Environ Resource Econ (2009) 44:307–326
DOI 10.1007/s10640-009-9287-3

Testing the Effectiveness of Certainty Scales, Cheap Talk, and Dissonance-Minimization in Reducing Hypothetical Bias in Contingent Valuation Studies

Mark Morrison · Thomas C. Brown

Accepted: 26 March 2009 / Published online: 3 June 2009

Abstract Stated preference methods such as contingent valuation and choice modeling are subject to various biases that may lead to differences between actual and hypothetical willingness to pay. Cheap talk, follow-up certainty scales, and dissonance minimization are three techniques for reducing this hypothetical bias. Cheap talk and certainty scales have received considerable attention in the literature, but dissonance minimization has not previously been experimentally tested. Using a four-way split sample design involving over 600 subjects, results from an actual referendum on provision of a quasi-public good were compared with three similar but contingent referenda employing the three bias-reducing techniques. Hypothetical bias was again present. Certainty scales, when properly calibrated, and dissonance minimization were found to be most effective in reducing the bias.

Keywords Contingent valuation · Hypothetical bias · Cheap talk · Certainty scales · Dissonance

1 Introduction

Over-estimation of willingness to pay has been a longstanding and important concern among practitioners of stated preference methods (Mitchell and Carson 1989), and is a manifestation of a much older concern about hypothetical bias (see for example LaPiere 1934). After reviewing evidence about the relationship between actual and hypothetical willingness to pay, the NOAA expert panel on contingent valuation concluded that "hypothetical markets tend to overstate willingness to pay for private as well as public goods" (Arrow et al. 1993, p. 4610). The existence of this hypothetical bias has since been verifie in many studies comparing

M. Morrison (✉)
Institute for Land, Water and Society, Charles Sturt University, Panorama Avenue, Bathurst, NSW 2795, Australia
e-mail: mmorrison@csu.edu.au

T. C. Brown
Rocky Mountain Research Station, U.S. Forest Service, Fort Collins, CO, USA

contingent to actual willingness to pay, and was recently summarized in meta-analyses by Little and Berrens (2004) and Murphy et al. (2005a). Little and Berrens took ratios from 42 studies; the average ratio was 3.13.[1] Murphy et al. reported several summaries of ratios; the summary that is most similar in methodology to that of Little and Berrens produced a mean ratio of 3.26 (with a median of 1.50).[2] In both meta-analyses the effects of various study design factors on the magnitude of the ratio were examined, including methods for avoiding hypothetical bias. Both studies found that these methods can be effective, although the evidence does indicate that they differ in their effectiveness.

Three main approaches have been used in the literature to avoid or adjust for hypothetical bias in estimating willingness to pay, with all three approaches appearing in the literature at approximately the same time.[3] The firs approach involves the use of follow-up certainly scales, which allow respondents to indicate how sure they are that they would actually pay the amount, or vote as, they just indicated they would. There are two basic versions of this approach.[4] The firs uses a numerical scale with labelled end points. For example, Li and Mattsson (1995) used a scale from 0 to 100% in 5% increments where 0% was labelled "absolutely uncertain" and 100% was labelled "absolutely certain", and Champ et al. (1997) used a 10-point rating scale with 1 labelled "very uncertain" and 10 labelled "very certain". The other version of the certainty scale approach offers two or more discrete options, each describing a level of certainty.[5] For example, Blumenschein et al. (1998) used a two-category qualitative scale that allowed respondents to indicate whether they were "probably sure" or "definitely sure" about their prior response. Similarly, Johannesson et al. (1998) used the categories "fairly sure" and "absolutely sure". With both versions, the certainty scale responses are used to switch some of the positive responses to negative.

These studies, and Little and Berrens in their meta-analysis, show that use of follow-up certainty scales can be effective at eliminating the difference between actual and hypothetical mean willingness to pay. The two versions each have their advantages. The advantage of the quantitative scale is that it offers respondents a fl xible way to indicate their certainty, as only the end points are labelled and there are various options between the end points. This fl xibility, however, comes at a cost in that the scale must be calibrated to the situation at hand. Studies using 10-point rating scales have shown that the appropriate certainty cut-off point, below which a yes response is best recoded to no, differs. Champ et al. (1997) and Blumenschein et al. (2001) found that including only those who circled 10 on their certainty

[1] Little and Berrens (2004) updated and extended the meta-analysis of List and Gallet (2001). Most of the 42 studies Little and Berrens examined reported more than one ratio. To avoid over-emphasizing a given study, the authors compiled the median ratio for each study and then computed the mean across the 42 medians. Note that a few of the ratios were for willingness to accept compensation rather than willingness to pay.

[2] Both meta-analyses include some ratios from studies that used methods to reduce hypothetical bias, such as certainty scales or cheap talk. Use of the methods, of course, lowered the numerators and therefore the ratios. When Murphy et al. (2005a) excluded from the sample the ratios resulting from bias-reducing methods, the mean ratio rose to 5.42, with a median of 2.66. Note that Murphy et al. do not include ratios from willingness to accept compensation bids.

[3] Some authors call these approaches "calibration" techniques, in the sense that they are used to calibrate willingness to pay. However, "calibration" is typically used in reference to a method or model (i.e., the method is calibrated so that it accurately estimates the value of the variable under investigation). We use calibration in this more traditional way, and speak of calibrating the methods used to avoid hypothetical bias so that they provide accurate estimates of actual willingness to pay.

[4] Ready et al. (1995) offer a completely different way of incorporating uncertainty, replacing the standard dichotomous choice (yes/no) format with a six-part response format with the labels definitely yes, probably yes, maybe yes, and then three similar categories for no.

[5] Ready et al. (2001) tried a hybrid of these two versions in the form of six fi ed categories some of which included quantifie (i.e., 95% sure) estimates of certainty.

scale as positive responses and all others as negative produced a mean willingness to pay that was equivalent to actual willingness to pay. In contrast, Ethier et al. (2000) and Poe et al. (2002) found that a cut-off of 7, and Champ and Bishop (2001) and Norwood (2005) found that a cut-off of 8, was needed to produce equivalent results. Thus, the selection of a cut-off remains somewhat arbitrary, although further studies may demonstrate the most commonly occurring point of calibration or devise a way to choose the appropriate point for a given context.

The other version of the certainty scale avoids the need for calibration, but commits to distinct categories. In three studies using private goods, Blumenschein et al. (1998), Blumenschein et al. (2001) and Blumenschein et al. (2008) found that when the "probably sure" yes responses were recoded to no, leaving only the "definitely sure" yes responses as yes, the null hypothesis of no difference between the corrected hypothetical treatment and the real payment treatment could not be rejected. However, in the Johannesson et al. (1998) study cited above, hypothetical bias was absent from the start and recoding the "fairly sure" yes responses to no reduced the percentages of yes responses significantly below the percentages in the real payment treatment at most bid levels. Thus it seems that more study is needed, including study with public goods, before we can have confidenc in the fi ed category version of the certainty scale approach.

The second approach that has been used to reduce hypothetical bias, now known as cheap talk, involves use of an entreaty to the respondent prior to presenting the valuation question. The "cheap talk" label was borrowed from experimental economics where it refers to communication between players prior to execution of an experiment. Here cheap talk refers to communication from the experimenter to the participant about things to consider when responding to a subsequent question.[6]

The firs tests of cheap talk (Cummings et al. 1995b; Loomis et al. 1996) used rather short cheap talk scripts that failed to lower hypothetical WTP sufficientl to match actual WTP. Cummings and Taylor (1999) then tried a more lengthy script, which takes about 5 min to read aloud. It has several elements including identificatio of the problem of hypothetical bias, a numerical example from a previous study demonstrating hypothetical bias, explanation of why hypothetical bias occurs, and a fina entreaty to consider opportunity costs and avoid hypothetical bias. Cummings and Taylor found this longer script to be surprisingly effective at reducing hypothetical bias in experiments using public good referenda with student subjects, although only one bid level ($10) was tested.

Cummings and Taylor's (1999) paper stimulated several other tests of cheap talk scripts. Brown et al. (2003), in another public good referendum study with student subjects, tested the Cummings and Taylor (1999) cheap talk script at several bid levels. They found that the script was effective at an $8 bid level, but was less effective or ineffective at lower bid levels.[7] Both List (2001), who used the full (Cummings and Taylor 1999) script in a second-price auction for a private good, and Lusk (2003), who used a shorter but largely similar script in a dichotomous choice mail survey valuing a private good, found that cheap talk lowered bids for inexperienced consumers but not for experienced ones. Aadland and Caplan (2003), using a very short script in a phone survey, also found evidence that people differ in their susceptibility to cheap talk. Murphy et al. (2005b), using the full script in an experiment eliciting individual donations subject to a provision point, found that cheap talk was quite

[6] We defin cheap talk as an entreaty to avoid a specifie hypothetical bias (typically, over-estimation of willingness to pay). The term has also been interpreted more broadly. For example, Aadland and Caplan (2006) use a script they call cheap talk that warns generally of bias without suggesting the direction of the bias.

[7] See also Ajzen et al. (2004). Aadland et al. (2007) offer an interesting explanation of this finding one that emphasizes anchoring and Bayesian updating in the face of preference uncertainty.

effective at reducing hypothetical bias at bid levels of $9 and higher but less effective at lower bid levels. Blumenschein et al. (2008), using a script similar to Cummings and Taylor's but without the numerical example, found that cheap talk was completely ineffective at reducing hypothetical bias for a private good in the health industry. Finally, List et al. (2006), using a shorter but similar script to Cummings and Taylor's in a choice study conducted at a sports card show, found that cheap talk removed bias for both experienced (dealers) and inexperienced (nondealers) traders.[8] Importantly, however, they found that subjects in the cheap talk treatment were less consistent in their choices than subjects in either the actual or the hypothetical treatment, suggesting that cheap talk may introduce new biases. In summary, the evidence indicates that cheap talk varies in its effectiveness, but is most effective when the script presents a compelling case for avoiding hypothetical bias, for public goods where respondents are relatively inexperienced with the good being valued, and for moderate to high bid levels—all contexts where hypothetical bias is likely to be greatest.

The third approach, available for use with dichotomous choice or referendum contingent valuation, presents the respondent with response categories beyond the simple yes and no responses. First used by Blamey et al. (1999) and Loomis et al. (1999), this approach, called "dissonance minimising" by Blamey et al., is based on the view that hypothetical bias or yea-saying results from cognitive dissonance, which is an "emotional state set up when two simultaneously held attitudes or cognitions are inconsistent or when there is a conflic between belief and overt behaviour" (Reber 1985, p. 129). As suggested by Brown et al. (1996), the standard dichotomous choice contingent valuation format may place respondents in the awkward position of choosing between two competing objectives: honestly responding to the bid level and indicating whether or not they favour provision of the good.[9] Respondents who favour the good but suspect that they would not pay as much as they are asked about must choose which objective to pursue, and may choose the latter. The dissonance minimizing (DM) format adds response categories that permit respondents to express support for a program without having to vote in favour of increased expenditure. For example, in the Blamey et al. (1999) study the extra response categories included "I support the [program] … but it's not worth $50 to me", "I support the [program] … but I cannot afford $50", and "I support the [program] but not if it requires a [fee] of any amount". By decoupling the choice of whether or not to support the environmental program from the commitment of dollars, this questioning format attempts to reduce dissonance and hypothetical bias.

The results presented by Blamey et al. (1999) and Loomis et al. (1999) were promising. Compared to the hypothetical dichotomous choice results, the DM format resulted in a much steeper bid curve, and there was greater evidence of construct validity. However, neither study included an actual payment treatment; therefore the extent to which this approach was successful in addressing hypothetical bias could not be ascertained.

The purpose of this paper is to investigate the relative effectiveness of these three approaches to reducing hypothetical bias. Using experimental economics procedures in a split sample design, groups of participants voted in actual or contingent payment referenda about provision of a quasi-public good. The four treatments—actual payment, hypothetical with follow-up certainty scale, hypothetical with cheap talk, and hypothetical with DM format—allow for fi e separate estimates of willingness to pay, as the hypothetical with certainty

[8] In another choice study, a mail survey of demand for two separate food items, Carlsson et al. (2005) used a very short script and found that cheap talk significantly lowered willingness to pay for seven of ten attributes. Effectiveness could not be tested, however, because a real payment treatment was not included.

[9] These competing objectives are in addition to those inherent in any dichotomous choice response, which plays the objective of obtaining the good against the objective of preserving disposable income (Ready et al. 1995).

scale treatment provides two estimates, one without the certainty correction and one with the correction.

In testing for hypothetical bias we acknowledge the perspective of Murphy and Stevens (2004) who noted that with homegrown values it is often not possible to know whether preferences obtained when respondents are required to pay represent true or "real" preferences. It may be that respondents in the actual payment treatment are understating their willingness to pay, which would invalidate the comparison with hypothetical willingness to pay. In our study, this might happen for at least two reasons. First, because we use an open referendum (i.e., the good can also be provided by people not taking part in the experimental referendum), free-riding is a possibility. We do however note the evidence provided by Taylor (1998), which demonstrated that moving from an open to a closed referendum did not significantly affect value estimates. That is, removing the possibility of free-riding (via a closed referendum) did not significantly alter willingness to pay. This findin is supportive of the validity of using actual willingness to pay in an open referendum as a reference point for determining the extent of hypothetical bias. Second, there is the possibility that some participants in the actual payment referendum who are inclined to otherwise vote yes may nevertheless vote no in order to avoid potentially forcing others to pay.[10] Because such behavior, if it were to occur in the actual payment condition, may also occur in the hypothetical conditions, this second possibility is perhaps not a significant concern.[11] However, this possibility remains a concern to the extent that such behavior is more salient in an actual than in a hypothetical referendum. Further tests are needed of both of these potential behaviors.

2 Description of Experiments

Students participated in sessions that lasted approximately half an hour, for which they were each paid 20 Australian dollars (at the time of the study, A$1 was equivalent to about US$0.70). The sessions had three steps: a trade-off experiment unrelated to the current topic; a questionnaire regarding participants' socio-demographics and general attitudes; and a referendum on provision of a quasi-public good, school childrens' breakfasts provided through the Red Cross.[12] The firs two parts served to help mitigate found-money effects on the referendum to follow. In the referendum step, the students participated in one of four treatments: actual payment, hypothetical payment followed by a certainty scale, hypothetical payment with cheap talk, and hypothetical payment with a DM format. The students each had a copy of the questionnaire and followed along as the moderator read the instructions aloud.

The questionnaire for the DM treatment is presented in the Appendix; the other questionnaires are identical to the DM questionnaire except for the differences described below. The questionnaires began with a brief outline of what the experiment would involve, followed by a description of the problem—children missing breakfast, particularly in disadvantaged areas, and the consequences thereof. A solution to this problem—the provision of children's breakfasts by the Red Cross—was then described. A justificatio for collecting funds through a referendum was also provided, as follows:

[10] This is a case of what Madariaga and McConnell (1987) call non-paternalistic or individualistic altruism.

[11] Subjects in the hypothetical referenda are asked to respond "as if we were actually giving you the opportunity to contribute".

[12] This trade-off experiment involved subjects trading off between various consumers products (e.g., chocolate bars, pens and visors for cars) to enable assessment of endowment effects. After this experiment and prior to the experiment reported in this paper, students completed a questionnaire that included a series of attitudinal and socio-demographic questions.

> Because the Red Cross relies on donations, the number of meals it can provide to children each year through the Breakfast Club depends on the funds received. If more money were given, the Red Cross would be able to provide meals to needy children who currently are not receiving these meals. So in a moment we are going to ask if you would like, as a group, to contribute money towards the Breakfast Club initiative.

Subjects were then told that if they wanted additional breakfasts to be provided they would need to make a donation. To provide balance to the information provided earlier, the subjects were also given reasons why they might not want to donate and were reminded of their budget constraint.

They were then given an explanation of how the referendum in the experiment would work, as follows (for the $15 hypothetical treatments):

> After you have each voted about whether you want to contribute towards providing a Red Cross breakfast, one of us will collect your voting slips.
>
> (1) If more than 50% of you had voted YES to this proposal, all of you would—if this experiment were for real—pay $15. I would have deducted $15 from the money that each of you earned for participating in these experiments and a cheque for this amount ($15 x ____ people = $_____) would be sent to the Australian Red Cross, who would send each of you a receipt for your contribution.
> (2) If 50% or fewer of you had voted YES on this proposal, no one would pay $15, and no money would be sent to the Australian Red Cross.

This description of how the referendum would work was used in each of the hypothetical treatments. In the actual payment treatment a similar description was used, but without the subjunctive language of the hypothetical treatments.

Next, for each treatment except cheap talk a short review was followed by the referendum question. In the cheap talk treatment the full Cummings and Taylor (1999) cheap talk script, complete with numerical example, was presented prior to the short review. We made only minor modification to the Cummings and Taylor script to allow for differences in the good that was valued.

For each of the treatments except that using the DM format, the referendum question had two response categories (vote yes or vote no). The DM format had four additional categories (the second through to the fift categories), as follows[13]:

- I would vote YES to this proposal that everyone contribute $15 to the Breakfast Club.
- I support the goal of the Breakfast Club, but I'm not prepared to pay $15 and thus would vote NO.
- I support the goal of the Red Cross Breakfast Club, but I cannot afford $15 and thus would vote NO.
- I support the goal of the Red Cross Breakfast Club, but I would prefer to save my money and contribute to another cause and thus would vote NO.
- I support the goal of the Breakfast Club initiative, but I would vote NO for the following reason_____.
- I would vote NO to this proposal that everyone contribute $15 to the Breakfast Club.

Loomis et al. (1999) used only three categories, the standard yes and no and a third stating that "I would vote for program _____ only if the cost to my household were less than $C per

[13] When modelling the data, each of the categories other than the yes vote (i.e., categories 2–6) were treated as no votes.

year." Blamey et al. (1999) used fi e categories, which subsume the three of the Loomis et al. study. Three of the Blamey et al. categories are similar to the firs three above, except for changes to reflec differences in the good and payment vehicle. We passed over Blamey et al.'s last two categories (support program but not if it requires a payment of any amount; do not support program at any cost) because they seemed either redundant with prior categories or inappropriate with the current good, and added the last three categories above, two of which (the fourth and fifth are an attempt to offer a more comprehensive list of reasons for why one might wish to vote no.

After the referendum question, in each of the hypothetical treatments a certainty question was asked. Similar to the majority of previous studies that have used certainty scales, a ten-point certainty scale was used, with the labels "very uncertain" and "very certain" positioned at either end of the scale.

Four focus groups were held to refin the draft questionnaires for each of the four treatments. The focus groups provided some initial qualitative insights about the likely effects of the cheap talk and DM treatments in reducing hypothetical bias. For the cheap talk treatment, some participants indicated that they thought it helped them to give more accurate responses, as in the following three examples: "Slowed down the decision—made me think a bit longer"; "Made me think how I'd really spend it"; and "I think it is a valid point, I changed my answers … would probably lead to more accurate answers". However, other participants thought that they were being inappropriately persuaded to vote no, as in the following two examples: "Seemed as if it were a persuasion to say no … If you're not saying no you're not being honest"; and "The part where it says if you vote yes you're being a good person … and then if I say no does that mean I'm a bad person?" Participants also indicated that the cheap talk script tended to devalue the good in addition to encourage them to answer realistically.

Participants generally reacted favourably to the DM format. The reduction in dissonance is evident in the following two quotes: "If you want people to say no if they really mean no it's giving them an option where they can say no and still feel good about themselves"; and "A relief … I can say why I'm not [going to vote yes] … it removes some of the guilt so you can give an honest answer". Participants did, however, have some concerns about the additional categories in the DM treatment. For example, they objected to the phrase in "but it's not worth $X to me" used in the Blamey et al. study. Participants thought the phrase was value laden and preferred use of the phrase "but I'm not prepared to pay $X", which was adopted in the fina questionnaire, as noted above.

3 The Sample

Students from four campuses and three faculties of Charles Sturt University, a regional university in New South Wales, participated in the experiments presented in this paper. Experiments were conducted at the end of classes, and student groups were randomly sampled until sufficien numbers were collected for each cell. A total of 610 undergraduate students were included in the fina sample. Students studying education were excluded from the sample because of the nature of the good. Students who were working full-time were also excluded to increase the homogeneity of the sample, although this did not influenc the fina results as indicated by regression analysis. The sample sizes for each of the treatments at each of the bid levels ($10, $15, and $20) are shown in Table 1.[14]

[14] For comparison purposes, in Australia a McValue Meal at McDonalds costs about $7, an adult ticket to a movie cinema costs about $12, a music CD costs about $20, and two home delivered pizzas cost about $25.

Table 1 Sample sizes

Treatment	Bid level			Total
	$10	$15	$20	
Actual payment	47	50	41	138
Hypothetical with certainty scale	47	47	57	151
Hypothetical with cheap talk	67	47	33	147
Hypothetical with DM format	61	59	54	174

Table 2 Socio-demographic and attitudinal characteristics of subjects in each treatment, by bid level

	Actual payment			H-certainty scale		
	$10	$15	$20	$10	$15	$20
% Female	51	40	63	45	45	32
% Work part-time	55	38	49	49	51	53
% Self-responsible	26	24	20	21	32	34
Interest	2.43	2.45	2.39	2.59	2.28	2.60
	H-cheap talk			H-DM format		
% Female	62	47	58	21	56	35
% Work part-time	49	40	58	47	34	61
% Self-responsible	22	21	15	31	36	27
Interest	2.51	2.45	2.45	2.65	2.48	2.59

Note: self-responsible refers to those who are responsible for their own tuition and living expenses, and interest is a 4-point rating scale question indicating respondents' interest in the Breakfast Club Program, where 1 is not at all interested and 4 is very interested

The socio-demographic and attitudinal characteristics of subjects in each of the treatments at each bid level are summarized in Table 2. The variables presented in this table are those that were found to be significant regressors in regression analysis that will be reported shortly. While responses to the attitudinal question INTEREST were relatively consistent across treatments, more substantial differences were found for the socio-demographic variables. In particular, the percentage of female students ranged from 21% to 63% while those working part-time ranged from 34% to 61%, indicating some socio-demographic heterogeneity across treatments.

4 Results

The percentages of respondents voting yes in the actual payment treatment, in the hypothetical treatment without correction, and with the three bias-reducing approaches are presented in Table 3 and Fig. 1. Three observations are apparent. First, the actual payment bid curve is relatively flat Log-linear modelling of the raw data indicates that the percentage of subjects supporting payment does not differ significantly across the bid levels ($\chi^2 = 4.467$, $p = 0.107$), though the bid variable is significant in one of our reported regression equations

Table 3 Percentage of subjects voting yes across treatments and bid levels

Bid level	Treatment				
	Actual payment	Hypothetical	H-cert7	H-cheap talk	H-DM format
$10	49	74	49	39	45
$15	46	57	43	36	41
$20	44	53	40	27	43
All	46	61	44	35	43

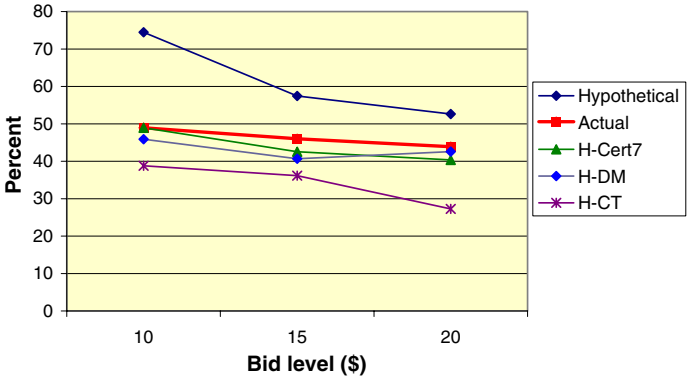

Fig. 1 Percentage of subjects voting yes across treatments and bid levels

(see Table 5). We expect that this has resulted from the good chosen for this experiment.[15] When selecting the good, students were surveyed to identify a public good that they would be most willing to pay for, and Red Cross children's breakfasts was the good indicated most often. It is possible that this insensitivity to cost reflect the importance of this public good to the subset of students willing to pay for its provision.

The second observation, which is consistent with many previous studies, is that the percentage of respondents voting yes in the hypothetical treatment without calibration (61% across all three bid levels) is greater than for the actual payment treatment (46%). The log-linear analysis indicated that the percentage of respondents voting yes is related to treatment ($\chi^2 = 20.904$, $p = 0.000$), and the existence of large adjusted residuals[16] for the hypothetical treatment in the log-linear analysis indicate differences for this treatment (see Table 4).

The third observation is that a certainty correction using a rating of 7 as the cut-off and the DM approach were both successful at eliminating hypothetical bias. Log-linear modelling

[15] An alternative explanation for the insensitivity to bid is that subjects' choices may indicate attitudes or value expression rather than preferences (Kahneman et al. 1999). However, we do not subscribe to this view as it is relevant for explaining bid insensitivity for hypothetical treatments. In our treatments we fin the opposite: that there is bid sensitivity for our hypothetical treatment, but not for our actual treatment.

[16] A residual is the difference between the observed and the expected frequency therefore large residuals indicate that yes responses were greater than expected. A standardised residual is the raw residual divided by the standard deviation of the observed counts. The adjusted residual is a modifie form of the standard residual, and under ideal circumstances are distributed normally. If normality occurs, adjusted residuals greater than 1.96 indicate a statistically significant effect at the 5% level. However, as these ideal conditions are not often met, Bakeman and Robinson (1994) recommend using a critical value of 2 for the 5% significance level.

Table 4 Adjusted residuals for treatments (excluding H-cert7) from the log-linear analysis (for vote = yes)

Bid level	Treatment			
	Actual payment	Hypothetical	H-cheap talk	H-DM format
Count	64	92	52	75
Expected count	64	70.1	68.2	80.7
Residual	0.0	4.1	−3.1	−1.0

results (small residuals) support this finding However, there is evidence that the use of the full cheap talk script has over-corrected, with the log-linear modelling also producing large residuals for this treatment.

To further test the significance of these observations we use regression analysis, which controls for differences in respondents' socio-demographic characteristics and attitudes. The results from two binary logit regressions are reported in Table 5. Pooled models are estimated with dummy variables used to capture treatment effects. Attempts were also made to interact the dummy variables with the cost variable, however as these interactions did not prove to be significant they have not been reported. The regressions have moderate explanatory power as indicated by the rho-square values. Rho-square is a pseudo-R^2 measure but it has a different metric, with values of 0.2–0.4 equivalent to values of 0.7–0.9 in linear regression (Louviere et al. 2000).

Contrary to the finding in the log-linear analysis, the coefficien for cost is significant at the 5% level in the firs regression equation. However, the coefficien for cost is not significant in the second regression equation, which includes the hypothetical treatment with the certainty correction treatment instead of the hypothetical treatment without calibration. This suggests that significance of the cost coefficien in the firs regression is due in part to the higher price sensitivity of the hypothetical treatment without calibration, and its replacement in the second equation with the certainty correction treatment—which was relatively insensitive to cost—has lead to the cost coefficien being insignificant.

In both equations several socio-demographic variables were found to be significant, which is indicative of construct validity. Two of these variables—SELF-RESPONSIBLE and WORK PART-TIME—are related to income. As expected, these variables indicate that if you are responsible for your own tuition and living expenses your willingness to pay will be lower, and if you work part-time your willingness to pay will be higher. The other two significant covariates are FEMALE and INTERESTED; these coefficient indicate that females and those who are interested in the good are more likely to vote in favour of the referendum.

The results from the regression analyses confir the observations reported earlier, namely that the DM and certainty-7 correction were effective at mitigating hypothetical bias. In both models the coefficien for the dummy variable representing the DM estimate is insignificant, indicating that there is no significant difference between the DM treatment and the actual payment (which is the missing category). The certainty-7 correction is included as a dummy variable in Model 2 in place of the dummy variable for the uncorrected hypothetical estimate in Model 1. The coefficien for this dummy variable is also insignificant, indicating that the use of the certainty-7 correction has dealt with the hypothetical bias.

The evidence from the regression analysis also confirm the observation that the use of the full cheap talk script has over-calibrated potential hypothetical bias. In both models the coefficien for the cheap talk dummy is significant at the 5% level. Of interest is the magnitude

Table 5 Regression models showing the effect of treatment, socio-demographic and attitudinal variables on voting behaviour

Variable	Model 1: all variables and data included	Model 2: CERT7 included instead of H
Cost	−0.054**	−0.036
	(0.023)	(0.023)
H-DM	−0.216	−0.205
	(0.253)	(0.252)
H-CT	−0.614**	−0.593**
	(0.260)	(0.259)
H	0.623**	
	(0.264)	
H-cert7		−0.162
		(0.261)
Self-responsible	−0.538***	−0.438**
	(0.214)	(0.212)
Work part-time	0.444**	0.508***
	(0.185)	(0.184)
Female	0.383**	0.431**
	(0.185)	(0.183)
Interest	0.608***	0.584***
	(0.114)	(0.113)
Constant	−1.499***	−1.829***
	(0.529)	(0.534)
Rho squared	0.095	0.073

Notes: standard errors are in parentheses
* Significant at 10% level, ** Significant at 5% level, *** Significant at 1% level
BID bid level, *H-DM* dissonance-minimizing treatment, *H-CT* cheap talk treatment, *H-Cert7* certainty-7 correction, *H* hypothetical treatment with no correction
Dependent variable in both models is vote (yes, no). For Model 2, the vote is recoded based on the certainty-7 correction
For both models the baseline responses are for actual payment treatment

of this coefficien in Model 1; it is about the same size as the coefficien for the hypothetical treatment without correction. The over-correction from cheap talk is about equal in magnitude to the hypothetical bias.

Regarding the certainty scale, it is of interest to examine the effects of various cut-offs. As Fig. 2 shows, using a cut-off of 8 (i.e., recoding as a no response those yes respondents with a certainty response of 7 or lower) will under-estimate the actual percentage voting yes. Using a cut-off of 6 or 7 yields a reasonable approximation of the actual payment percentages, with the cut-off of 7 slightly out-performing the cut-off of 6.

Using the results from Model 1, mean willingness to pay was estimated for each of the treatments, except for certainty 7 (Table 6). Mean willingness to pay was calculated using the approach recommended by Haab and McConnell (2002, p. 34) for use with a linear random utility model. That is, $E_\varepsilon(WTP_j|\alpha, \beta, z_j) = \frac{\alpha z_j}{\beta}$, where z_j is a vector of characteristics related to individual j, α is vector of parameters and β is the coefficien of the cost parameter.

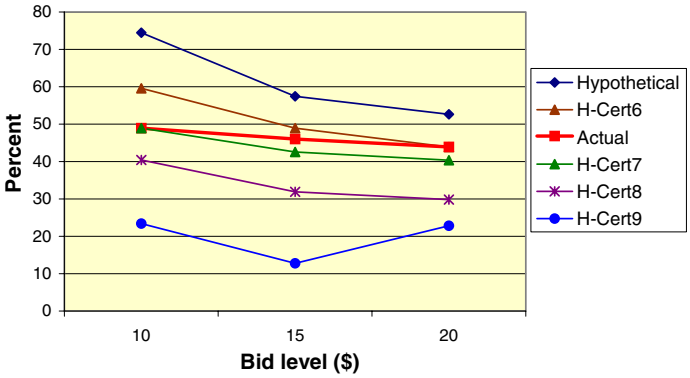

Fig. 2 Percentage of subjects voting yes at various certainty scale cut-offs

Table 6 Mean willingness to pay (WTP) across treatments

	Treatment			
	Actual payment	Hypothetical	H-cheap talk	H-DM format
Willingness to pay estimated using logit model				
Mean WTP	$12.34	$23.92	$0.92	$8.33
Ratio relative to actual payment	–	1.94	0.07	0.67
Willingness to pay estimated using Turnbull estimator				
Mean WTP	$9.39	$12.95	$7.05	$8.75
Ratio relative to actual payment		1.38	0.75	0.93

Mean willingness to pay of the hypothetical treatment without correction exceeded the actual payment mean by a ratio of 1.94, which is substantially lower than the mean ratio reported by both Little and Berrens and Murphy et al. in their meta-analyses, but higher than the median of 1.50 reported by Murphy et al. The DM treatment produced a slightly lower mean bid than that of the actual payment treatment, but the difference was not significant. Finally, the CT treatment produced a substantially lower mean bid than that of the actual payment treatment. This low bid for the CT treatment results from the formula used to calculate mean willingness to pay, which is the same formula as that used to calculate median willingness to pay. As only about 40% of respondents voted in favour of supporting the referendum at $10, a low median (and mean) willingness to pay is implied.

Given the relatively disparate willingness to pay estimates produced using the results from the binary logit model, willingness to pay was also calculated using a non-parametric Turnbull estimator (Table 6). The ratio of hypothetical to actual willingness to pay (1.38) was lower when using this non-parametric approach. The over-correction with the cheap talk treatment was also lower, with the cheap talk treatment producing a mean willingness to pay that was 75% of the willingness to pay in the actual payment treatment.[17]

[17] Mean willingness to pay for the CT treatment is higher when using the Turnbull estimator because the entire density of the bid-response function is used to calculate willingness to pay, in contrast to the parametric approach, which is influence by the willingness to pay of the median respondent.

Table 7 Distribution of responses across categories in the H-DM treatment

Response category	Bid level		
	$10	$15	$20
I would vote YES to this proposal that everyone contribute $X to the Breakfast Club.	28	24	23
I support the goal of the Breakfast Club, but I'm not prepared to pay $X and thus would vote NO.	5	3	2
I support the goal of the Red Cross Breakfast Club, but I cannot afford $15 and thus would vote NO.	11	15	15
I support the goal of the Red Cross Breakfast Club, but I would prefer to save my money and contribute to another cause and thus would vote NO.	9	12	10
I support the goal of the Breakfast Club initiative, but I would vote NO for the following reason _____	3	4	0
I would vote NO to this proposal that everyone contribute $X to the Breakfast Club.	6	1	4
Total	62	59	54

A fina question to consider is whether all four additional categories were needed for the DM treatment. The distribution of responses across each of the additional categories is shown in Table 7. The most commonly chosen DM categories were the second and third ones, that is "I support the goal of the Red Cross Breakfast Club, but I cannot afford $15 and thus would vote NO" and "I support the goal of the Red Cross Breakfast Club, but I would prefer to save my money and contribute to another cause and thus would vote NO". The remaining two categories, particularly the fill-in-the-blan category ("I support the goal of the Breakfast Club initiative, but I would vote NO for the following reason _____") were much less frequently chosen and could potentially be dropped without affecting the effica y of this approach.

5 Discussion

This study offers promising results about the ability to adjust for hypothetical bias. First, the DM approach as implemented was effective in mitigating hypothetical bias. There was also some evidence that less than four additional categories may be needed when using the DM approach. Compared to both certainty scales and cheap talk, this approach has received limited attention in the literature. The finding from this study suggest that more research is warranted to examine the generailizability of the DM result across alternative public goods, across cultures, and in fiel trials. Second, further evidence has been provided about the effectiveness of follow-up certainty scales. The findin that a cut-off rating of 7 produces estimates that match actual payments in this study adds to the majority of studies that have found a cut-off of either 7 or 8 is needed to equalise hypothetical and actual willingness to pay.

Unlike the other two approaches, cheap talk over-corrected for hypothetical bias. This was unexpected, as the Cummings and Taylor (1999) script when used with public goods has generally performed well at higher bid levels and been at worst ineffective at lower bid levels. The consistent over-correction in this case may reflec cultural differences between the United States and Australia in terms of tendencies for yea-saying behaviour. Indeed the

finding of Sinden (1988), the only other study conducted in Australia that compared hypothetical and actual willingness to pay, suggests that hypothetical bias may be less pervasive in Australia than in the United States. Possibly a shorter and weaker entreaty would have been more effective at correcting for hypothetical bias. However, the apparent sensitivity of cheap talk to cultural changes, to the experience trading or knowledge about the good, and even to bid levels, and the long length of the most successful scripts, suggest that other approaches such as the use of certainty scales and DM formats may offer practitioners more reliable and practical means for reducing hypothetical bias.

That said, it is clear that all three approaches can eliminate hypothetical bias if properly calibrated and worded for the given context, and that each approach has its shortcomings and strong points. Summarizing the shortcomings, (1) cheap talk and categorical certainty scales have been found to over-correct in some situations, such as when no bias is present, and under-correct in others, such as with experienced consumers; (2) the DM format cannot be used with open-ended responses; (3) quantitative certainty scales require selection of the cut-off point, and the appropriate cut-off seems to differ across contexts; and (4) cheap talk, categorical certainty scales, and the DM format may all need some calibration in the form of wording changes to suit a given context.

We now take a step back and ask what this and related studies tell us about hypothetical bias. Although there have been many tests for hypothetical bias and quite a few successful attempts to make it go away, there has been relatively little effort to understand the reason for hypothetical bias, at least as it occurs with monetary bidding. One study that did attempt to explain hypothetical bias and to focus on individual differences, by Ajzen et al. (2004), asked subjects, in the context of public good referenda, a substantial set of questions about their attitudes and beliefs regarding the good (scholarships for needy students) and paying for the good.[18] They posited two hypotheses as possible explanations for hypothetical bias and then used responses to the attitude and belief questions to test the hypotheses. The firs hypothesis was that hypothetical bias is simply due to the subset of respondents who are on the fence, for whom the less demanding hypothetical context allows a positive response when the more demanding actual payment context does not do so. The other hypothesis is that the hypothetical and real payment contexts are fundamentally different, generating quite different beliefs and attitudes about the good and its provision, with the hypothetical payment context generating more positive beliefs and attitudes about the good and about paying for the good than the real payment context. Examining responses to extensive questionnaires fille in by the bidding participants, Ajzen et al. concluded that the latter hypothesis was more likely to be the explanation for the hypothetical bias they observed. Specifically, they found that subjects' beliefs about what others (family members and friends) would think if the subjects spent their money to support the scholarship effort—beliefs that were more negative in the actual payment condition than in the hypothetical payment condition—seemed to most clearly separate the two groups.[19]

[18] Another exception is the work of Champ et al. (1997) and Champ and Bishop (2001). Comparing attitudes expressed by respondents in an actual payment treatment with attitudes of respondents in a hypothetical payment treatment, they found that respondents to the hypothetical payment question who indicated high certainty on a quantitative follow-up certainty scale were quite similar to respondents in the other treatment who actually paid for the good, and dissimilar to those of lower certainty.

[19] The Ajzen et al. study was complicated by the fact that they examined bidding for a public good, where a complex combination of motives or behaviors may explain the observed bias. Studies similar to Ajzen et al.'s are needed that focus on private goods, where the possible reasons for the bias are fewer and therefore easier to detect.

We suggest that these two hypotheses are not mutually exclusive, and specifically that the difference in beliefs that separates the hypothetical from the real payment condition only matters for those people who are ambivalent in the hypothetical payment situation. We come to this conclusion by reviewing the evidence from induced value and private good studies. First, because hypothetical bias is absent in induced value experiments (Taylor et al. 2001; Vossler and McKee 2006; Mitani and Flores 2007), where the value of the good is in monetary terms (i.e., is precise and quantitative), we conclude that hypothetical bias is not a habitual and unavoidable response, and suggest that hypothetical bias may be related to the lack of precision introduced when the good is not simply provided at a given monetary amount. Second, because the bias often occurs for private goods, we must conclude that public-spirited motives or attempts to encourage provision of the good (which may play a role in hypothetical bias observed in public good valuations) are not a necessary condition for bias to occur.[20] Thus, we suspect that the bias is related to uncertainty about one's willingness to pay for the good.[21] Some subjects in bidding experiments act as if they have only a vague notion of their willingness to pay for the good, as if their willingness to pay were represented by a range rather than a point. This is not a new idea; Opaluch and Segerson (1989), Dubourg et al. (1997), Gregory et al. (1995), Ready et al. (1995), Wang (1997), Murphy and Stevens (2004), and undoubtedly others have suggested that imprecision of preference can be characterized by a preference range. As an example of the existence of this preference range, Loomis and Ekstrand (1998) found that respondents were more uncertain about their votes in the mid level bid range, but had greater certainty at high and low bids. Using this range metaphor for the ambivalence across various prices that we observe, if the posited price falls above the range, a no response to a dichotomous choice bid question is forthcoming, but if the posited price falls within the range, subjects asked a hypothetical question have little or no incentive to do other than say yes. It is only when they are faced with the possibility of a real economic commitment that subjects are forced to consider the risk of paying too much (and what they and others would think of them paying so much) for the good, leading them to narrow the range and perhaps say no.

Valuation of public goods introduces another possible cause, or set of related causes, of hypothetical bias—that respondents may wish to express support for the good, may wish to be seen as or to think of themselves as public-spirited, or may wish to encourage provision of the good (to the extent that they think their response may make provision more likely, and especially if they think they will escape payment). The success of the DM treatment that we report here suggests that such motives exist.

In summary, we tested three methods for reducing hypothetical bias in referendum contingent valuation and found considerable success, but our efforts contribute only modestly to understanding the underlying causes of hypothetical bias. Not content to leave it there, we examined the work of others, especially induced value and private good experiments, in an attempt to isolate a viable hypothesis for why hypothetical bias occurs. These less compli-

[20] Cummings et al. (1995a) suggested that subjects in their experiments on bidding for a private good may have interpreted the hypothetical dichotomous choice question as asking whether they, the subjects, would *ever* pay $x for the good. However, private good studies wherein attempts were made to avoid this interpretation by urging subjects to respond as if the purchase were occurring "here and now" nevertheless found substantial bias (Blumenschein et al. 1998; Blumenschein et al. 2001; Johannesson et al. 1998).

[21] Murphy and Stevens (2004) raise the possibility that the reason for hypothetical bias with private goods is quite different from that for public goods—that in bidding for private goods people may actually report their true willingness to pay when the question is hypothetical but truncate their bids at the market price (net of transaction costs) when payment is for real. We discount this possibility based on the observation that in experiments using commonly available market goods mean hypothetical bids tend to be much below the market price of the goods and real payment bids tend to be much closer to the market price.

cated bidding scenarios point to a parsimonious hypothesis, focusing on value imprecision, which appears to have considerable support in the literature—that the lack of consequences allows respondents in hypothetical payment situations to respond based on what are likely to be quite vague notions of willingness to pay, and that the imposition of economic consequences encourages more careful consideration of one's willingness to pay in an effort to avoid paying what seems, or what may later seem, like too much. This, however, does not account for motives that may also influenc bidding in public good experiments, such as the desire to express support for the good or to nurture a self image of generosity. Such motives may widen the range in willingness to pay beyond what would exist for a private good, and open the door for bias reducing techniques that addresses such motives.

Appendix: Dissonance Minimising Questionnaire

Consumer Behaviour Experiment

This is a study about the decisions people make. We will be describing to you a problem facing many children in Australia, and a non-profi organisation that is working on this problem. You are not actually being given the opportunity to contribute to this organization today. We will be presenting you with a hypothetical situation. Even though payment of money in this experiment is hypothetical, we ask that you respond to the questions as if they involved real cash payments.

The Problem: Children Missing Breakfast

It is estimated that one in four Australian school children have an inadequate breakfast or no breakfast at all. These rates are higher in schools in disadvantaged areas, as children from lower socio-economic backgrounds are six times more likely to miss breakfast than children from higher socio-economic backgrounds (O'Dea, Journal of Children's Nutrition and Physical Activity Study, The University of Sydney, 2001).

Many nutritionists consider breakfast to be the most important meal of the day, assisting in children's physical, mental and emotional development. Children who miss breakfast are less able to concentrate, are more prone to fidgeting and f nd learning difficul by mid-morning. Children therefore need an adequate breakfast to participate well at school.

The Solution

The *Australian Red Cross NSW* has started a Breakfast Club initiative that provides about 80,000 breakfasts a year to school children in 26 centers in New South Wales. The program runs in disadvantaged schools across New South Wales, with *Red Cross* volunteers providing children with cereal, toast and juice before they start class.

This initiative has:

- provided better nutrition for children in a comforting and supportive environment;
- lead to an improvement in the children's concentration in the classroom, leading to children achieving better grades; and
- encouraged children to develop vital social and living skills.

The Breakfast Club is funded solely by donations from the Australian public and *relies* mostly on the work of volunteers.

Because the *Red Cross* relies on donations, the number of meals it can provide to children each year through the Breakfast Club depends on the funds received. If more money were given, the *Red Cross* would be able to provide meals to needy children who currently are not receiving these meals. So in a moment we are going to ask if you would like, as a group, to contribute money towards the Breakfast Club initiative.

Donating to the Red Cross Breakfast Club Initiative

If you would like additional breakfasts to be provided to needy school children through the *Red Cross* Breakfast Club, you and the others in the room would be need to donate.

It costs $15 to buy *Red Cross* breakfasts for one week for a child living in a disadvantaged area.

There are reasons why you might want to make a donation towards the Breakfast Club, and there are reasons why you might not want to make a donation. For instance:

- The Breakfast Club does provide disadvantaged children with healthy breakfasts each week.
- But you might think the cost of donating to the Breakfast Club is more then you are willing to spend on it. Or you might have some other reason for not making a donation.

When thinking about whether or not you would support the Breakfast Club, keep in mind your available income and all of the other things you have to spend money on.

How Contributions Would Have Worked

In a moment we are going to hold a secret ballot to decide whether or not you and each other participant in the room would contribute $15 to buy *Red Cross* breakfasts for additional children.

Here's how it will work. After you have each voted about whether you wanted to contribute towards providing a *Red Cross* breakfast, one of us will collect your voting slips.

1. If more than 50% of you had voted YES to this proposal, all of you would—if this experiment were for real – pay $15
   I would have deducted $15 from the money that each of you earned for participating in these experiments and a cheque for this amount ($15 x ____ people = $_____) would be sent to the *Australian Red Cross,* who would send each of you a receipt for your contribution.
2. If 50% or fewer of you vote YES on this proposal, no one would pay $15, and no money would be sent to the *Australian Red Cross*.

Review

We have described to you a program run by the *Australian Red Cross* to provide breakfasts to children in disadvantaged areas. The program is funded by donations from the public, so the number of breakfasts provided each week depends on the donations received.

In a moment you will be voting to decide whether each of you would contribute—if this experiment were for real—$15 towards providing a *Red Cross* breakfast for one week for an additional child in a disadvantaged area.

While this experiment is hypothetical, we would like you to respond to this question as if we were actually giving you the opportunity to contribute to the *Red Cross* today.

But before you do, I want to make it clear that the $20 participation fee that you earned today is your money. You've spent your time helping us in our research, and you've earned it! You were told that the money is yours, so believe it!

Any questions?

Voting Slip

Proposal: Everyone here in this room will contribute $15 to provide *Red Cross* breakfasts for additional children in a disadvantaged area.

Please choose one of the following options:

☐ I would vote YES to this proposal that everyone contribute $15 to the Breakfast Club.
☐ I support the goal of the Breakfast Club, but I'm not prepared to pay $15 and thus would vote NO.
☐ I support the goal of the Red Cross Breakfast Club, but I cannot afford $15 and thus would vote NO.
☐ I support the goal of the Red Cross Breakfast Club, but I would prefer to save my money and contribute to another cause and thus would vote NO.
☐ I support the goal of the Breakfast Club initiative, but I would vote NO for the following reason_____
☐ I would vote NO to this proposal that everyone contribute $15 to the Breakfast Club.

References

Aadland D, Caplan AJ (2003) Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias. Am J Agric Econ 85(2):492–502

Aadland D, Caplan AJ (2006) Cheap talk reconsidered: new evidence from CVM. J Econ Behav Organ 60:562–578

Aadland DM, Caplan AJ, Phillips OR (2007) A Bayesian examination of information and uncertainty in contingent valuation. J Risk Uncertain 35: 149–178

Ajzen I, Brown TC, Carvajal F (2004) Explaining the discrepancy between intentions and actions: the case of hypothetical bias in contingent valuation. Pers Soc Psychol Bull 30(9):1108–1121

Arrow K, Solow R, Portney PR, Leamer EE, Radner R, Schuman H (1993) Report of the NOAA panel on contingent valuation. Fed Regist 58:4601–4614

Bakeman R, Robinson BF (1994) Understanding log-linear analysis with ILOG: an interactive approach. Erlbaum, Hillsdale, NJ

Blamey RK, Bennett JW, Morrison MD (1999) Yea-saying in contingent valuation surveys. Land Econ 75(1):126–141

Blumenschein K, Johannesson M, Blomquist GC, Liljas B, O'Connor RM (1998) Experimental results on expressed certainty and hypothetical bias in contingent valuation. South Econ J 65(1):169–177

Blumenschein K, Johannesson M, Yokoyama KK, Freeman PR (2001) Hypothetical versus real willingness to pay in the health care sector: results from a fiel experiment. J Health Econ 20:441–457

Blumenschein K, Blomquist GC, Johannesson M, Horn N, Freeman P (2008) Eliciting willingness to pay without bias: evidence from a fiel experiment. Econ J 118(535):114–137

Brown TC, Ajzen I, Hrubes D (2003) Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation. J Environ Econ Manage 46(2):353–361

Brown TC, Champ PA, Bishop RC, McCollum DW (1996) Which response format reveals the truth about donations to a public good?. Land Econ 72(2):152–166

Carlsson F, Frykblom P, Lagerkvist CJ (2005) Using cheap talk as a test of validity in choice experiments. Econ Lett 89:147–152

Champ PA, Bishop RC (2001) Donation payment mechanisms and contingent valuation: an empirical study of hypothetical bias. Environ Resource Econ 19:383–402

Champ PA, Bishop RC, Brown TC, McCollum DW (1997) Using donation mechanisms to value nonuse benefit from public goods. J Environ Econ Manage 33:151–162
Cummings RG, Taylor LO (1999) Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. Am Econ Rev 89(3):649–665
Cummings RG, Harrison GW, Rutstrom EE (1995a) Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible?. Am Econ Rev 85(1):260–266
Cummings RG, Harrison GW, Taylor LO (1995b) Can the bias of contingent valuation surveys be reduced? Evidence from the laboratory. Georgia State University, Atlanta
Dubourg WR, Jones-Lee MW, Loomes G (1997) Imprecise preferences and survey design in contingent valuation. Economica 64:681–702
Ethier RG, Poe GL, Schulze WD, Clark J (2000) A comparison of hypothetical phone and mail contingent valuation responses for green-pricing electricity programs. Land Econ 76(1):54–67
Gregory R, Lichtenstein S, Brown TC, Peterson GL, Slovic P (1995) How precise are monetary representations of environmental improvements?. Land Econ 71(4):462–473
Haab TC, McConnell KE (2002) Valuing environmental and natural resources. Edward Elgar, Cheltenham
Johannesson M, Liljas B, Johansson P-O (1998) An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisons. Appl Econ 30:643–647
Kahneman D, Ritov I, Schkade D (1999) Economic preferences or attitude expressions? An analysis of dollar responses to public issues. J Risk Uncertain 19(1–3):203–235
LaPiere RT (1934) Attitudes vs. actions. Social Forces 13(2):230–237
Li C-Z, Mattsson L (1995) Discrete choice under preference uncertainty: an improved structural model for contingent valuation. J Environ Econ Manage 28:256–269
List JA (2001) Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from fiel auctions for sportscards. Am Econ Rev 91(5):1498–1507
List JA, Gallet CA (2001) What experimental protocol influenc disparities between actual and hypothetical stated values? Environ Resource Econ 20:241–254
List JA, Sinha P, Taylor MH (2006) Using choice experiments to value non-market goods and services: evidence from f eld experiments. Adv Econ Anal Policy (Berkeley Electronic Press) 6(2), article 2
Little J, Berrens R (2004) Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. Econ Bull 3(6):1–13
Loomis J, Ekstrand E (1998) Alternative approaches for incorporating respondent uncertainty when estimating willingness to pay: the case of the Mexican spotted owl. Ecol Econ 27(1):29–41
Loomis J, Brown T, Lucero B, Peterson G (1996) Improving validity experiments of contingent valuation methods: results of efforts to reduce the disparity of hypothetical and actual WTP. Land Econ 72(4):450–461
Loomis J, Traynor K, Brown T (1999) Trichotomous choice: a possible solution to dual response objectives in dichotomous choice contingent valuation questions. J Agric Resour Econ 24(2):572–583
Louviere JJ, Hensher DA, Swait JD (2000) Stated choice methods: analysis and applications. Cambridge University Press, New York
Lusk JL (2003) Effects of cheap talk on consumer willingness-to-pay for golden rice. Am J Agric Econ 85(4):840–856
Madariaga B, McConnell KE (1987) Exploring existence values. Water Resour Res 23(5):936–942
Mitani Y, Flores N (2007) Hypothetical and real bias in threshold public goods experiments. University of Colorado, Boulder
Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Resour Future, Washington, D C
Murphy JJ, Stevens TH (2004) Contingent valuation, hypothetical bias, and experimental economics. Agric Resour Econ Rev 33(2):182–192
Murphy JJ, Allen PG, Stevens TH, Weatherhead D (2005a) A meta-analysis of hypothetical bias in stated preference valuation. Environ Resource Econ 30:313–325
Murphy JJ, Stevens TH, Weatherhead D (2005b) Is cheap talk effective at eliminating hypothetical bias in a provision point mechanism? Environ Resource Econ 30:327–343
Norwood FB (2005) Can calibration reconcile stated and observed preferences? J Agric Appl Econ 37: 237–248
Opaluch JJ, Segerson K (1989) Rational roots of "irrational" behavior: new theories of economic decision-making. Northeast J Agric Econ 18:81–95
Poe GL, Clarke J, Rondeau D, Schulze WD (2002) Provision point mechanisms and f eld validity tests of contingent valuation. Environ Resource Econ 23:105–131
Ready RC, Whitehead JC, Blomquist GC (1995) Contingent valuation when respondents are ambivalent. J Environ Econ Manage 29:181–196

Ready RC, Navrud S, Dubourg RW (2001) How do respondents with uncertain willingness to pay answer contingent valuation questions? Land Econ 77(3):315–326
Reber AS (1985) The penguin dictionary of psychology. Penguin Books, London
Sinden JA (1988) Emperical tests of hypothetical bias in consumers' surplus surveys. Aust J Agric Econ 32(2/3):98–112
Taylor LO (1998) Incentive compatible referenda and the valuation of environmental goods. Agric Resour Econ Rev 27(2):132–139
Taylor LO, McKee M, Laury SK, Cummings RG (2001) Induced-value tests of the referendum voting mechanism. Econ Lett 71(1):61–65
Vossler CA, McKee M (2006) Induced-value tests of contingent valuation elicitation mechanisms. Environ Resource Econ 35(2):137–168
Wang H (1997) Treatment of "don't-know" responses in contingent valuation surveys: a random valuation model. J Environ Econ Manage 32(2):219–232