Can Stated Measures of Willingness-to-Accept be Valid? Evidence from Laboratory Experiments

P. Lloyd-Smith[1]; V. Adamowicz[2]

1: University of Saskatchewan, Agricultural and Resource Economics, Canada, 2: University of Alberta, Resource Economics and Environmental Sociology, Canada

Corresponding author email: patrick.lloydsmith@usask.ca

Abstract:

Willingness-to-accept (WTA) questions have been largely abandoned in stated preference empirical work in favor of eliciting willingness-to-pay (WTP) responses, mainly due to perceived unreliability of questions that ask respondents for compensation amounts. This paper reassesses whether stated WTA welfare measures can be valid in public and private good contexts. We conduct two sets of laboratory experiments to analyze whether elicitation format, survey design and framing, and follow-up questions can generate truthful responses. For public goods, we adapt the existing WTP incentive compatibility theoretical framework to the WTA context and test the theory using an experiment involving voting. Results are consistent with the WTP literature and suggest that WTA values can be valid as long as responses have consequences for respondents. For the private good experiment, we focus on whether respondents are motivated to affect the price or the provision of the good. We find that strategic behavior is present and in the direction expected by theory. Survey framing and the use of follow-up questions can provide bounds on the value estimates. These findings raise potential concerns with the use of non-incentive compatible elicitation mechanisms in WTA contexts.

Acknowledegment:

JEL Codes: Q24, Q51

#1445

Can Stated Measures of Willingness-to-Accept be Valid? Evidence from Laboratory Experiments

January, 2018

Abstract

Willingness-to-accept (WTA) questions have been largely abandoned in stated preference empirical work in favor of eliciting willingness-to-pay (WTP) responses, mainly due to perceived unreliability of questions that ask respondents for compensation amounts. This paper reassesses whether stated WTA welfare measures can be valid in public and private good contexts. We conduct two sets of laboratory experiments to analyze whether elicitation format, survey design and framing, and follow-up questions can generate truthful responses. For public goods, we adapt the existing WTP incentive compatibility theoretical framework to the WTA context and test the theory using an experiment involving voting. Results are consistent with the WTP literature and suggest that WTA values can be valid as long as responses have consequences for respondents. For the private good experiment, we focus on whether respondents are motivated to affect the price or the provision of the good. We find that strategic behavior is present and in the direction expected by theory. Survey framing and the use of follow-up questions can provide bounds on the value estimates. These findings raise potential concerns with the use of non-incentive compatible elicitation mechanisms in WTA contexts.

Keywords: willingness-to-accept; stated preference; consequentiality; strategic behavior

1 Introduction

Asking people questions about their preferences has a long, controversial history in economics. There is an analogous parallel between the traditional predisposition amongst economists towards working with revealed preference (RP) data over stated preference (SP) data and the penchant among SP practitioners to estimate willingness-to-pay (WTP) rather than willingness-to-accept (WTA) values. The preferences for WTP measures by SP practitioners is evident in both stated and revealed sources. An influential source of the SP WTP elicitation format is the 1993 report of the NOAA Blue Ribbon Panel on Contingent Valuation (Arrow et al., 1993). This report outlines a set of best practice guidelines for value elicitation surveys that are "met by the best CV surveys and *need to be present* in order to assure reliability and usefulness of the information that is obtained [emphasis added]". In terms of choosing between WTP and WTA, the report states: "[t]he willingness to pay format should be used instead of the compensation required because the former is the conservative choice". Many authors have used these stated guidelines as justification for completely sidestepping WTA and instead estimating WTP in its place as evident in the relative empirical prevalence of the two measures.

The substantial empirical evidence suggests a clear 'revealed preference' for WTP studies among practitioners. A search of the 3,643 primary valuation studies in the Environmental Valuation Reference Inventory (EVRI) database finds that WTP studies outnumber WTA studies 14 to 1.[1] These differences are stark and suggest a large 'prevalence gap'. Moreover, WTA as a welfare measure is rarely challenged on purely theoretical grounds. In fact, WTP and WTA are both introduced in economics textbooks as useful theoretical welfare measures that economists use to think about and assess value (Freeman III et al., 2014).

This preference for WTP would appear innocuous if practitioners are only interested in situations where WTP is the proper welfare measure to use or if differences between WTP and WTA values estimates are relatively small. However, there are many contexts where eliciting WTA is clearly the conceptually more appropriate measure such as compensatory natural resource damage assessments or payment for ecosystem services (PES) schemes. Furthermore, there is a large theoretical and empirical literature demonstrating the presence of a WTP-WTA gap. Theoretical explanations of the WTP-WTA gap are based on neoclassical

[1]The search of the EVRI website (www.evri.ca, accessed on November 30, 2016) returns 2,589 studies that provide WTP estimates versus only 181 studies that estimate WTA values.

reasons such as the income effect (Willig, 1976), substitution effects (Hanemann, 1991), commitment costs (Corrigan et al., 2007), and behavioral economics explanations such as loss aversion (Kahneman and Tversky, 1979) and reference dependence (Knetsch, 2010). The empirical evidence on the WTP-WTA gap has been summarized by a recent meta-analysis of 76 studies by Tuncel and Hammitt (2014). They find that the average ratio of WTA to WTP value estimates is around 3 across all types of goods and over 6 for environmental goods. In sum, there are compelling theoretical explanations and a wide body of empirical evidence suggesting that WTP and WTA values are not always close to each other; thus using WTP estimates as proxies for WTA values may yield misleading welfare calculations and policy advice (Interis, 2014).

The main explanation for why, even in the face of evidence suggesting a WTP-WTA gap, practitioners continue to elicit WTP in lieu of WTA responses is that the latter are perceived as unreliable. The reasons for the perceived unreliability of WTA responses include the difficulties with scenario rejection and protest bids, the relative lack of respondent's experience with receiving compensation compared to making purchases, the fact that WTA responses are not bounded by income, and the strong perception of hypothetical bias in many empirical applications of WTA (Villanueva et al., 2017). Some of these reasons are unique to the WTA context but many are shared with WTP measures, although there is a worry that these factors are accentuated in a WTA context.

These reasons can be usefully grouped into two categories: the lack of incentive compatibility and non-conforming or invalid responses. Incentive compatibility implies that it is in the respondent's best interest to reveal true preferences when making choices or responding to survey questions. In general, the elicitation of WTA is not thought to be incentive compatible. For example, returning to the NOAA report, it states that "respondents would give unrealistically high answers to [willingness to accept] questions" and suggests that WTP responses are more incentive compatible compared to WTA responses. In their influential practitioner's guide to the econometrics of non-market valuation, Haab and McConnell (2002) state that SP approaches are generally perceived to be not able to elicit WTA responses because responses are not incentive compatible.

Carson and Groves (2007) describe how the incentive compatibility of the SBC question depends on whether the good is public or private. While the SBC question is incentive compatible for public goods under certain conditions, somewhat ironically, the incentive

compatibility of SBC questions does not hold for private goods. In a WTP context, when facing a take-it-or-leave-it offer, the optimal response differs depending on whether the private good is an existing good on the market or a new good entering the market. Specifically, in the context for existing goods, respondents have the incentive to respond no to take-it-or-leave-it offers to indicate more price sensitivity. We can call this effect the price bias. On the other hand, respondents should say yes and indicate less price sensitivity for new goods to increase the chance of having the option to purchase the new good in the future. This effect can be called the provision bias. In most private good stated preference applications it is not clear how respondents think of these 'price' and 'provision' biases and the extent to which these two opposing effects on choice behavior cancel each other out is an open question (Doyon and Bergeron, 2016). We re-examine these issues in the WTA context by both designing treatments that place participants in price or provision contexts and by using follow-up questions to identify the participant's motivation.

The purpose of this paper is to assess whether stated WTA can be "'rescued"' as a useful empirical welfare measure. To study these issues, we examine the validity of WTA value responses elicited in both public and private good experiments. Both experiments use the participant's time as the good to be valued which mitigates "house money" effects (Harrison, 2007). In the public goods experiment, we examine the SBC format which has been shown both theoretically and empirically to be incentive compatible in the WTP context (Vossler et al., 2012; Carson et al., 2014). We show that the mechanism perspective of incentive compatibility formalized by Vossler et al. (2012) can easily be extended to the WTA context and we derive analogous conditions for incentive compatibility. In many practical contexts, these conditions are less restrictive than their WTP counterparts but also raise several unique concerns. We provide the first test of the theory in a WTA context using an experiment involving voting to accept payment to give up a public good using the SBC format. Results provide support for the incentive compatibility of WTA responses to SBC questions for public goods as long as responses have consequences for respondents.

For the private good experiment, we assess and compare different types of strategic behavior on decisions in a WTA context, focusing on whether respondents are motivated to affect the price or the provision of the good. The overall spirit of the private good experiment is to mimic a SP survey for setting up a PES scheme. We choose a PES scheme, as this is a context where WTA is the only meaningful measure: it is clearly a private good (i.e. landowner has a voluntary choice to engage in a contract involving payment for private action), and there are

several examples in the literature of such elicitation (Porras and Hope, 2005; Horne, 2006; Espinosa-Goded et al., 2010; Southgate et al., 2010; Kaczan et al., 2013). In the experiment, the program is the opportunity to receive payment to give up an hour of time at a later date. We use this program as it is a good that all individuals "own", contracts can be written to implement the transaction, and we can determine to offer the contract or not to respondents. Participants are first presented with a SP question and are then given the opportunity to submit an offer to participate in the program.

There are three main results for the private good experiments. First, strategic behavior is present in SP questions for private goods for some but not all participants and follow-up questions can be helpful in identifying strategic behavior. Second, strategic behavior biases are in the directions expected by theory with price (provision) motivated individuals less (more) likely to accept offers. Third, explicit framing of the questionnaire and the use of follow-up questions can provide useful bounds on value estimates. These findings raise potential concerns with the use of non-incentive compatible elicitation mechanisms in WTA contexts and provides an alternative approach to eliciting values in private good contexts. The results reinforce the need for special consideration to be given to strategic behavior in WTA surveys that value private goods, such as PES programs.

2 Experiment Implementation

We conducted two experiments at a University using participants from the University's Experimental Database. Participants are a mix of undergraduate and graduate students from all Faculties as well as campus staff. A total of 13 experimental sessions were conducted involving 202 participants.[2] The experiments were conducted between March 4 and March 16, 2016. Subjects only participate in the experiments once. The private and public good experiments were conducted as part of the same session, and for all sessions the private good experiment was conducted first, followed immediately by the public good experiment.

One of the potential problems with economic experiments is the influence of money/goods provided by the experimenters. Participants may treat participation fees or goods that are given differently than their own money/goods in making decisions (Harrison, 2007).

[2]Experiment group sizes ranged from 12 to 19 and respondents were paid a $20 show-up fee. To minimize possible experimental biases, all experiments were conducted by the same individual, wearing the same shirt.

This 'house money' or 'gift' effect may limit the researcher's ability to generalize results from the laboratory setting to the outside world. There is empirical evidence suggesting individuals treat small one-time financial gains differently than their regular income (Keeler et al., 1985). Thaler and Johnson (1990) attribute these effects to a 'mental accounting' framework where small, one-time windfall gains are placed in a 'mad money' account with a higher marginal propensity to spend. Harrison (2007) finds that house money does affect behavior of participants in a public good contribution experiment. One approach to side-stepping these issues in WTA experiments is to use goods that respondents already own (Cash, 2015). This approach can work well in private goods contexts because participants can voluntarily accept money to give up the good, but faces challenges for public goods.

To mitigate these gift effects in the experiments, we use the participant's own time as the good to be given up in exchange for money. The private good considered in this experiment is to give up one hour of the participant's time at a within one week of the experiment's date. During the hour, participants would be working for one hour in the department's library helping sort and organize books. In the public good setting, we cannot coerce respondents to give up their time outside of the experiment. As a solution, we use the last 30 minutes of time in the experiment as the public good to be valued.

3 WTA for Public Goods

For public goods, the SBC referendum style question has long been held up as an example of an incentive compatible elicitation mechanism (Mitchell and Carson, 1989; Harrison, 2006). However, it was not until Vossler et al. (2012) that a formal model was developed of the incentive compatibility properties of binary choice questions that incorporated the recent insights on the importance of consequentiality. Vossler et al. (2012) develop an explicit game theoretic model to describe the conditions informally sketched out in Carson and Groves (2007) under which respondents have the incentive to reveal their true preferences. In the WTP context, they identify four sufficiency conditions for truthful voting 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.

The first two conditions ensure that at least the costs and possibly the attributes of the project enter the respondent's utility function. The third condition eliminates the possibility for votes for one project to affect the probability of different projects being implemented. The fourth condition states that a yes vote increases the probability of the project being implemented at least some of the time. Carson et al. (2014) extend Vossler et al. (2012) theoretical results by relaxing the expected utility assumption and find that the incentive compatibility properties hold. Both Vossler et al. (2012) and Carson et al. (2014) provide empirical evidence using field experiments supporting the theoretical results on the incentive compatibility of WTP.

Neither Vossler et al. (2012) nor Carson et al. (2014) discuss WTA contexts, but we can extend the theoretical model by changing condition (ii). In the WTA context, the sufficient conditions for a truthful vote according to the participant's preference between a single project and the status quo are:

(i) the participants care about the outcome;

(ii) the authority can enforce voters to give up the good;

(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.

Condition (ii) in the WTA context is the authorities can enforce voters to give up the good rather than payment as in the WTP context. Theoretically, this condition ensures that the outcomes are binding for participants and the good (or its attributes) and payment enter the participant's utility function. Practically, however, there can be important differences in the restrictiveness of this condition. In the WTP case, condition (ii) is actually quite restrictive as different payment vehicles may not be binding for different people. For example, if the payment vehicle is income tax increases, many individuals do not pay income taxes and thus may not view these costs as real for themselves. Thus, payment consequentiality has also been raised as an important component of the incentive compatibility of survey responses (Herriges et al., 2010). On the surface, condition (ii) in the WTA also appears quite restrictive as the authority may not have the ability to take certain goods away from people against their will. However, many environmental goods are public goods such as air

quality or natural areas and the authorities already effectively own or control the access to the good or the quality of the good. In other cases where the authorities have less control over the good, legally enforceable contracts can be written for payment.

3.1 Public Good Experiment Structure

The format of the public good experiment adapts the WTP experiment described in Carson et al. (2014) to the WTA context. We follow the script in Carson et al. (2014) as closely as possible to provide comparable results. The good in our experiment is the participant's time and participants are told they would collectively decide what to do with the last 30 minutes of time in the experiment. A vote is taken on whether everyone in the group would spend 30 minutes filling out a survey or leave immediately after the vote.

The experiment consists of three treatments. In the baseline "real" treatment, participants are told that if more than 50% of the people in the group voted in favor, then everyone would receive $3 to give up 30 minutes of time to fill out the survey. If 50% or fewer people in the group voted no, then everyone could leave immediately after the vote and receive no additional money. In the hypothetical treatment, the vote is not binding and participants are told that "*regardless of the vote outcome, no one will receive an additional $3 or have to stay the extra 30 minutes to fill out the questionnaire*". The "consequential" treatment included a probabilistic referendum that set the probability that the referendum would be binding to 50%. In this treatment, a two-step referendum format is used. The first step consists of participants voting using the same simple majority voting rule to determine if the referendum passes as the real treatment. If the referendum passes, the second step determines if the referendum is binding with a flip of a coin. If the referendum binds, every participant is paid the $3 and has to give up the 30 minutes of time to fill out the survey. If the referendum does not pass or does not bind, no compensation is paid and all participants can leave immediately after the vote.

3.2 Hypotheses and Analysis

The experimental design allows us to test several hypotheses pertaining to the response of voting behavior to the different treatments. The three hypotheses to be tested are formally

stated and then discussed. Let p be the probability that the treatment is consequential.

1. Hypothesis 1: The percentage in favor at p = 0.5 (consequential) is equal to that of p = 0 (hypothetical)

2. Hypothesis 2: The percentage in favor at p = 0.5 (consequential) is equal to that of p = 1 (real)

3. Hypothesis 3: The percentage in favor at p >0 (consequential and real) is equal to that of p = 0 (hypothetical)

In light of the incentive compatibility framework, we expect to reject Hypothesis 1 as hypothetical (i.e. inconsequential) and consequential voting behavior need not be the same. We expect to not reject Hypothesis 2. According to the theory of consequentiality, participants should treat consequential and real votes similarly. We expect to reject Hypothesis 3 for the same reasons as Hypothesis 1.

3.3 Public Good Experiment Results

Table 1 presents the voting distribution for the three treatments. The voting results show that only 46.8% (37/79) voted yes in the hypothetical treatment compared to 63.9% (39/61) in the consequential treatment and 69.4% (43/62) in the real treatment. The near 50-50 voting split for the hypothetical treatment suggests that participants may have behaved as though they are flipping a coin due to the inconsequential nature of the decision. This inconsequential voting pattern has been found in previous studies (Cummings et al., 1997; Burton et al., 2007).

Table 1: Public good experiment voting summary statistics

Treatment	Number of Sessions	Number of Participants	Vote (%)
Hypothetical	5	79	46.8
Consequential	4	61	63.9
Real	4	62	69.4

To statistically test our hypotheses, we use different subsets of the data to estimate probit regression models of the probability of a yes response against different treatment dummy

variables. Table 2 shows the results for the three hypotheses. The first column compares the hypothetical and consequential treatment data and includes a dummy variable for the hypothetical treatment to test Hypothesis 1. We reject the hypothesis at the 0.05 level that responses in the two treatments are equal. The second column in Table 2 shows the results for Hypothesis 2. The results show that there are no statistically significant differences in voting behavior between the consequential and real treatments. The final column of Table 2 includes data for all three treatments and a single dummy variable for the responses from the hypothetical treatment. Results are consistent with Hypothesis 3 and show that voting behavior in the hypothetical treatment is statistically significant different from voting behavior in the other two treatments.

Table 2: Probit regression results for public good experiment hypotheses

	Hypothesis 1	Hypothesis 2	Hypothesis 3
Hypothetical treatment	-0.436^{**}		-0.510^{***}
	(0.217)		(0.183)
Consequential treatment		0.149	
		(0.234)	
Constant	0.357^{**}	0.357^{**}	0.431^{***}
	(0.164)	(0.164)	(0.117)
Treatment included			
Hypothetical	Y	N	Y
Consequential	Y	Y	Y
Real	N	Y	Y
Observations	140	123	202
Log Likelihood	-94.48	-78.09	-132.89

Note: Standard errors are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01

Taken together, the results are consistent with our theoretical expectations regarding the importance of the incentive compatibility of WTA elicitation mechanisms. However, the results so far do not control for participant characteristics. The random assignment of respondents to treatment should render such controls unnecessary and a pair-wise comparisons between the different treatment sub samples confirms this result. As a robustness check, we estimate models with such controls none of the results substantially change with respondent

characteristic controls.

Another consideration in comparing hypothetical and binding referendums is the variance associated with the choices (Haab et al., 1999; Carson et al., 2014). We use various probit model specifications, including the heteroskedastic probit model, to statistically test the effect of allowing the different treatments to affect both the mean and variance of the underlying WTA distributions. The results of these tests corroborate the results presented in Table 2.

4 WTA for Private Goods

As mentioned above, the incentive compatibility of SBC questions does not hold for private goods in the majority of cases (Carson and Groves, 2007). When facing a take-it-or-leave-it offer for a private good, respondents have the incentive to respond no to indicate more price sensitivity for existing goods (i.e. the price bias) and have the incentives to say yes to increase the chance of having the option to purchase the new good in the future (i.e. the provision bias). These biases were studied in a WTP laboratory experiment by Lusk et al. (2007) who find support for strategic behavior responses for some people.

For private goods, while Carson and Groves (2007) do not discuss incentive compatibility in the WTA context, we hypothesize that the strategic behavior biases would be similar to their WTP counterparts. Specifically, it would be optimal to say no to questions involving existing offers for goods to increase perceived necessary compensation levels and to say yes to potential new offers for goods to increase the chance of the offers being introduced. Using a PES scheme as an example, respondents to a survey for a new scheme may have the incentive to understate their WTA to increase the likelihood of the program being implemented, for which later on they can decide whether to participate. Alternatively, if the PES scheme is already in operation and a SP survey is administered to assess the potential to tinker with the compensation levels, respondents have the incentive to overstate their WTA.

There are three types of empirical strategies available to improve the incentive compatibility of WTA responses in private goods contexts: different elicitation mechanisms (Lusk and Shogren, 2008), auction cheap talk scripts (Krishna et al., 2013; Kanjilal, 2015), and learning rounds and rationality spillovers (Chilton et al., 2011). Lusk and Shogren (2008) review

the experimental literature considering the use of different elicitation mechanisms such as Vickrey auctions, BDM mechanisms and the random incentive system. There is also a substantial body of theoretical and empirical work in the field using conservation auctions (Latacz-Lohmann and Van der Hamsvoort, 1997; Latacz-Lohmann and Schiizzi, 2005). Most of this work has focused on the optimal design of auction mechanisms with the aim of reducing information rent and/or improving environmental outcomes. Cheap talk scripts have also been used to increase the validity of WTA responses. Krishna et al. (2013) introduce the concept of competitive bidding into their survey of compensation payments to Indian farmers to make the responses more incentive compatible. Respondents were told that the government had a limited budget and only the least-cost providers would be selected to participate in the PES scheme. Another example of the use of cheap talk script is in Kanjilal (2015) who use a description of the auction mechanism that would be used in a real program before eliciting WTA responses from farmers in Saskatchewan and Alberta. Similar to the use of cheap talk scripts to reduce hypothetical bias, we do not know if incentive compatibility cheap talk scripts can completely induce truthfully responses. With the exception of auction cheap talk scripts, the main limitation with these existing approaches is that they do not easily transfer to the SP survey setting.

4.1 Private Good Experiment Structure

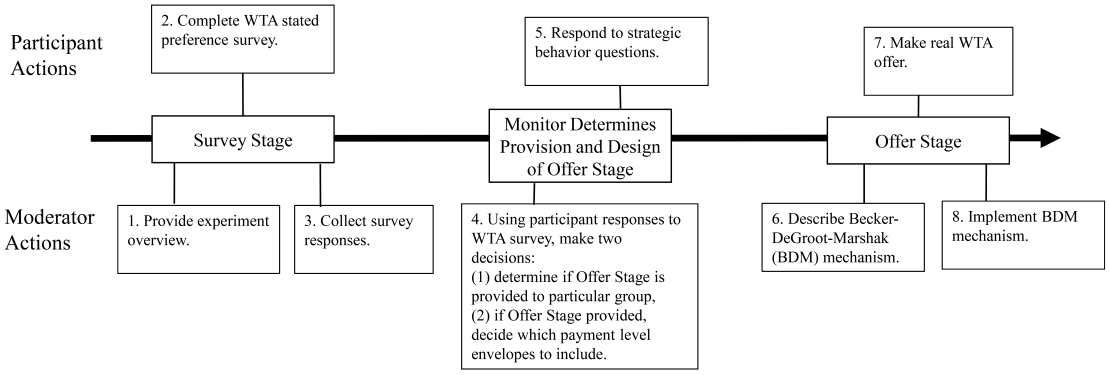
The objective of the second experiment is to assess and compare different types of strategic behavior on choice behavior in a WTA private good setting. The overall spirit of the private good experiment is to mimic a SP survey for setting up a PES scheme. The scenario is modeled on the introduction of a new voluntary program that will pay participants compensation for undertaking some costly activity. Participants are told how their responses to the SP survey will be used to determine whether the program is offered to the group and the expected program payment levels. If the program is offered to the group, participants then have a second choice of whether to participate in the program and at what price.

To identify strategic behavior in WTP settings, follow-up questions have been proposed as one method to identify strategic behavior (Lusk et al., 2007; Doyon and Bergeron, 2016). However, using these responses in statistical models of behavior is challenging due to potential endogeneity concerns. To complement the use of follow-up questions, we also use different treatments that introduce a clear price or provision framing to the use of the SP

survey. These exogenous treatments help provide an alternative means of assessing the effects of strategic behavior on choice decisions.

The experiment itself consists of two stages: a Survey Stage and an Offer Stage. The Survey Stage includes a SP question and a description of how responses to this question will be used by the moderators. After the Survey Stage, the moderators decide whether to provide the Offer Stage to the particular experiment group or not and what payment levels to include in the Offer Stage. If the Offer Stage is provided to the group, respondents make an offer of the minimum amount that they would be willing to accept to sell an hour of their time. The private good considered in this experiment is to give up one hour of the participant's time to work in a library sorting and organizing books at a later date. Figure 1 provides an overview of the two experiment stages as well as the sequential actions by participants and the moderator. We describe each stage and action in detail below.

Figure 1: Overview of private good experiment and actions by participants and moderators



4.1.1 Survey Stage

The Survey Stage starts with the moderator providing a brief overview of the experiment including an overview of Offer Stage (Step 1). Participants are informed that the Offer Stage may not be provided to every group and different payment level may be presented to different groups. Participants are then asked to respond to a SP question and are told how their responses will be used (Step 2). Specifically, participants are told that

*Responses in the **Survey Stage** will help the researchers decide whether or not to present the **Offer Stage** to this group as well as determine which payment level envelopes to provide*

*in the **Offer Stage**, depending on our budget.*

The choice question uses the stochastic payment card (SPC) approach. The SPC approach is closely related to the multiple-bound discrete choice (MBDC) but allows respondents to use a combination of words and numerical values to more easily express themselves (Wang and Whittington, 2005). We employ a SPC approach as it is well suited to the context, parallels the consequential ("real") treatment, and provides considerable value information for each individual. The SPC question used in the experiment is provided in Figure 2. All responses in the Survey Stage are collected by the moderator (Step 3).

4.1.2 Moderator Determines Provision and Design of Offer Stage

The moderator analyzes the responses to the Survey Stage and makes two decisions (Step 4). The first decision is whether to provide the Offer Stage to this particular experimental group or not. The second decision is what payment levels to include in the Offer Stage. The moderator has 15 different envelopes containing different payment amounts ranging from \$5 to \$20 and decides on 10 to include in the Offer Stage. Note that there is no explicit link between the participant's individual response to the Survey Stage and the Offer Stage. Instead, the aggregate information from the Survey Stage is used to inform the Offer Stage design.

While the moderator is analyzing the Survey Stage responses, participants are asked two different questions to identify the presence of strategic behavior in the Survey Stage (Step 5). The first question format focuses on the motivation of the respondents:

Which of the following motives were important when deciding to accept or not accept the offer at each payment level? Select all that apply.

- *A. Ensure the offer is worth it for myself given the payment levels presented*
- *B. Ensure the Offer Stage is provided to my group*
- *C. Ensure the payment levels in the Offer Stage are favorable for myself*

If participants selected more than one motivation, they are asked a follow-up question on which one is the most important. The first motivation option identifies participants who

Figure 2: Example of the stochastic payment card design

Suppose you have been given the opportunity to be paid money to give up one hour of your time to work in the department's library helping sort and organize library books. The work can be completed any weekday between 9 am and 5 pm and the work must be completed within two weeks of today's date (March 24, 2016). Please think about a suitable time when you would complete the work before you respond to the question. You would be paid once you've completed the hour of time.

How likely would you be to accept the offer to give up one hour of your time if the payment amount is....? Please select a response for each payment amount (one response per row).

Payment for 1 Hour of Time	Definitely Yes (100% chance)	Probably Yes (75% chance)	Not Sure (50% chance)	Probably No (25% chance)	Definitely No (0% chance)
$5	☐	☐	☐	☐	☐
$10	☐	☐	☐	☐	☐
$15	☐	☐	☐	☐	☐
$20	☐	☐	☐	☐	☐
$25	☐	☐	☐	☐	☐
$30	☐	☐	☐	☐	☐
$35	☐	☐	☐	☐	☐
$40	☐	☐	☐	☐	☐

are motivated by whether the program is beneficial to them, and not necessarily strategic behavior. The second option identifies 'provision motivation' and participants who are focused on the provision of the Offer Stage. The third option identifies 'price motivation' and participants who aim to affect the payment levels in the Offer Stage.

The second strategic behavior question format focuses on the perceived price and provision

consequences of the Survey Stage choices. Using a 1 to 5 scale, where 1 is not at all and 5 is definitely taken into account, participants are asked a price consequence question, "*to what extent do you think your choices will be taken into account for determining which payment levels are provided in the Offer Stage*", and a provision consequence question, "*to what extent do you think your choices will be taken into account for determining the chances of being provided the Offer Stage*".

4.1.3 Offer Stage

The moderator describes the open-ended BDM mechanism with a random payment amount as the elicitation mechanism (Step 6). Participants make a monetary offer of the minimum dollar amount they would be willing to accept to give up an hour of their time as described earlier (Step 7). The offers are compared to a randomly drawn payment level contained in one of the 10 envelopes held by the moderator (Step 8). No information on the range or distribution is provided to the participants. There are two possible outcomes. Offers are accepted if the amount they indicate is less than or equal to the randomly drawn payment level. In this case, participants would receive the randomly drawn payment level in dollars and have to give up an hour of their time. Offers are not accepted if the amount they indicate is more than the randomly drawn payment level. In this later case, they do not receive any money and do not have to give up an hour of time.

4.1.4 Treatments

1. Consequential treatment: This treatment serves as the baseline and the text before the Survey Stage describes how responses will be used to inform the decision to present the Offer Stage to the group and the design of payment levels provided. The Offer Stage is a real binding contract that may or may not be offered to participants depending on their responses to the Survey Stage.

    - Text before Survey Stage: *Responses in the Survey Stage will help the researchers decide whether or not to present the Offer Stage to this group as well as determine which payment level envelopes to provide in the Offer Stage, depending on our budget.*

2. Hypothetical treatment: This treatment is exactly the same as the consequential treatment, but the Offer Stage is hypothetical. The following text is provided at the beginning of the Offer Stage: *You will not actually have to give up the time nor will you receive the money. But we ask that you make choices as if you were actually making a real money decision on whether to accept the payments and give up an hour of your time.*

    - Text before Survey Stage: Same as consequential

3. Price framing treatment: This treatment fixes the provision of the Offer Stage and the text before the Survey Stage tells participants that their responses will be used to design the payment levels (i.e. compensation amounts) offered in the Offer Stage. The Offer Stage is a real binding contract.

    - Text before Survey Stage: *Responses in the Survey Stage will help the researchers determine which payment level envelopes to provide in the Offer Stage, depending on our budget.*

4. Provision framing treatment: In this treatment, the payment level envelopes are selected before the experiment instead of deciding which payment levels to include in the Offer Stage based on Survey Stage responses. The text before the Survey Stage tells participants that their responses will be used to determine whether or not the Offer Stage will be offered. The Offer Stage, if provided, is a real binding contract.

    - Text before Survey Stage: *Responses in the Survey Stage will help the researchers decide whether or not to present the Offer Stage to this group, depending on our budget. The payment levels potentially being provided in the Offer Stage have already been determined.*

4.2 Private Good Experiment Results

Table 3 provides a summary of participants and monetary WTA offers in the private good experiment treatments. The Monetary Offer column is the mean of the values provided by participants in the Offer Stage. For all the non-hypothetical treatments, a total of 6 offers were accepted and 5 individuals followed-up to receive payment and work in the department

for one hour each.

Table 3: Willingness-to-accept to give up an hour of time in the private good experiment

Treatment	Number of Participants	Mean Monetary WTA Offer	Range of WTA Offers (min - max)
Consequential	65	\$19.2	(\$1 to \$40)
Hypothetical	42	\$16.5	(\$5 to \$30)
Price framing	49	\$17.9	(\$5 to \$40)
Provision framing	46	\$17.8	(\$5 to \$30)
Total	202	\$18.0	(\$1 to \$40)

To assess the issues surrounding strategic behavior in the private good context, we present the results as answers to three questions:

1. To what extent is strategic behavior present in SP questions eliciting values for private goods?

2. How does the strategic behavior of respondents affect private good choice decisions?

3. Can the treatments and strategic behavior questions provide informative bounds on WTA?

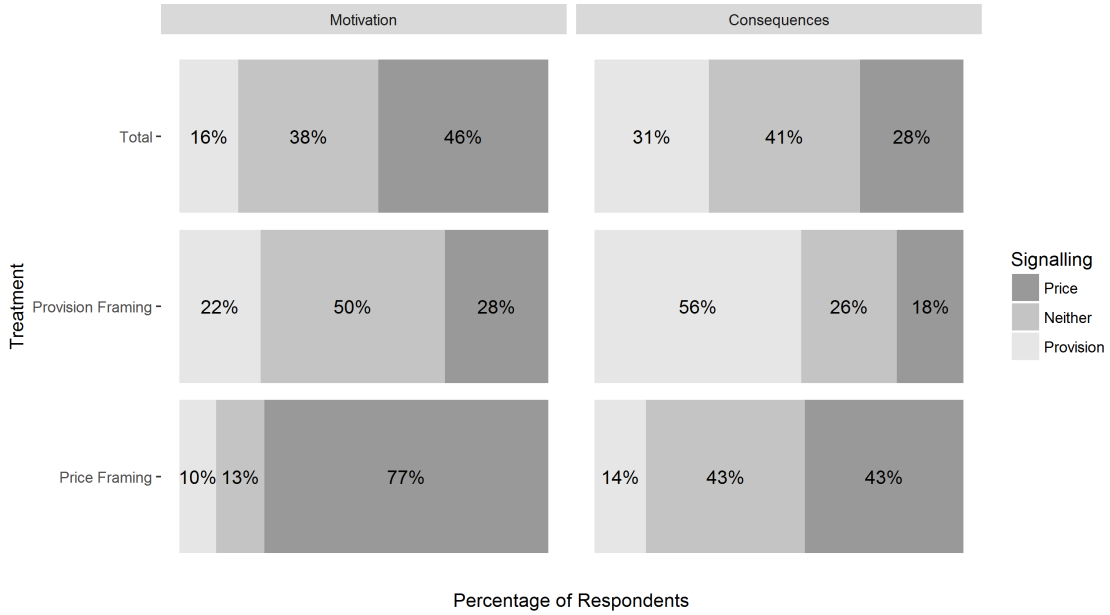
To what extent is strategic behavior present in SP questions eliciting values for private goods?

To answer this question, we examine how the responses to the strategic behavior follow-up questions varied by treatment.We compute the percentage of respondents who perceive their responses as influencing the prices or provision of the Offer Stage and summarize the results in Figure 3. The left panel of Figure 3 reports the percentage of respondents who are price or provision motivated in each treatment. These motivation questions are intended to identify if the respondent thought their Survey Stage responses were being used primarily to ensure the Offer Stage is provided to their group (Provision motivation) or used to set the payment levels provided in the Offer Stage (Price motivation). Across all treatments, 46% of participants indicated they are motivated by price considerations while 16% indicated a provision motivation, with the remaining 38% indicating neither price nor provision motivations. As expected, the percentages of price and provision motivated individuals are highest in their corresponding treatments with 77% of respondents in the price framing treatment and 22% of respondents in the provision framing treatment indicating price and provision motivations,

respectively.

Using the consequences follow-up questions, we code individuals as price consequential if they indicate a larger number for the price consequentiality question compared to the provision consequentiality question and vice versa for provision consequential. The right panel in Figure 3 report the percentage of individuals who are price and provision consequential or neither. Around 28% and 31% of respondents are price and provision consequential, respectively, across all the treatments. As with the motivation follow-up question, the percentages of price and provision consequence focused individuals are highest in their corresponding framing treatments.

Figure 3: Identifying strategic behavior in the private good experiment



Notes: Numbers are percentages of respondents in each treatment group. The left panel uses the motivation strategic behavior follow-up question and presents the percentage of respondents who indicated they are price or provision motivated or neither in the treatments. The right panel uses the consequences strategic behavior follow-up question and reports the same numbers.

To more formally examine the impact of the framing treatments on responses to the two strategic behavior follow-up questions, we estimate a set of probit models using price and provision strategic behavior responses as the dependent variable and include dummy variables for the different treatments. Table 4 presents the results. For the motivation follow-up question, the provision framing treatment had a negative and significant effect on responses to the price motivation question and a positive and insignificant effect on responses to the

provision motivation question. As expected, the price framing treatment had a positive and significant effect on price motivated responses and a negative and significant effect on provision motivated responses.

Table 4: Probit models of framing treatments on strategic behavior

	Motivation		Consequences	
Parameter	Price	Provision	Price	Provision
Hypothetical treatment	-0.134	0.427*	-0.374	0.718***
	(0.250)	(0.254)	(0.276)	(0.265)
Provision framing treatment	-0.586**	0.035	-0.421	1.038***
	(0.246)	(0.254)	(0.273)	(0.260)
Price framing treatment	0.504**	-0.723**	0.322	-0.17
	(0.254)	(0.293)	(0.243)	(0.286)
Constant	0.253	-0.547***	-0.502***	-0.898***
	(0.157)	(0.164)	(0.163)	(0.181)
Observations	202	202	201	201
Log Likelihood	-129.06	-112.6	-115.09	-110.78

Note: Standard errors are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01. For all models, the Consequential treatment is the omitted reference category.

The last two columns of Table 4 present the probit model results using responses to the consequences strategic behavior follow-up questions as dependent variables. We can see that the framing treatments are not as effective in influencing the strategic behavior of respondents compared to their motivation as captured in the motivation questions. Only the provision framing treatment has a statistically significant impact on provision consequences. Given the lack of response of the consequences follow-up question to the different framing treatments, we only present the results using the motivation follow-up questions in subsequent sections.

Taken together, these results suggest that a significant portion of the sample was behaving strategically in the Survey Stage. Furthermore, the strategic behavior follow-up questions appear to be identifying price and provision signaling as the framing treatments affect strategic behavior in the expected direction.

How does the strategic behavior of respondents affect private good choice deci-

sions?

For the second question, we investigate how the framing treatments and strategic behavior follow-up questions affect the decision of whether to accept a given payment level in the Survey Stage. The polychotomous responses for each payment level are converted into a binary variable using the 'Probably Yes (75%)' as the lower bound cut-off for a 'yes' response. We use these converted binary responses and estimate a random effects logit model to account for multiple responses per individual. Table 5 presents these random effects logit model results. The first column shows the results using only a constant term and the payment amount. As expected, the higher the payment amount, the more likely an individual is to give up an hour of their time. The second column includes dummy variables for the hypothetical, provision framing, and price framing treatments with the consequential treatment being the reference category. The provision framing treatment variable is positive and significant implying that participants in this treatment are more likely to accept the program. This result is consistent with the provision signaling hypothesis as respondents had an incentive to underbid in this treatment to increase the likelihood of being presented with the Offer Stage. The price framing treatment variable is close to zero and is not significant.

The third column adds the strategic behavior follow-up questions to the model while acknowledging that these variables are potentially endogenous. We also include interaction terms between these motivation responses and the payment amount because these two motivations will have different impacts on the price sensitivity of respondents. Specifically, provision motivated respondents will want to indicate less price sensitivity while price motivated respondents will want to indicate that they are more price sensitive. The results support the provision and price signaling hypotheses. The positive and significant coefficient for the provision motivation variable indicates that these respondents are more likely to accept the program while the negative and significant coefficient for the price motivation variable suggests the opposite. Furthermore, the interaction term between the payment amount and provision motivation is negative, as expected, as these respondents are less price sensitive. The positive and significant coefficient for the payment amount and price motivation interaction term suggests that these respondents are more price sensitive. Overall, it appears that strategic behavior did influence private good choices in the directions expected by theory.

Can the treatments and strategic behavior questions provide informative bounds

Table 5: Random effects logit models of the decision to accept payment for time

Parameter	(1)	(2)	(3)
Constant	-4.919^{***}	-5.381^{***}	-5.460^{***}
	(0.370)	(0.458)	(0.634)
Amount	0.329^{***}	0.329^{***}	0.329^{***}
	(0.021)	(0.021)	(0.030)
Hypothetical treatment		0.751	0.812
		(0.460)	(0.502)
Provision framing treatment		1.203^{***}	1.265^{**}
		(0.454)	(0.492)
Price framing treatment		0.138	0.110
		(0.436)	(0.493)
Provision motivation			2.263^{***}
			(0.785)
Price motivation			-2.084^{***}
			(0.777)
Amount*Provision motivation			-0.125^{***}
			(0.038)
Amount*Price motivation			0.133^{***}
			(0.042)
Observations	1,612	1,612	1,612
Participants	202	202	202
Log Likelihood	−490.56	−486.20	−467.44
AIC	987.11	984.39	954.88

Note: Standard errors are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01
AIC = Akaike Information Criteria. The dependent variable is whether the participant would accept the payment level using 'Probably Yes (75%)' and 'Definitely Yes (100%)' responses as yes responses. The provision and price motivation variables represent whether the participant indicated these are motivations in responding to the survey. For all models, the Consequential Treatment is the omitted reference category

on WTA?

For the third question, we use the different framing treatments and strategic behavior questions to provide bounds on WTA value estimates from the Survey Stage. We consider the framing treatments without the potential endogenous follow-up questions and use the model estimates from the second column of Table 5. As a lower bound, the WTA for the participants in the provision framing treatment is estimated to be \$12.68 (standard error (se): 1.047). As an upper bound, the WTA for participants in the price framing treatment is estimated to be \$15.92 (se: 1.000). These lower and upper bound estimates of WTA can be used to inform the potential range of WTA values.

5 Conclusion

Economists have largely abandoned SP WTA welfare measures in empirical work. In this paper, we reassess whether this abstinence is justified. We report the results of two WTA experiments involving private and public good contexts. Both experiments use time as the good to be valued which mitigates house money effects and provides a useful public goods context. For public goods, we adapt the WTP incentive compatibility proof outlined in Vossler et al. (2012) to the WTA contexts. These adapted incentive compatibility conditions highlight two key differences between the WTP and WTA context. First, payment consequentiality in the WTA context is quite different compared to the WTP context. Finding and defining a credible payment vehicle that applies to the population of interest has remained a challenge in the WTP context (Johnston et al., 2017). In the WTA context, these challenges may be less of an issue because people have a greater incentive to receive rather than provide payment. However, there may be remaining concerns about the government actually following through on its pledges to pay citizens or reduce their taxes. Second, the WTA context involves the government or relevant authority having the ability to enforce voters to give up the good and enforcement could be an issue. However, governments often already control the access, quality, or quantity of many public goods such as air or water quality and in many cases can credibly determine the final amount of good provided. Conversely, in the case of a damage assessment after an event, the good is effectively already taken away and the WTA question only relies on the perceived consequentiality of payment if not cleaned up. Depending on the context, these two differences can have important practical impacts on the viability of eliciting either welfare measure.

The adapted WTA incentive compatibility theory for public goods is tested in a laboratory experiment that places people in a hypothetical, consequential, or real treatment. The results provide strong support for the use of consequential SBC WTA questions as long as the probability that the responses are taken into account is greater than zero. While the experimental set-up abstracts away from many issues with eliciting WTA questions in practice, Whittington et al. (2017) provides some guidance on asking WTA questions in SP surveys.

The second part of the paper focuses on eliciting WTA values for private goods. The experimental results suggest that strategic behavior is present in SP questions eliciting values for private goods. Identifying strategic behavior is not straightforward however and the approach in this paper used both explicit survey framing and follow-up questions. While the explicit framing approach benefits from randomly placing people in separate price and provision framing treatments, this may not be appropriate or viable in all private good contexts. The use of strategic behavior follow-up questions suffers from many of the same issues that plague the increasing use of perceived consequentiality questions (Herriges et al., 2010). Both benefit from being relatively simple to append to a usual SP survey, but using these responses in data analyses raises issues of potential endogeneity. Furthermore, while a strategically behaving participant may not voluntarily reveal this information in the follow-up question, participants who chose to misrepresent their motivations makes detecting the influence of strategic behavior more difficult. Conducting a large field experiment with both survey framing and follow-up questions would allow the endogeneity of these questions to be controlled for using an instrumental variable strategy.

The second main result of the private good experiment is that strategic behavior affects private good decisions in the directions expected by theory. Participants who are price motivated are less likely to accept the program and show greater price sensitivity while provision motivated individuals are more likely to accept the program. The results also demonstrate how survey framing and follow-up questions can be used to provide bounds on WTA values.

The results of this paper have implications for the burgeoning use of SP methods in designing PES schemes. There is a lack of clarity in the literature on how PES schemes fit into the public/private good distinction and the appropriate role of SP elicitation methods. In developing PES schemes, SP methods have been used to either estimate the WTP of users

of the services for the program or the WTA of landowners to accept payment.[3] Part of the confusion stems from the fact that PES schemes are often targeted at providing public goods for the users such as carbon sequestration or water quality improvements. However, from a landowner's standpoint, the payments and activities associated with these programs more closely mimic private goods.[4] As outlined above, the private/public good distinction is important for incentive compatibility and if we interpret PES schemes to be private goods for landowners then SP research has exclusively used non-incentive compatible elicitation mechanisms. For example, Liu et al. (2014) use a SP survey to elicit the WTA of landowners in Iowa for adopting perennial strips. They conduct a convergent validity test of a traditional multinomial choice question, a modified multinomial choice question suggested by Carson and Groves (2007) where all but one of the alternatives are implemented, and a SBC question.[5] For public goods, the latter two elicitation schemes are incentive compatible, but none of the three elicitation mechanisms are generally incentive compatible for private goods. Other SP research on estimating WTA for PES schemes have used elicitation mechanisms that are not necessarily incentive compatible such as binary choice questions (Southgate et al., 2010), open-ended questions (Southgate et al., 2010), and choice experiments (Porras and Hope, 2005; Horne, 2006; Espinosa-Goded et al., 2010; Kaczan et al., 2013). While there is no commonly accepted survey-based elicitation mechanism that is incentive compatible for private goods, the results of this study show how explicitly incorporating strategic behavior framing into the design of the survey and the use of strategic behavior follow-up questions can help inform and control for biases in value estimates.

References

Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R., and Shuman, H. (1993). Report of the NOAA Panel on Contingent Valuation. Technical report.

Burton, A. C., Carson, K. S., Chilton, S. M., and Hutchinson, W. G. (2007). Resolving

[3] In the developing world context, Whittington and Pagiola (2012) review and assess the use of CV in the design of PES and solely focus on WTP studies in their evaluation due to a paucity of WTA studies.

[4] The WTA PES schemes can be considered private goods from a landowner's perspective because private actions (i.e. restoring a wetland) are being compensated with individual payments and both are excludable and rival. Of course, the benefits of restoring a wetland may have broader public good-like benefits.

[5] The results in Liu et al. (2014) suggest no significant difference in estimates between the two multinomial choice questions while the results diverge between the multinomial and SBC formats.

questions about bias in real and hypothetical referenda. *Environmental and Resource Economics*, 38(4):513–525.

Carson, R. T. and Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, 37(1):181–210.

Carson, R. T., Groves, T., and List, J. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists*, 1(1):171 – 207.

Cash, S. (2015). The Facebook Suicides: Using Second-Price Auctions to Determine Attachment and Addiction to Facebook.

Chilton, S., Jones-Lee, M., McDonald, R., and Metcalf, H. (2011). Eliciting reliable willingness-to-accept responses.

Corrigan, J. R., Kling, C. L., and Zhao, J. (2007). Willingness to pay and the cost of commitment: An empirical specification and test. *Environmental and Resource Economics*, 40(2):285–298.

Cummings, R. G., Elliott, S., Harrison, G. W., and Murphy, J. (1997). Are Hypothetical Referenda Incentive Compatible? *Journal of Political Economy*, 105(3):609–621.

Doyon, M. and Bergeron, S. (2016). Understanding Strategic Behavior and Its Contribution to Hypothetical Bias When Eliciting Values for a Private Good. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 64(4):653–666.

Espinosa-Goded, M., Barreiro-Hurl, J., and Ruto, E. (2010). What Do Farmers Want From Agri-Environmental Scheme Design? A Choice Experiment Approach. *Journal of Agricultural Economics*, 61(2):259–273.

Freeman III, A. M., Herriges, J. A., and Kling, C. L. (2014). *The Measurement of Environmental and Resource Values: Theory and Methods*. RFF Press, New York, New York, 3rd edition.

Haab, T., Huang, J., and Whitehead, J. (1999). Are hypothetical referenda incentive compatible? A comment. *Journal of Political Economy*, 107(1):186–196.

Haab, T. and McConnell, K. E. (2002). *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*. Edward Elgar Publishing.

Hanemann, W. M. (1991). Willingness to pay and willingness to accept: How much can they differ? *The American Economic Review*, 81(3):635–647.

Harrison, G. W. (2006). Making choice studies incentive compatible. In Kanninen, B. J., editor, *Valuing Environmental Amenities Using Stated Choice Studies*, number 8 in The Economics of Non-Market Goods and Resources, pages 67–110. Springer Netherlands.

Harrison, G. W. (2007). House money effects in public good experiments: Comment. *Experimental Economics*, 10(4):429–437.

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

Horne, P. (2006). Forest owners acceptance of incentive based policy instruments in forest biodiversity conservation a choice experiment based approach. *Silva Fennica*, 40(1):169–178.

Interis, M. (2014). A challenge to three widely held ideas in environmental valuation. *Journal of Agricultural and Applied Economics*, 46(3):347–356.

Johnston, R. J., Boyle, K. J., Adamowicz, W. V., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., and Vossler, C. A. (2017). Contemporary Guidance for Stated Preference Studies. *Journal of the Association of Environmental and Resource Economists*, 4(2):319–405.

Kaczan, D., Swallow, B. M., and Adamowicz, W. L. V. (2013). Designing a payments for ecosystem services (PES) program to reduce deforestation in Tanzania: An assessment of payment approaches. *Ecological Economics*, 95:20–30.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–91.

Kanjilal, M. (2015). *Costs of Wetland Restoration in the Canadian Prairies: Comparison of Simulated Opportunity Costs and farmers' WTA*. PhD thesis, University of Alberta; Department of Resource Economics and Environmental Sociology.

Keeler, J. P., James, W. L., and Abdel-Ghany, M. (1985). The relative size of windfall income and the permanent income hypothesis. *Journal of Business & Economic Statistics*, 3(3):209–215.

Knetsch, J. L. (2010). Values of gains and losses: Reference states and choice of measure. *Environmental and Resource Economics*, 46(2):179–188.

Krishna, V. V., Drucker, A. G., Pascual, U., Raghu, P. T., and King, E. D. I. O. (2013). Estimating compensation payments for on-farm conservation of agricultural biodiversity in developing countries. *Ecological Economics*, 87:110–123.

Latacz-Lohmann, U. and Schiizzi, S. (2005). Auctions for Conservation Contracts: A Review of the Theoretical and Empirical Literature (Project No: UKL/001/05). Report to the Scottish Executive Environment and Rural Affairs Department: Project No: UKL/001/05.

Latacz-Lohmann, U. and Van der Hamsvoort, C. (1997). Auctioning Conservation Contracts: A Theoretical Analysis and an Application. *American Journal of Agricultural Economics*, 79(2):407–418.

Liu, C.-C., Herriges, J. A., Kling, C. L., Secchi, S., Nassauer, J. I., and Phaneuf, D. J. (2014). A comparison of value elicitation question formats in multiple-good contingent valuation. *Frontiers of Economics in China*, 9(1):85–108.

Lusk, J. L., McLaughlin, L., and Jaeger, S. R. (2007). Strategy and Response to Purchase Intention Questions. *Marketing Letters*, 18(1/2):31–44.

Lusk, J. L. and Shogren, J. F. (2008). *Experimental Auctions: Methods and Applications in Economic and Marketing Research*. Cambridge University Press.

Mitchell, R. and Carson, R. T. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*. RFF Press.

Porras, I. and Hope, R. (2005). Using stated choice methods in the design of payments for environmental services schemes.

Southgate, D., Haab, T., Lundine, J., and Rodrguez, F. (2010). Payments for environmental services and rural livelihood strategies in Ecuador and Guatemala. *Environment and Development Economics*, 15(01):21–37.

Thaler, R. H. and Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, 36(6):643–660.

Tuncel, T. and Hammitt, J. K. (2014). A new meta-analysis on the WTP/WTA disparity. *Journal of Environmental Economics and Management*, 68(1):175–187.

Villanueva, A. J., Glenk, K., and Rodrguez-Entrena, M. (2017). Protest Responses and Willingness to Accept: Ecosystem Services Providers Preferences towards Incentive-Based Schemes. *Journal of Agricultural Economics*, pages n/a–n/a.

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

Wang, H. and Whittington, D. (2005). Measuring individuals' valuation distributions using a stochastic payment card approach. *Ecological Economics*, 55(2):143–154.

Whittington, D., Adamowicz, W., and Lloyd-Smith, P. (2017). Asking Willingness-to-Accept Questions in Stated Preference Surveys: A Review and Research Agenda. *Annual Review of Resource Economics*, 9(1):null.

Whittington, D. and Pagiola, S. (2012). Using contingent valuation in the design of payments for environmental services mechanisms: A review and assessment. *World Bank Research Observer*, 27(2):261–287.

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