.05] to sell a $2 bank note. One possible explanation is that some participants fail to recognize the mechanism's second price rule, bidding instead as if they are participating in a more familiar first price auction (against one random bidder).

Valuation surveys also require correct game form recognition by respondents. Carson and Groves' (2007, p. 184) "face-value" property alludes to this requirement: "respondents always correctly understand and answer the question being asked." This is typically a maintained assumption in valuation research. Yet, the empirical evidence based on measured consequentiality beliefs suggests that this assumption may not hold for a significant fraction of respondents. Where a respondent's beliefs about the implications of their answer do not match the IC conditions set out by the researcher, answers to valuation questions may not reflect truthful preferences over the choices presented in the survey.

Current approaches for identifying game form misconceptions in valuation surveys focus on measuring consequentiality. So-called "policy consequentiality" and "payment consequentiality" questions have received particular attention. The empirical evidence we document indicates that incorporating answers to those questions enhances the validity of welfare measures. Nevertheless, consequentiality questions asked in previous research focus on a narrow subset of conditions required for truthful preference revelation. They also appear to confound pure notions of consequentiality and other possible sources of game form misconceptions. These ambiguities can make it challenging to interpret answers and analyze results on solid theoretical grounds.

One important contribution of this paper is a discussion of the need for a new and expanded theory-driven approach for measuring and incorporating beliefs in data analyses. We develop a detailed proposal to this effect that refines the recent work of Vossler et al (2025). We

argue that such an approach can enhance the validity of stated preference methods and help better comprehend respondent choices.

2. Theoretical foundations

A better understanding of responses to valuation questions and an increased validity of valuation surveys can be achieved by approaching survey design and data analysis with a clear conceptual understanding of the requirements for the truthful revelation of preferences. In this section, we discuss three distinct conceptual issues from a theoretical perspective.

Consequentiality: A respondent's answer to a valuation question has the potential to affect her utility. In other words, the respondent's answer can influence the likelihood of different outcomes (e.g., status quo, project, or policy implementation), which, in turn, the respondent cares about. If a choice made by a respondent has no possible consequence on her utility, economic methods cannot predict what option will be chosen, nor interpret stated choices.

Incentive compatibility: a set of institutional rules that (assuming consequentiality) provide respondents with the theoretically correct incentives for the truthful revelation of preferences. Incentive compatibility is achieved when the valuation question(s) and supporting information provided are structured in such a way that respondents find it in their best interest to answer the valuation question truthfully.

Adherence to game form: The extent to which a respondent's beliefs about how her answer will ultimately affect her welfare are consistent with the actual mechanism presented to her in the survey.

There are two reasons for separating consequentiality from incentive compatibility and beliefs. First, consequentiality is conceptually distinct from the rules of the allocation

6

mechanism. Even if the respondent believes that her answers are consequential, her answers do not necessarily represent her true preferences over the options presented to her. Second, it is likely that past empirical work that controls for consequentiality beliefs captures elements of beliefs about *how* the survey leads to changes in her welfare. If beliefs about the rules of the game do not match the IC structure of the mechanism presented, game form misconceptions (i.e., failures to adhere to the game form presented in the survey) arise. Such failures can have important implications for the validity of welfare measures.

2.1 The valuation survey as a game

Consider a mechanism presented in a survey designed to elicit preferences for a public good. Using standard game theory, this "game" can be described in the following fashion:

- Players: There are $n = 1, ..., N$ respondents.
- Strategies: Each player is asked to choose a strategy (action, message, answer), $s \in S$, among the available choices. For a single binary choice (SBC) question, the possible strategies may be to indicate a vote for or against the proposal, e.g., $S=\{$Yes, No$\}$. In the case of a single multinomial question with three possible answers, the strategy set may be denoted $S=\{$Option 1, Option 2, Option 3). In a DCE involving four binary choice questions, each involving a choice between a policy (Yes) and the status quo (No), the relevant strategy space is the set of all possible combinations of four responses (e.g., $\{$Yes, Yes, Yes, No$\}$; $\{$No, Yes, No, No$\}$, etc.). With an open ended question, the strategy space might be any non-negative number.[3]

---

[3] We might consider that the respondent can also abstain from answering the question(s). This might be irrelevant to the researcher if such answers are discarded, but these responses may indicate important deviations from game form.

- Allocation rules: The survey describes a set of rules (R) for the allocation of the public good. These rules determine how the individual choices (strategies) of the N respondents translate into an outcome. One familiar rule, in the case of an SBC referendum, is to implement the outcome (a new policy or the status quo) based on a majority vote rule. Without loss of generality, we describe the possible outcomes as the provision of a good $x \in X$ at an individual cost $c \in C$. Hence, the rules of the mechanism define how a respondent's choice is combined with the choices of others to determine the outcome of the exercise: $R(s_1 \ldots s_N) \rightarrow (x, c)$.[4] If the mechanism admits uncertainty, this mapping can be thought of as aggregating choices and returning a vector of probabilities with which each possible outcome will be implemented: $Pr(x, c) = f(R(s_1 \ldots s_N))$.

- Payoffs: Players have a utility function defined over the (x, c) pairs.

- Player Types. Since players do not know other players' preferences, this is a game of incomplete information. Each player is of a "type" defined by their true preferences. Players must then hold beliefs about the distribution of types among the N players. Though rarely discussed in the valuation literature, beliefs about the distribution of other players' types can play a central role in establishing, for instance, that valuation questions with more than two choice options are not IC.

This description of the game is what we refer to as its "game form".

---

[4] Both the good and the cost may vary across respondents, and in this case x and c could be written as vectors that contain individual-specific elements. The analysis is not substantially affected by adding heterogeneity.

2.2 Consequentiality: a necessary pre-condition for economic analysis

Carson and Groves (2007, p.187) define the consequentiality of an advisory survey as

follows:

> "If a survey's results are seen by the agents as potentially influencing an
> agency's actions and the agent cares about the outcome of that action, the agent
> should treat the survey questions as an opportunity to influence those actions.
> In such a case, standard economic theory applies and the response to the
> question should be interpretable using mechanism design theory concerning
> incentive structures."

What Carson and Groves (2007) describe (but also Cason and Plot 2014; Smith 1982; and

Bergstrom and Stoll 1989) is the essential requirement that there exists a mapping between an

agent's choice(s) and her utility. This mapping can be factually true or in beliefs alone; for

present or future utility; with certain effects or probabilistically. We break down this mapping

into three essential components:

(M1) The agent sees an opportunity to influence the survey results;

(M2) The survey results influence the agency's actions; and

(M3) The agent cares about the outcomes of the agency's actions.

If this chain is broken at any point, a respondent's choice will have no impact on her welfare.[5]

Regardless of whether the allocation rules are explicitly stated or perceived,

consequentiality must hold for each respondent. Formally, consequentiality is established when

the following two conditions hold:

1) for each player n, choosing different strategies (s_n) leads to different outcomes. If the

rules of the mechanism aggregate individual choices to determine an outcome directly

---

[5] People are often offered payment to complete the survey. Respondents could also be deriving warm glow from their response. These incentives make the survey consequential from the respondent's perspective. However, if the respondent only cares about the act of participating with no additional consequences associated her responses to valuation questions, her choices cannot be meaningfully interpreted as expressions of her preferences.

such that $R(s_1 \dots s_N) \to (x, c)$, we can write this condition as $\Delta R(\bullet)/\Delta s_n \neq 0$. Alternatively, if the mechanism is probabilistic in that individual choices translate in changes in the probability of each outcome, $f(R(s_1 \dots s_N)) \to Pr(X, C)$, we write the condition as $\left(\Delta f(R(\bullet)) / \Delta s_n \neq 0 \right)$. Either way, this condition formalizes that M1 and M2 hold.

2) the agent has well-formed preferences over the outcomes in (X, C), i.e., she cares about possible outcomes (M3). She can have a ranking over discrete choices, be non-satiated in the goods of X, or she can simply be bound by a budget constraint. In this last case, paying (or receiving) different amounts (c) would modify the consumption bundle she could purchase (i.e., c enters her indirect utility function).

2.3 Incentive compatibility

Here, we focus on the theoretical properties of mechanisms that could be embedded in a valuation survey. When studying the theoretical properties of a mechanism, it is necessarily assumed that the choices are consequential (i.e., can affect utility). The objective is then to establish the properties of the game.

Of particular interest is the concept of incentive compatibility. Under an IC mechanism, being truthful is a weakly or strictly dominant strategy for all players. The reason why it is so important to pay attention to the incentive compatibility property of the mechanism is that without it we cannot interpret observed choices as reflecting true preferences, and the incentive structure can thus lead to systematic biases.

Most mechanisms that can be deployed to allocate public goods (as well as many popular mechanisms used in SP surveys) are not IC, at least without imposing potentially severe

restrictions on player preferences (Gibbard 1973; Satterthwaite 1975). For instance, voluntary contribution mechanisms where players are asked to give an amount of their choice towards the provision of a public good are not IC. From a theoretical perspective, these mechanisms generally lead to free-riding because: marginal benefits are obtained from the contributions of others (removing individual incentives to fund those units); and the marginal values to an individual associated with the provision of at least some units of the public good are smaller than the cost of provision despite the social marginal value being greater.

Mechanisms where the preferences of more than two players over three or more options must be aggregated are also not IC (see Arrow's impossibility theorem; Gibbard-Satterthwaite theorem). To see this, imagine a player facing three options. Suppose that she prefers Option 1 to Option 2, and that Option 3 is strictly worse than both. It is easy to construct examples where a player's optimal choice is to vote for Option 2 even if it is not her first choice. For example, if Option 1 is low in the preferences of other players and our voter can influence whether Options 2 or 3 is selected, it can be optimal to indicate a preference for Option 2 to avoid Option 3 (her worst outcome). Choosing her favorite outcome is not a dominant strategy and her vote does not reflect her true favorite option.

The simplest IC mechanism for collective decisions is an SBC referendum whereby participants must choose between two alternatives. Carson and Groves (2007), Vossler, Doyon, and Rondeau (2012), and Carson, Groves and List (2014) discuss the sufficiency conditions for incentive compatibility in detail: (i) A vote between only two possible outcomes; (ii) One of the two outcomes will be implemented; and (iii) A vote for one option increases the likelihood of that option being implemented (and reduces the likelihood of the other). Consequentiality is of course also necessary.

A binary or two-option DCE, with each question being a vote between a potential policy and the status quo, can also be IC (Vossler, Doyon and Rondeau 2012). When each vote is viewed in isolation, the conditions described above for the SBC mechanism must hold. In addition, however, the allocation rule must maintain independence between choice questions. A vote for one policy in one question cannot affect the likelihood that some other policy not under consideration in that question will be implemented. Furthermore, at most one of the various possible policies described in the DCE can be implemented.

It is easy to devise allocation rules that violate these added conditions. For example, suppose the allocation rule stipulates that of all policies considered in the different questions of the DCE, the one that received the most votes in favor will be implemented. Then, voting for one policy in one question effectively decreases the probability that all other policies in the survey will be implemented. As in the case of a single choice between three or more options, it can then be optimal for a respondent to vote strategically and not reveal her true preferences in at least some of the choice questions.

2.4 Beliefs at the intersection of consequentiality and game form

Respondents to advisory surveys can approach valuation scenarios with any prior beliefs. They might think that they can broadly influence the quantity, quality, location, or any other characteristic of the good presented to them in the survey. They might also believe that their answers could influence the amount they will have to pay for future policies. They might believe that their "opinion" influences policymaking beyond the immediate proposal in a survey.

Even when an IC mechanism is deployed, respondent beliefs can get in the way of the truthful revelation of preferences. Imagine a respondent facing an IC advisory referendum on

whether to plant trees along a specific road on the way out of town. Further imagine that the respondent believes that her answer will not only help decide whether those trees are planted, but that it will also affect the likelihood that trees will be planted on her own street or elsewhere in the city. Clearly, this respondent believes that her vote is consequential, and the mechanism presented to her was IC. Yet, she thinks that her answer can affect decisions about planting trees that are not part of the survey scenario. As a result, her answer cannot be interpreted as truthfully revealing her preferences over the options presented to her.[6] This respondent suffers from a game form recognition problem.

To solidify concepts, we restate the sufficiency conditions required to ensure valid SP surveys in terms of respondent beliefs. For accurate game form recognition in a SBC referendum, a respondent must hold the following beliefs:

(B1) The respondent believes that her vote can only influence the likelihood of the two possible outcomes presented in the survey;

(B2) The respondent believes that voting for one option increases the likelihood of that option being implemented and decreases the likelihood of the other option; and

(B3) The respondent believes that the voting outcome can affect her welfare.

Deviations from these beliefs are game form misconceptions. Where there are game form misconceptions, answers cannot safely be interpreted as revealing the respondent's preferences over the options presented in the survey.

---

[6] Theory does not rule out the possibility that respondents with game form misconceptions are nevertheless truthfully revealing preferences. As an example, a respondent may think that her vote not only influences the policy asked about, but other policy decisions as well. If she believes that her influence over other policies is of second-order importance relative to the options in the survey, her answers may still truthfully reveal her preferences. However, there is no guarantee that this is indeed the case.

In Figure 1, we superimpose the three components of influence, M1 to M3, as described in Section 2.1, on beliefs B1 to B3. M1 to M3 are general links mapping a vote to utility without restrictions imposed by incentive compatibility conditions. M1 captures whether an individual believes her vote affects survey results, M2 is whether she believes the survey influences policy decisions, and M3 is whether she cares about the any outcome. Beliefs adhere to an IC game form if, along that chain, only two options are in play (B1), voting for an option increases the chance that this option is implemented (B2), and the respondent's belief about how her utility might be affected is restricted to those two options (B3).

<INSERT FIGURE 1 HERE>

This characterization of beliefs is useful in interpreting the results of studies that have used survey questions to measure beliefs. Past elicitations of consequentiality beliefs have been heavily influenced by Herriges et al. (2010). They introduced separate notions of "policy consequentiality" and "payment consequentiality": "First, the respondent must believe that the results of the survey might influence an outcome they care about, a condition we refer to as policy consequentiality. Second, the respondent must perceive that there is some probability that they will have to pay, or payment consequentiality" (p. 67).

There are now many examples of policy consequentiality questions in the literature (see Zawojska et al. 2021), and a small number of payment consequentiality questions. One popular policy consequentiality question (see Zawojska et al. 2021) asks "How likely do you think it is that the results of this survey will influence [decisions or general changes in policy]?" A similar question asks respondents to agree or disagree with a statement such as "I believe that the results of this survey could affect decisions about [topic of the survey]." Both questions target M2. It is

14

notable, however, that they ignore whether the individual believes that her own vote affects survey results. Hence, these questions technically fail to establish consequentiality.

Other policy consequentiality questions do ask the respondent about the impact of her own vote. One recurring question asks: "To what extent do you believe that your choice will be taken into account by [the agency]?" This question arguably combines the individual's influence on the survey results (M1) with M2. Zawojska et al. (2021) report that such consequentiality questions that cover both M1 and M2 result in different welfare estimates than those that only correspond to M2.

Zawojska et al. (2021) and Borger et al. (2021) suggest that policy consequentiality questions measure the influence of a vote (M1 plus M2) or survey (M2 only) on the delivery of the good (x) only. This is not a consensus, however. Good survey scripts spend a lot of time trying to convince participants that they will have to pay if the good described in the policy is provided. If a respondent is then asked to what extent she believes that her vote will be taken into consideration by policy makers, is she really thinking exclusively about the delivery of x or about the bundle (x, c) that she just voted on? A respondent who thinks that she or the survey can only influence the cost component but who interprets the policy consequentiality question to mean influence over the bundle (x, c) should answer that her vote was consequential, just as someone who believes that she has influence over both components. Including a separate question on "payment consequentiality" does not help resolve this ambiguity.

There has been considerable variation in the wording of consequentiality questions, and interpreting answers to them can be challenging. Moreover, past questions provide an incomplete picture of respondent beliefs. In particular, they ignore beliefs over the conditions for incentive compatibility of the mechanism (B1-B3). We believe that expanding the measurement of beliefs

beyond existing policy and payment consequentiality questions can yield deeper insights into respondent behavior and survey validity.

Nevertheless, past practices provide an opportunity to demonstrate how a theory-driven approach can help derive testable hypotheses, identify inconsistent behavior, and ultimately advance research into the validity of SP methods. As an example, imagine a respondent to an SBC survey who is asked to vote for one of two options $(x_1, c_1) \neq (x_2, c_2)$. Assume that the respondent's beliefs adhere to B1-B3.

Suppose that the researcher asks two belief questions: 1) whether she believes that her vote can influence the provision of the good (i.e., $\Delta x / \Delta s_n \neq 0$); and 2); whether she believes that her vote can influence the cost that she would have to pay (i.e., $\Delta c / \Delta s_n \neq 0$).[7] Answers are on a Likert scale. With B1 to B3 holding, theory predicts that the respondent agrees that her answer can influence both attributes. She should also provide identical responses on the Likert-scale. To see this, note that by B1, she believes that the only possible outcomes are (x_1, c_1) or (x_2, c_2). Since the good and cost in each bundle are inseparable in the survey mechanism, the respondent cannot possibly believe that her vote can have different degrees of influence on the provision of x and on its cost c. If her beliefs adhere to game form, her vote has equal influence on x and c, and her beliefs should be consistent with this feature of the mechanism. If a respondent answers the two questions differently, she must think that it is possible to get it possible to obtain x_i at a different cost than c_i, for example. This is a failure of adhering to the game form. Hence, past consequentiality questions are not just about ensuring

---

[7] We intentionally do not define $\Delta x / \Delta s_n \neq 0$ as "policy consequentiality", nor do we define "payment consequentiality" as $\Delta c / \Delta s_n \neq 0$. We believe that the existing literature using these concepts and the questions typically deployed are too vague to make these distinctions. Adopting these definitions here would likely exacerbate confusion and potentially lead to misinterpretations of past results. For the same reason, we also refrain from attaching new labels to $\Delta x / \Delta s_n \neq 0$ and $\Delta c / \Delta s_n \neq 0$. Our goal here is to illustrate how a clearly formulated theory of game form recognition in an SBC survey can be used to generate testable hypotheses and interpret results.

that consequentiality hold. They can pick up deviations from game form recognition as well, with the implication that truthful revelation of preferences between (x_1, c_1) and (x_2, c_2) cannot be guaranteed even when an IC mechanism is deemed consequential.

Violations of game form recognition are perhaps the single most difficult hurdle for DCEs. Even where each question is a binary choice to accept or reject a profile of attributes at a cost, incentive compatibility requires not only that respondents take each question in the game form presented to them, but also that they treat each question entirely independently of each other. It would not be surprising if some respondents believe that their sequence of answers, taken as a whole, is what influences the delivery and cost of a public good. In such a case, respondents are holding beliefs that differ from the game form and this casts doubt on the validity of each of their individual choices. We discuss ways to improve adherence to game form and detect game form violations below.

3. Does theory matter in practice?

In this section we briefly review the large empirical literature that compares different value elicitation procedures. We also summarize emerging research concerned with inducing and incorporating belief measures into welfare analysis. The evidence suggests that: (1) using different mechanism designs often leads to important disparities in welfare estimates; (2) survey design can alter respondent beliefs; and (3) welfare estimates can vary considerably across respondents with differing beliefs. We argue that many patterns in the data can be explained with economic theory but that current practices provide an incomplete picture of respondents' beliefs. As a result, theory-driven survey designs could increase both the theoretical and empirical validity of valuation methods.

An important caveat to this discussion is that interpreting empirical tests requires that one maintain what are often untestable assumptions. Take for example a test of whether two mechanisms yield equal WTP estimates. As we will see, most studies do not measure or control for consequentiality beliefs. It is therefore possible that similarities or differences between mechanisms are due to a lack of consequentiality, and without consequentiality, there is no theoretical basis to predict behavior. As another example, some studies control for respondent beliefs by dropping those who indicate that the survey is inconsequential. If the restricted sample is not representative of the population, issues of sample selection may confound the results of empirical tests.[8]

3.1 Mechanism design

There is a long history of testing whether different elicitation mechanisms give rise to similar welfare estimates. As these tests do not incorporate a criterion measure (for example, a revealed preference estimate based on an IC mechanism), statistical differences do not necessarily indicate that one approach is superior, or that either approach approximates actual demand. Instead, this literature often relies on comparisons with an SBC format because of its desirable incentive properties.

During the 1980s and 1990s, most tests of elicitation effects compared open-ended format or related formats (e.g., payment card) to an SBC format. Carson and Groves (2007) argue that open-ended and related formats are likely to engender two possible beliefs regarding how responses determine outcomes (i.e., beliefs about M1 and M2). First, a respondent might

---

[8] It is also relevant to note that empirical tests require the use of statistical procedures that embed additional assumptions. As the results from individual studies may be artifacts of invalid econometric procedures, or insufficient sample sizes, where possible we base conclusions on multiple studies.

believe that she can influence the amount paid for the provision of the good. This should lead to a downward bias in WTP. Second the respondent might think that her response can influence the provision of the good. In this case, indicating a WTP that exceeds her valuation is optimal if she expects the true cost of the good to be less than her true WTP. In contrast, if she expects the cost to exceed her true WTP, she has an incentive to under-report her demand.

Champ and Bishop (2006) summarize the results from studies that compare an SBC elicitation with an open-ended question, a payment card, or both. Of the 20 comparisons from published SP studies, the SBC format yields a significantly higher WTP in 14 cases and no detectable difference in the remaining six comparisons. More recent studies find similar results (e.g., Vossler and Holladay 2018). Assuming that an SBC mechanism is a reasonable benchmark of actual demand, the stylized fact from those comparisons is that open-ended and related formats underestimate demand. While the direction of the theoretical bias associated with open-ended and payment card formats is ambiguous, a downward bias is sensible since the potential strategic opportunity to influence the price is likely more salient with these mechanisms.

It is understood that the double-bounded binary choice format is predicted to give rise to biases.[9] Part of the argument is theoretical, and can be linked to game form misconceptions. Relative to the first valuation question, the second one offers the same good at a different price. Under these circumstances, respondents should be expected to see an opportunity to influence the price of the good. The stylized fact is that WTP estimates based on the first and second question are imperfectly correlated, and WTP estimates tend to be lower than for an SBC survey. Studies by Vossler and Holladay (2018) and Vossler and Zawojska (2020) suggest that

---

[9] This format involves a sequence of two binary choice questions that are identical except for the cost of the proposal. The cost presented in the second question is higher (lower) if the respondent answered "yes" ("no") to the first question.

differences between an SBC, and the open-ended, payment card, or double-bounded dichotomous choice elicitation formats can be mitigated by improved mechanism design, specifically by providing more precise information on the allocation rule.

Contemporary studies in environmental economics predominantly employ DCEs. Lloyd-Smith, Zawojska, and Adamowicz (2020) survey the literature that compares DCEs with various alternative elicitation formats. From the papers reviewed, we identify eight comparisons between a multinomial DCE and an SBC format. SBC-based welfare measures are higher in two cases, lower in three studies, and equal in the remaining three applications.[10] Two studies compare a two-option DCE to an SBC question, and both support a conclusion of equal WTP.

A related result is that varying the number of options to choose from in each question of a DCE often leads to differences in welfare estimates (Weng et al. 2021). It is important to note that theory does not provide a prediction on the direction of the bias. When choice sets include more than two options, whether the respondent votes strategically in any given question depends on the respondent's preferences and her beliefs about the likelihood that her favorite option will not be selected regardless of how she votes.

Some useful insight can be gleaned from a recent study by Vossler and Zawojska (2023). They compare an SBC, a two-option DCE, and a three-option DCE using real payment experiments that employed allocation rules that promote truthful responses. Results show that the SBC and two-option DCE return equal WTP estimates, which corroborates the limited SP evidence. The three-option DCE produced statistically different estimates, which is also

---

[10] These figures are based on our reading of these papers, which differs from Lloyd-Smith, Zawojska, and Adamowicz (2020) with respect to Adamowicz et al. (1998). For that study, the finding of equal WTP estimates pivots on the assumption that the intercept of the DCE model is ignored when calculating (total) WTP, which is not a commonly accepted practice. We also note that some of the papers that conclude that WTP estimates are similar overall nevertheless report statistical differences for a subset of comparisons.

generally consistent with the SP evidence. The authors attribute differences to two phenomena. Relative to the two-option DCE, a higher fraction of three-option DCE respondents always select the status quo, which the authors attribute to cognitive burden. Some differences are driven by framing effects: The authors find that the WTP for a good tends to increase (decrease) when it is paired with a relatively smaller (larger) good. Such behavioral pattens, along with broader incentive compatibility concerns, caution against the use of multinomial DCEs.

Several studies compare coercive and voluntary payment vehicles. Stithou and Scarpa (2012) and Hassan, Olsen and Thorsen (2018) describe this literature in some detail and provide their own evidence. These articles document nine studies where voluntary payments lower mean WTP estimates, and four studies with no significant difference.[11] The findings are consistent with the free-riding incentives of a non-IC voluntary contributions mechanism.

3.2 Respondent beliefs

The theory of game form recognition presses the point that respondent beliefs can affect stated preferences. Researchers have begun to explore the extent to which survey design can promote desirable beliefs, and the degree to which beliefs correlate with welfare measures. In their literature review on policy consequentiality, Zawojska et al. (2021) identify 37 studies that elicit values for public goods using mechanisms that involve a coercive payment vehicle. Twenty-eight of those studies find a statistically significant effect of beliefs on stated preferences, three do not find significant effects, and six do not provide relevant test results. In

---

[11] Hassan, Olsen and Thorsen (2018) list two studies that they claim find higher WTP estimates with voluntary payments. One citation reflects an oversight as that study did not compare payment vehicles. The other study (which is Stithou and Scarpa (2018)) finds that WTP is statistically equal across payment vehicles, and only when analyzing the subset of respondents with strictly positive WTP do they uncover WTP differences.

studies that find a statistical difference, almost all report that WTP increases with beliefs of greater policy consequentiality.[12]

Far fewer studies analyze payment consequentiality questions. Zawojska, Bartczak, and Czajkowski (2019) and Kabaya (2021) find that WTP decreases with payment consequentiality. If those who view the elicitation as payment inconsequential believe they will not have to pay anything for a new policy, but nevertheless perceive that they can influence provision, this result is consistent with theoretical expectations. In this case, the survey remains consequential. But respondents are deviating from the scenario, which is a failure of game form.

Börger et al. (2021) find the opposite effect in a study where they asked explicitly whether the respondent believed that she would have to pay the exact cost stated in the survey. One possible explanation for why WTP increases with payment consequentiality is that those reporting low consequentiality believe that they would have to pay a higher cost than the stated amount (once again, this should be understood as a failure of game form recognition – not of consequentiality per se). Vossler and Holladay (2018) and Welling, Zawojska, and Sagebiel (2022) control for whether respondents jointly hold policy and payment consequentiality beliefs and they find the overall effect of consequentiality on WTP to be positive.

In our opinion, more convincing evidence comes from studies that have used exogenous sources of variation to alter consequentiality beliefs and validate the use of stated belief measures. A prominent example is Herriges et al. (2010), who presented some respondents with a magazine article and a letter from a state official documenting how a previous survey on a related topic influenced policy. This predictably increased stated beliefs about the policy consequences of the new survey, and resulted in higher WTP estimates.

---

[12] In some cases, authors do not test for WTP differences, but instead find that those with consequentiality beliefs are more likely to select a policy option rather than the status quo, which is indicative of higher WTP.

Vossler et al. (2023) find that consequentiality beliefs for both WTP and WTA decreases with a change in the political climate (i.e., a likely transition of the political party in power). This change made it less plausible that the policy described in the survey would be implemented. Overall, higher consequentiality increased WTP and decreased WTA. The WTA/WTP ratio was between 4.8 and 6.5 for respondents who thought that the survey was unlikely to be consequential. The ratio fell below 2 for respondents who instead believed that the survey is somewhat or very likely to be consequential.

One way of promoting consistent beliefs is to develop questionnaires that actively promote consequentiality and game form recognition. Several studies have tested the effects of using short information scripts (sometimes referred to as "consequentiality scripts") to heighten consequentiality beliefs shortly before the valuation scenario(s). Bulte et al. (2005) find that alerting respondents that study results "will be made available to policymakers, and could serve as a guide for future decisions" significantly decreases WTP. In contrast, Czajkowski et al. (2017), Drichoutis et al. (2017), Oehlmann and Meyerhoff (2017), and Kabaya (2021) observe either small or insignificant WTP changes resulting from their scripts.[13] Kabaya (2021) finds that a payment script increases payment consequentiality beliefs and decreases WTP.

Two studies examine the effects of providing information on a potential allocation rule that is consistent with incentive compatibility. Vossler et al. (2025) conduct a split-sample test of an "independence script" designed to encourage respondents to treat each referendum in a binary DCE independently. Among those receiving the script, WTP increases with the spatial scope of the good. In contrast, those who did not receive the script show evidence of a negative scope

---

[13] For at least three of these four studies, the "script" is provided in addition to either an acknowledgement of government funding and/or a statement at the beginning of the survey suggesting that results will be shared with policymakers. As such, these small or insignificant effects should not be interpreted as meaning that promoting consequentiality is unimportant; instead, multiple pleas for consequentiality may not be necessary in some settings.

effect, a result that is inconsistent with economic theory. Those receiving the script are also much more likely to state that they treated the valuation scenarios of the DCE independently from one another, and are more likely to hold consequentiality beliefs. In a study with a single open-ended WTP question, Vossler and Holladay (2018) find that respondents who were randomly assigned a script describing an IC allocation rule are more likely to believe in both policy and payment consequentiality. Furthermore, elicited values among those receiving the script are statistically equal to those assigned to an SBC version of the survey.

Finally, we have identified two studies that examine whether controlling for consequentiality beliefs has any bearing on external validity tests. Vossler and Watson (2013) compare the results of an SP survey using an SBC mechanism with that of a naturally-occurring public referendum. Most respondents were unaware of the upcoming referendum. The proportion of "yes" votes in the survey is lower than the proportion cast in the actual referendum. However, the proportions are not statistically different when the sample is restricted to those who hold policy consequentiality beliefs.[14] The authors argue that WTP models based only on those with consequentiality beliefs provide greater evidence of construct validity. Vossler, Doyon, and Rondeau (2012) compare SP survey responses with those from parallel revealed preference treatments in which participants had the opportunity to fund actual tree-planting programs. Revealed and stated preferences were elicited using a binary DCE. Overall, WTP estimated from the SP survey exceeds the field experiment results by approximately 30%. The difference is eliminated for respondents who hold more than "weak" policy consequentiality beliefs.

While the above results have taught us much about the behavior of respondents, it is critical to acknowledge that the consequentiality belief questions deployed in past research have

---

[14] The authors elicited beliefs using a five-point Likert scale and coded all but those respondents selecting the lowest option ("not taken into account") as holding policy consequentiality beliefs.

almost exclusively targeted the M1 and M2 nodes of the consequentiality chain described in Figure 1. As such, they focus on an incomplete definition of basic consequentiality. Perhaps more importantly, they do not control for beliefs about the incentive compatibility conditions of the mechanism deployed. Without a more complete inventory of beliefs, it is always possible to take any empirical observation and construct a particular game form misconception to explain it.

4. Closing the gap between theory and practice

4.1 Current state of practice

It is clear from the recent guidelines of Johnston et al. (2017) and the empirical studies reviewed above that many SP researchers accept that the incentive properties of a valuation survey matter, and that they depend critically on various aspects of survey design and implementation. To gauge the extent to which the theoretical work on incentive properties has translated into practice, we collected information from 128 SP studies published in five leading environmental economics journals over the period 2018-22.[15,16]

Table 1 identifies, in rank order, the elicitation format used in these studies. Nearly half the studies (48 percent) employ three-option DCEs. The second most popular elicitation format is the SBC at 14 percent, followed by the two-option DCE (13 percent), and DCEs with four or more choice options (11 percent). Open-ended and payment card formats are each used in less than 10 percent of studies.

---

[15] The journals are *Environmental & Resource Economics*, *Journal of Environmental Economics and Management*, *Journal of the Association of Environmental and Resource Economics*, *Land Economics*, and *Resource and Energy Economics.*
[16] Study-specific information and references are provided in the Appendix.

Table 1. Utilization of elicitation formats in contemporary stated preference research

Elicitation format	% of studies
Discrete choice experiment, choice sets with three options	48
Single binary choice	14
Discrete choice experiment, choice sets with two options	13
Discrete choice experiment, choice sets with four or more options	11
Payment card	7
Other, attribute-based format[*]	6
Other, non-attribute-based format[**]	5
Open ended	4

Notes: Percentages are based on a set of *n*=128 studies published in five leading environmental economics field journals over the period 2018 to 2022. Percentages add up to over 100% as a small set of studies incorporate multiple methods. [*] Includes contingent ranking, best-worst scaling, and related repeated-question formats. [**] Includes multiple-bounded discrete choice, double-bounded binary choice, and other non-attribute-based methods.

The prevalence of DCEs is not entirely surprising since, compared to single-question surveys, they provide the opportunity to value multiple goods in a cost-effective manner. Perhaps more surprising is that most studies rely on elicitation formats, such as multinomial DCEs, that are not IC. Vossler et al. (2023) report that for a set of WTA-WTP comparison studies, most of which were conducted prior to 2010, fewer than 10 percent used an SBC format, and none used a DCE. This evidence, along with the data from our literature review, suggests a positive trend towards the use of theory-supported elicitation formats, but progress has been slow.

Table 2 shows that ten percent of studies ask a policy consequentiality question, and just four percent ask both payment and policy consequentiality questions. Most of those studies use the answers to control for beliefs in the data analysis. A slightly larger fraction of articles (14 percent) report removing "protest" respondents from their analyses. Many studies who drop "protest" respondents provide incomplete details on how those respondents are identified. If protest respondents were in fact those identified as deviating from game form based on follow-up

questions, at most 23 percent of studies can be considered as having measured and controlled for respondent beliefs in some way.

Table 2. Study attributes related to incentives and beliefs

Study attribute	% of studies
Article mentions survey incentive properties	47
Article cites Carson and Groves (2007)	30
Consequentiality beliefs considered in data analysis	9
Removed "protest" respondents in data analysis	14
Asked policy consequentiality question[*]	10
Asked policy and payment consequentiality questions[*]	4

Notes: Percentages are based on a set of 128 articles published in five leading environmental economics field journals over the period 2018 to 2022. [*] Categorizations are based on what is reported in the articles, and denoted entries should therefore be interpreted as lower bounds.

To assess whether authors may at least be considering survey incentive properties, we also recorded: (1) whether the published article includes any mention or discussion of survey incentive properties; and (2) whether the published article cites the seminal article by Carson and Groves (2007). One or both of those signals are present in about half (49 percent) of the studies. Of course, the presence of either or both signals does not guarantee that close attention was paid to the incentive properties of the survey during the design stage.

4.2 Selecting and implementing the elicitation mechanism

4.2.1 The elicitation format

Among the small set of mechanisms for which incentive compatibility conditions have been established, the SBC mechanism is IC under the weakest assumptions. When framed as a referendum, the mechanism is simple to explain and familiar to most. Yet, when researchers

wish to obtain values for multiple goods within one study, deploying an SBC mechanism requires large samples and can be prohibitively costly. A reasoned compromise is to use a two-option DCE, which is IC when additional assumptions are made.

Beyond the theoretical argument, there are important empirical reasons for constraining choice sets to two options. Adding alternatives increases cognitive burden (i.e., complexity), which can increase the incidence of undesirable behavior such as: (1) reverting to general decision heuristics that give rise to deviations from utility-maximizing choices; (2) opting out of the valuation exercise (e.g., non-response or persistently selecting the status quo); and (3) framing effects that make the valuation for a good conditional on the included alternatives. Increasing the number of alternatives also amplifies experimental design challenges by exponentially increasing the full factorial of possible choice sets. The counterargument is that adding alternatives improves the precision of welfare estimates. This claim may be true from a pure statistical power perspective, but only if one is willing to assume that the loss of incentive compatibility and the added complexity do not bias responses away from true preferences. The theoretical and empirical evidence strongly rejects this assumption.

4.2.2. The allocation rule

The allocation rule that aggregates responses into a policy decision is central to the theoretical property of a mechanism and requires parallel respondent beliefs. A precise allocation rule is typically absent from valuation scenarios because researchers can rarely make a guarantee on whether or how survey results will be used. It might be argued that it is natural for respondents to believe that an SBC vote in favor of a policy increases the chance it is implemented (B2). However, non-SBC surveys leave much open to the respondents'

imagination. As discussed in Section 2, it is not difficult to imagine reasonable ways in which survey results could be used that gives rise to a loss of incentive compatibility.

To help mitigate possible beliefs that threaten the incentive compatibility of a mechanism, we recommend that valuation scenarios include information on a credible way that authorities *may* interpret survey responses - one that is consistent with the allocation rule of an IC mechanism such as a majority rule referendum. Two recent papers illustrate this idea: Vossler and Holladay (2018) and Vossler et al. (2025). Both papers provide information scripts that describe allocation rules that would ultimately translate responses into a "yes" or "no" vote on a single good at a specific cost (i.e., as in an SBC mechanism). Both studies provide evidence that these information scripts improve validity.

An important part of the allocation rule in a valuation survey is the method by which a payment might be collected (or compensation paid). Incentive compatibility relies on the coercive imposition of the selected outcome. This calls for a payment vehicle such as a mandatory fee in a utility bill or a tax, but they are by no means perfect in an advisory survey. Some individuals may realize that they are not subject to the payment scheme (income below the taxable threshold) and can avoid payment. Such respondents almost certainly violate game form beliefs B1-B3.

Coercive payment vehicles are particularly problematic in developing countries. Many SP studies in these settings omit any mention of a payment vehicle (Whittington and Pagiola 2012). Hassan, Olsen and Thorsen (2018) propose an alternative approach, which is to fund a public good through a reduction in government subsidies for ordinary consumer goods. They find significant WTP differences across subsidy, donation, and income tax payment vehicles. Care must therefore be taken to assess both the credibility and efficacy of the payment vehicle

before a survey is administered. Controlling whether respondents are subject to the payment scheme could help detect beliefs that do not adhere to game form.

4.2.3. Promoting consequentiality

The theoretical and empirical importance of ensuring that respondents hold beliefs that are consistent with consequentiality strongly suggest that researchers would do well to emphasize the consequentiality of SP surveys in deliberate ways. In his influential book on survey methods, Dillman (1978) indicates that a survey should stress its usefulness for society, and an important strategy for doing so is to promise to provide the results to government officials (and then actually carry this out). From our experience, we find that public organizations operating in areas related to the surveys we conduct are typically interested in our study when we contact them. Published welfare estimates, including those not funded by the government, tend to be used by various government agencies to estimate benefits as part of regulatory analyses (see Petrolia et al. 2021). Thus, it can still be valid to stress the policy relevance of surveys that are not expressly designed to inform public decision making. Following Herriges et al. (2010), a promising approach may be to discuss how past surveys have been used as inputs to policy analysis. This could be supplemented with a specific example or two, and/or links or references to relevant government websites or news articles.

On the flip side of the coin, researchers should refrain from framing the valuation exercise as "hypothetical" (e.g., as in standard "cheap talk" scripts that discuss the problem of hypothetical bias). Purely hypothetical scripts can only weaken the link between respondent choices and outcomes. As Mitchell and Carson (1989, p. 238) stress:

> "Questionnaires that unnecessarily call attention to the hypothetical character of the
> valuation exercise, either directly by using wording such as "pretend" or "assume a

hypothetical situation" when eliciting the WTP amount, or indirectly by the use of a highly abstract or implausible payment vehicle, run the risk of encouraging overbidding by respondents who want the good to be provided. This situation would be exacerbated by wording that emphasized the importance of the respondents' answers to policymakers."

One way to promote consequentiality of the payment mechanism is to remind the respondent of her budget constraint. This has long been advocated as good practice (e.g., Mitchell and Carson 1989). An early example comes from Loomis, Gonzalez-Caban, and Gregory (1994, p. 502):

> "[…]. Money you spend on the fire program would reduce the amount of money your household will have available to spend on the other environmental problems mentioned as well as on the everyday products you buy."

Bateman and Langford (1997) ask respondents about their budget constraint, compelling them to think about how much money they spend on related goods. Scripts of this sort effectively remind the respondent that the policy has a cost that enters her indirect utility function and that it could alter their future consumption.

4.3 Toward a comprehensive approach to belief measurement

Recent studies rarely assess whether respondent beliefs adhere to incentive compatibility conditions beyond establishing generic M1 and M2 consequentiality. Figures 2 and 3 provide potential belief elicitation instruments for an SBC and a two-option DCE, respectfully. In the latter, choice sets ask respondents to select between a proposed policy and the status quo. These instruments represent alterations of the "belief inventory" approach of Vossler et al. (2025). The DCE belief instrument could readily be used for multinomial DCEs, but would not identify the well-known issue of voting strategically within a choice set.

<INSERT FIGURE 2 HERE>
<INSERT FIGURE 3 HERE>

The belief instruments each present six statements within a response matrix and ask respondents to indicate their level agreement with each one by selecting from "disagree", "neutral", and "agree". The first statement is like several of the policy consequentiality questions that have appeared in the literature, and gauges whether respondents thought that their responses and/or the survey results would influence outcomes (M1 and M2). This could be interpreted by the respondent as meaning influence over good provision (e.g., environmental protection), influence over payment (e.g., increase in taxes), or both. The second statement provides a way to measure whether changes in environmental quality enter the respondent's utility function (M3). We do not explicitly mention policy cost, and instead safely assume that a change in disposable income unambiguously impacts well-being. Respondents who disagree with either or both statements signal that the basic necessary condition of consequentiality is unlikely to hold.

The remaining statements are associated with the allocation rules and thus the game form itself. The third statement is motivated by B2, and measures whether the respondent perceives that her vote on a policy could influence whether that same policy is implemented. The fourth statement asks whether the respondent instead perceived that her vote would be used to determine whether to implement a different policy. This would signal a violation of B1. The last two statements measure whether the perceived provision and cost adhere to the information presented to the respondent in the valuation scenario(s). These are supporting items that aid in the interpretation of responses to other belief statements. They aim to identify whether in the "policy" as viewed by the respondent matches the "policy" described in the survey. Collectively, answers to the last four statements imply whether B3 holds.

Agreement with all statements provides evidence that the person holds beliefs that are consistent with the incentive compatibility conditions. Other patterns of answers suggest a lack

of consequentiality or game form violations. For respondents indicating that consequentiality holds but who exhibit departures from the game form, it may be informative to ask follow-up questions. For instance, where deviations from the stated improvements or costs are detected, such questions would aim to provide indications of whether the perceived level (of a good or its cost) is higher or lower than the stated amount.

While we hope that the belief measurement instruments prove useful to researchers, there are surely ways to improve upon them prior to use. First, the belief statements can further be polished by providing specific references to the information provided in the survey. For example, "environmental improvements" should be replaced with "water quality improvements", "land conservation outcomes", etc. At least for an SBC survey, the last statement could include the cost amount faced by the respondent. For an SBC survey, the third statement could be revised to signify a directional effect consistent with whether the respondent voted "yes" or "no". Second, while the intent was to promulgate a user-friendly measurement instrument, researchers may wish to rework the belief statements into a set of standalone questions or include additional response options (e.g., "strongly agree", "strongly disagree"). Third, while we suspect that additional statements may burden respondents, doing so may nevertheless prove fruitful to the understanding of motives and beliefs. For instance, Vossler et al. (2025) include a measure that gauges whether people voted in the survey as they would have in a public election, i.e., a revealed preference benchmark. This is potentially useful but omitted here as it does not directly relate to the theory.[17]

---

[17] For most applications involving DCEs, given the nature of the within-survey policy variation in attributes (e.g., less/more aggressive environmental protection) it is not technically feasible to deploy multiple policies at the same time. However, for instance in the case where policies described would impact distinct geographic areas, the possibility that multiple policies could be implemented arises. In such a case, it may be worthwhile to include a statement/question that provides a measure of this belief.

In cases where most respondents seemingly adhere to game form, belief questions remain useful as they help substantiate claims about the theoretical validity of the survey instrument. In other cases, they provide a means to adjust welfare estimates to account for potential biases. There are however important identification challenges to consider when one seeks to identify the causal effect of beliefs on welfare (see Börger et al. 2021). We close this paper by reflecting on identification challenges and related research needs.

One approach for aiding identification is to vary the information provided to respondents in ways that should alter beliefs about the mechanism. In such a case, information treatments serve as instrumental variables (IVs) (e.g., Herriges et al. 2010). The drawback of such an approach is that, by definition, more respondents will have game form violations relative to a case where all respondents receive the best-designed survey. We recommend that such approaches be used only when researchers have the budget to ensure that a sufficiently large sample receives the best information. The reason for this is two-fold. First, there is no guarantee that the experimentally-generated IV is sufficiently strong. Second, even with a reasonable identification strategy, policymakers and other researchers may be suspicious of IV approaches, and could prefer relying on estimates based only on those receiving the best information.

When experimental sources of variation are unavailable or impractical, an alternative approach is to rely on other information collected in the same survey. Unfortunately, demographic information has been shown to be generally poor at explaining beliefs. The construction of new survey questions that can be used to generate IVs remains an important area of research (as a recent attempt, see Zawojska et al. 2021).

When evidence suggests that an IV approach is warranted, care must be taken when interpreting belief-conditional estimates. As emphasized in the causal inference literature, one

cannot generally interpret IV-based estimates as representative of the population unless one assumes that the effect of treatment (in our setting, holding desired beliefs) is homogenous. Otherwise, the results are "local" effects that only apply to a segment of the population. Recently proposed estimation methods (e.g., Abrevaya and Hu 2023) provide potential solutions to this issue. To our knowledge, such methods have not yet been applied to SP data. In the case of DCEs, an alternative approach to IV estimation is to identify strategic voting based on the sequence of respondent choices, and adjust for the effects of strategic behavior through econometric model specification (e.g., Day et al. 2012). Further developments along these lines represent yet another important area of future research to ensure a more complete understanding and integration of beliefs into the measurement of nonmarket values.

References

Abrevaya, Jason, and Haiqing Xu. 2023. Estimation of treatment effects under endogenous heteroskedasticity. *Journal of Econometrics* 234: 451-478.

Adamowicz, Wiktor, Peter Boxall, Michael Williams, and Jordan Louviere. 1998. Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *American Journal of Agricultural Economics* 80(1): 64-75.

Bateman, Ian J., and Ian H. Langford. 1997. Budget constraint, temporal and ordering effects in contingent valuation studies. *Environment and Planning A* 29(7): :1215-1228.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak. 1964. Measuring utility by a single-response sequential method. *Behavioral Science* 9(3): 226-232.

Bergstrom, John C., and John R. Stoll. 1989. Application of experimental economics concepts and precepts to CVM field survey procedures. *Western Journal of Agricultural Economics* 7(1): 98-109.

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

Bull, Charles, Pascal Courty, Maurice Doyon, and Daniel Rondeau. 2019. Failure of the Becker-DeGroot-Marschak mechanism in inexperienced subjects: New tests of the game form misconception hypothesis. *Journal of Economic Behavior & Organization* 159: 235-253.

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

Carson, Richard T., and Theodore Groves. 2007. Incentive and informational properties of preference questions. *Environmental and Resource Economics* 37: 181-210.

Carson, Richard T., Theodore Groves, and John A. List. 2014. Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists* 1(1/2): 171-207.

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

Champ, Patricia A., and Richard C. Bishop. 2006. Is willingness to pay for a public good sensitive to the elicitation format? *Land Economics* 82(2): 162-173.

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

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

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

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York, NY: Wiley-Interscience.

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

Gibbard, Allan. 1973. Manipulation of voting schemes: A general result. *Econometrica* 41(4): 587-601.

Hassan, Suziana, Soren Boye Olsen, and Bo Jellesmark Horsen. 2018. Appropriate payment vehicles in stated preference studies in developing countries. *Environmental and Resource Economics* 71: 1053-1075.

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

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

Kabaya, Kei. 2021. Opposite impacts of policy and payment consequentiality treatments on willingness-to-pay in a contingent valuation study. *Journal of Environmental Economics and Policy* 10(2): 175-188.

Lloyd-Smith, Patric, Ewa Zawojska, and Wiktor Adamowicz. 2020. Moving beyond the contingent valuation versus choice experiment debate: Presentation effects in stated preference. *Land Economics* 96(1): 1-24.

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

Mitchell, Robert Cameron, and Richard T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. New York and London: Resources for the Future.

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

Petrolia, Daniel R., Dennis Guignet, John Whitehead, Cannon Kent, Clay Caulder, and Kelvin Amon. 2021. Nonmarket valuation in the Enviornmental Protection Agency's Regulatory Process. *Applied Economic Perspectives and Policy* 43(3): 952-969.

Plott, Charles R., and Kathryn Zeiler. 2005. The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *The American Economic Review* 95(3): 530-545.

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

Smith, Vernon L. 1982. Microeconomic systems as an experimental science. *The American Economic Review* 72(5): 923-955.

Stithou, Mavra, and Riccardo Scarpa. 2012. Collective versus voluntary payment in contingent valuation for the conservation of marine biodiversity: An exploratory study from Zakynthos, Greece. *Ocean & Coastal Management* 56: 1-9.

Vossler, Christian A., Stéphane Bergeron, Maurice Doyon, and Daniel Rondeau. 2023. Revisiting the gap between the willingness to pay and willingness to accept for public goods. *Journal of the Association of Environmental and Resource Economists* 10(2): 413-435.

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

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

Vossler, Christian A., David A. Keiser, Catherine L. Kling, and Daniel J. Phaneuf. 2025. Information scripts and the incentive compatibility of discrete choice experiments. *Journal of the Association of Environmental and Resource Economists* 12(2).

Vossler, Christian A., and Sharon B. Watson. 2013. Understanding the Consequences of Consequentiality: Testing the Validity of Stated Preferences in the Field. *Journal of Economic Behavior and Organization* 86: 137-147.

Vossler, Christian A., and Ewa Zawojska. 2020. Behavioral Drivers or Economic Incentives? Toward a Better Understanding of Elicitation Effects in Stated Preference Studies. *Journal of the Association of Environmental and Resource Economists* 7(2): 279-303.

Vossler, Christian A., and Ewa Zawojska. 2023. Revealed preference tests of discrete choice experiments for valuing public goods: The effects of varying the number of choice options and choice questions. Paper presented at the AERE 2023 Summer Conference, Portland, Maine.

Welling, Malte, Ewa Zawojska, and Julian Sagebiel. 2022. Information, Consequentiality and Credibility in Stated Preference Surveys: A Choice Experiment on Climate Adaptation. *Environmental and Resource Economics* 82(1): 257-283.

Weng, Weizhe, Mark D. Morrison, Kevin J. Boyle, Peter C. Boxall, and John Rose. 2021. Effects of the number of alternatives in public good discrete choice experiments. *Ecological Economics* 182: 106904.

Whittington, Dale, and Stefano Pagiola. 2012. Using contingent valuation in the design of payments for environmental services mechanism: A review and assessment. *The World Bank Research Observer* 27(2): 261-287.

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

Zawojska, Ewa, Pascal Gastineau, Pierre-Aleandre Mahieu, Beoît Chèze, and Anthony Paris. 2021. Measuring policy consequentiality perceptions in stated preference surveys. Working paper.
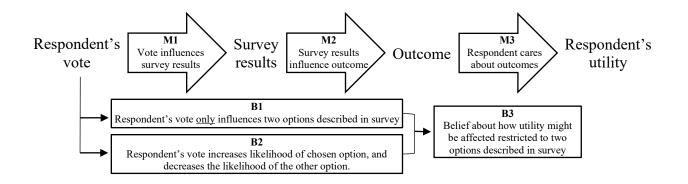
Figure 1. Diagrammatic representation of belief conditions for truthful revelation of preferences using a single binary choice referendum. M1 to M3 are beliefs that characterize the mapping between a respondent's vote and her utility (related to consequentiality). B1 to B3 are beliefs required for accurate game form recognition.

It is important for us to correctly interpret your vote on the policy proposal. Please think about why you voted the way you did. With this in mind, please indicate your level of agreement with the following statements.

	Disagree	Neutral	Agree
I voted as if my choice and those of other survey respondents will be used to inform policy makers.	○	○	○
I care about whether a policy to improve environmental quality, such as the one described in the survey, is implemented or not	○	○	○
I voted as if my choice could influence whether the policy, as described in terms of environmental improvements and cost, is implemented or not.	○	○	○
I voted as if my choice could **not** influence whether other policies not described in the survey might be implemented (for example, policies that differ in terms of environmental improvements or cost).	○	○	○
I voted as if the policy, if implemented, will achieve the stated environmental improvements.	○	○	○
I voted as if my household will have to pay the stated cost of the policy, if implemented.	○	○	○

Figure 2: Belief elicitation instrument for a single binary choice elicitation. Adapted from Vossler et al. (2025)

It is important for us to correctly interpret your votes on the policy proposals. Please think about why you voted the way you did. With this in mind, please indicate your level of agreement with the following statements.

	Disagree	Neutral	Agree
I voted as if my choices and those of other survey respondents could be used to inform policy makers.	○	○	○
I care about whether a policy to improve environmental quality, such as those described in the survey, is implemented or not	○	○	○
I voted on each proposal independently, i.e., as if my vote on one policy could only influence whether that policy is implemented.	○	○	○
I voted as if my choice could **not** influence whether other policies not described in the survey might be implemented (for example, policies that differ in terms of environmental improvements or cost).	○	○	○
I voted as if the policies, if implemented, will achieve the stated environmental improvements.	○	○	○
I voted as if my household will have to pay the stated costs of the policies, if implemented.	○	○	○

Figure 3: Belief elicitation instrument for a two-option discrete choice experiment. Adapted from Vossler et al. (2025)