The Australian Journal of Agricultural and Resource Economics

Journal of the Australian Agricultural and Resource Economics Society

Australian Journal of Agricultural and Resource Economics, 61, pp. 76–94

Convergent validity in contingent valuation: an application to the willingness to pay for national climate change mitigation targets in Germany*

Reinhard Uehleke†

This stated preferences survey determines the willingness to pay (WTP) for climate change mitigation policies using a representative sample of the German population. WTP is compared across three valuation question formats in a split sample design: the dichotomous choice (DC) referendum, the dissonance minimizing (DM) referendum and the two-way payment ladder (TWPL). The influence of multinational cooperation on WTP is assessed by variation in the hypothetical scenarios. We demonstrate that the DM referendum and the TWPL, two question formats that induce similar response incentives, yield equal mean WTPs. Multinational cooperation did not change WTP in any of the question formats. Implications for current contingent valuation practice are discussed.

Key words: climate change mitigation, contingent valuation, hypothetical bias, question format, willingness to pay.

1. Introduction

Past contingent valuation studies have regularly shown that different question formats yield different willingness to pay (WTP) values for hypothetical public goods (Champ and Bishop 2006) as well as for real donation public goods (Brown *et al.* 1996). In general, the dichotomous choice (DC) referendum leads to a larger WTP than the open-ended or the payment ladder format. Few studies have found convergent validity, for example Reaves *et al.* (1999). Thus, the choice of question format is a crucial factor for any contingent valuation (CV) study.

The DC referendum is the only clearly strategic incentive compatible format (Carson and Groves 2007) and has been able to predict agreement rates in several real community referenda (Loomis 2014). However, in the presence of cost and value uncertainties surrounding many long-term environmental projects, such as the reduction of greenhouse gas emissions, other formats may be better suited to assess WTP, because in these cases, the

* This research was financed by the German Federal Ministry of Education and Research (FKZ 01LA1108B).

† Reinhard Uehleke (email: reinhard.uehleke@uni-rostock.de) Department of Business Administration, Leipzig University of Applied Sciences, Leipzig, Germany. Present address: Department of Agricultural Economics, University of Rostock, Justus-von-Liebig-Weg 2, 18059, Rostock, Germany.

doi: 10.1111/1467-8489.12148

DC referendum might produce a larger hypothetical bias due to behavioural responses such as yea-saying (Kanninen 1995; Boyle *et al.* 1996; Ready *et al.* 1996; Blamey *et al.* 1999) and respondent uncertainty (Ready *et al.* 2001; Flachaire and Hollard 2007; Hanley *et al.* 2009).

The dissonance minimising (DM) referendum (Blamey *et al.* 1999) tries to directly address yea-saying by providing additional answer categories to the yes/no/would not vote options of the DC referendum, so that subjects can express their support for the program without having to state that they are willing to pay the posted amount. Hence, the DM format gives the option to support the program, while stating that one cannot afford, or is not willing to pay the given amount. Another format, the two-way payment ladder (TWPL) (Bateman *et al.* 2002), gives respondents the possibility to express WTP as an interval instead of a point value.

This study contributes to the literature about convergence validity in CV settings by directly comparing WTP for Germany's climate change mitigation targets under three question formats: the classic single DC referendum, the DM referendum and the TWPL. Based on the literature about response incentives in CV, we conclude that the DM format and the payment ladder induce qualitatively identical psychological and strategic response incentives and demonstrate that they indeed yield equal WTPs. This result improves understanding about the influence of question format on WTP responses and provides useful guidance of when to use which question format. The study furthermore contributes to the international literature on WTP for CO_2 reduction policies (Berrens *et al.* 2004; Li *et al.* 2004; Akter and Bennett 2011; Kotchen *et al.* 2013) and the growing literature on WTP for climate change mitigation in Germany (Achtnicht 2011, 2012; Löschel *et al.* 2013; Diederich and Goeschl 2014). But while the latter elicit hypothetical and real WTP values from a voluntary contribution perspective, we provide the first empirical evidence about WTP for Germany's CO_2 reduction target using a large representative sample of the German population.

The mean WTP for an increase in the 2020 national climate mitigation target from 30 per cent to 40 per cent CO_2 reduction is €403 (DC referendum) and €205 (DM referendum) per household per year. Maximum WTP in the TWPL is between €202 and €257. WTP in the DM referendum is not significantly different from the maximum WTP of the TWPL. Finally, we find that there is no influence of multinational cooperation on WTP across all formats.

The article is structured as follows: section 2 examines the literature about incentive compatibility and hypothetical bias, section 3 presents the hypotheses of the study, section 4 describes the study design and section 5 depicts the results. Section 6 closes the article with a discussion about implications for CV studies.

2. Contingent valuation and incentive compatibility

Two rival explanations dominate the discussion about the sources of hypothetical bias in CV studies: the strategic behaviour (Carson and Groves

2007) and the yea-saying (Blamey *et al*. 1999) or response uncertainty (Ready *et al*. 2001; Flachaire and Hollard 2007) arguments.

2.1 Consequentiality and mechanism design theory

Carson and Groves (2007) derive two basic conditions for incentive compatibility that are necessary for an unbiased WTP estimate from stated preference surveys: the survey must be consequential, meaning that respondents have to believe that the results at least gradually influence the policy decision, and the payment rule must be coercive to prevent free-riding behaviour. The DC referendum format presents a specified and coercive payment. Therefore, it is incentive compatible for the evaluation of a public good. Continuous-response formats, such as the open-ended format or the payment card, however, do not specify project costs and therefore signal that the cost allocation is not fixed, so that the optimal response depends on *perceived* costs and how the agency is believed to use the information. Carson and Groves (2007) conclude that in most cases, these formats pose an incentive to strategically underbid one's true WTP. Accordingly, WTP assessed with the DC format is expected to be closer to true WTP than the WTP assessed with a payment ladder or other continuous formats, and observed differences in WTP can be explained by the different response incentives of the formats.

Recent CV surveys confirm that the DC format can successfully predict outcomes of real referenda over a set of community projects (Vossler and Kerkvliet 2003; Vossler *et al*. 2003; Johnston 2006; Vossler and Watson 2013). The prediction success of these hypothetical referenda is ascribed to the degree of consequentiality evoked by these surveys.

Experimental evidence on the impact of consequentiality on stated preferences supports this result. Cummings and Taylor (1998) demonstrated that the degree of consequentiality must be rather large to mitigate hypothetical bias of a majority vote, while Carson *et al*. (2014) found that any positive probability of influence on the decision is sufficient to produce as-if-real information. The contradiction of these findings might originate from the different experimental goods. Cummings and Taylor used an open referendum on a local quasi-public good – the provision of a Citizen Guide to poor Hispanic families in Albuquerque, New Mexico – to secure these citizens' access to information about groundwater quality in their neighbourhood, while Carson *et al*. used a closed referendum on a publicly provided private good, namely sports memorabilia. These goods imply different value uncertainties to respondents with the value uncertainty for the Citizen Guide being larger than that of the sports memorabilia among specialised collectors. It is conceivable that with increasing value uncertainty, the degree of consequentiality must increase as well in order to be sufficient for responses to mimic real responses. However, the author is not aware of any existing experimental evidence concerning this hypothesis. Vossler and

Evans (2009) additionally reveal that in a typical advisory referendum on a local public good (classroom recycle container) hypothetical bias is caused by respondents who view their vote as inconsequential.

This result evoked hopes that CV estimates could be corrected by a simple ex-post consequentiality adjustment. But recent field surveys (Herriges *et al.* 2010; Vossler and Watson 2013; Interis and Petrolia 2014) show that respondents in a DC referendum, who ex-post assess the survey as inconsequential, have a statistically *lower* WTP than those who assess the survey as at least minimally consequential, which is at odds with theoretical predictions and experimental evidence.

2.2 Respondent uncertainty and yea-saying

In nonmarket valuation, respondent uncertainty most generally refers to the confidence regarding how well the hypothetical response matches an actual valuation decision (Moore *et al.* 2010). This uncertainty can have many sources (see e.g. Martínez-Espiñeira and Lyssenko 2012) and is argued to be at least partly responsible for hypothetical bias (Champ and Bishop 2001). Ready *et al.* (2001) and Flachaire and Hollard (2007) found that respondents in the DC format state yes to amounts when they are relatively unsure that they would really pay, whereas in open formats, such as the payment ladder, they state values that they are relatively sure they would really pay.

Furthermore, respondents to the DC format may be faced with the two conflicting aims of wanting to answer the question truthfully and to signal a positive attitude towards the project. This form of cognitive dissonance can boost hypothetical bias through yea-saying, which occurs if respondents neglect the costs in favour of a supportive expression to the program (Brown *et al.* 1996; Loomis *et al.* 1996; Green *et al.* 1998; Blamey *et al.* 1999). The presence of yea-saying in the DC referendum format may produce WTP values that are further away from true WTP than WTP derived by other formats, despite the fact that the latter are not strategic incentive compatible. So far, the only external validation study by Morrison and Brown (2009) showed that the DM format can successfully eliminated hypothetical bias.

3. Response incentives in the survey and hypotheses

The global public good under consideration is the reduction of CO_2 emissions from 30 per cent to 40 per cent by 2020 (compared to 1990) in Germany in order to mitigate climate change. This contribution to the global public good is surrounded by large value uncertainty due to the wide ranges of estimators for marginal damage of CO_2 emissions and free-riding behaviour within the international community. Therefore, the public good of interest in this research shares some characteristics with the experimental goods of Morrison and Brown (2009) and Vossler and Evans (2009). The provision of children's breakfasts and a classroom recycle bin are characterised as well as the

national CO2 emission reduction by a low marginal utility of the group contribution, value uncertainty, free-riding of outsiders and moral duties to act.

For this case study, the incentive properties of the DM and payment ladder formats can be analysed using mechanism design theory as outlined in Carson and Groves (2007). The introduction letter emphasised that the survey was financed by the Federal Ministry of Education and Research and the CV scenarios closely resembled real policy options as they are discussed by politicians and in the media. Because of government funding and the presence of the topic in the media, the survey gives an impression of being at least weakly consequential. In the DM format, payment is coercive as in the DC format. The additional categories of the former, however, might signal that cost distribution is not fixed and hence the optimal response depends on the expectations of the respondent about the probability of provision.

Consider agents with a WTP larger than the posted price: a strategic incentive to underbid exists in the DM format if agents believe that the good will be supplied with large probability. Then it is optimal to state that one is supporting the program, but not willing to pay the posted price, in order to force the government to decrease costs assessed against the respondent or to distribute costs differently. Yet, if the agent believes that the probability of provision increases with the proportion of approval for the posted price, he must balance the increased probability that the good will be provided against a potential downward shift in costs. The optimal response now equals the posted price, if the increase in probability is more important to the respondent than the potential cost reduction. The incentives in the payment ladder are similar to those of the DM format. The payment ladder does not encourage yea-saying, because a participant can express general support for the policy by choosing a low bid amount. The ladder also signals, perhaps more strongly than the DM format, that costs are not fixed, because no costs are specified. The optimal response now depends on *perceived* costs. If WTP is larger than perceived costs, one must again balance an increased provision probability for large stated WTPs against the potential upward shift in costs. The optimal response equals perceived costs, because otherwise the surplus might be extracted. If the WTP is smaller than the perceived costs, the optimal response would be zero, as any other response increases the likelihood of provision at costs larger than utility.

The comparison of WTP from the DM and the payment ladder formats provides a base for testing the procedural variance, because they imply equal response incentives but are conceptually different. We use the TWPL instead of a regular payment ladder to gain more flexibility for hypotheses testing. For example, if respondents do not know their exact WTP but an interval in which their WTP lies, it is unclear what point of the interval the DM format measures. On the one hand, it could measure the lower end of a possible WTP interval if people say yes to the bid amount only when they are really sure they would pay the amount and otherwise lean towards the support options.

On the other hand, they could state yes to the posted price, even if they are not sure whether they would pay. They could also state yes if the posted price is somewhere in the middle of the possible WTP interval. Depending on the response behaviour, WTP in DM will either be close to the lower end, the middle point or the upper end of the observed interval of the TWPL. The last possibility is that besides the behavioural and strategic bias discussed so far, there exist other important drivers of response behaviour that affect WTP differently across different formats. Then we would observe that WTP_{DM} is very different from WTP of the payment ladder. To summarise, the following hypotheses will be tested:

H1: $WTP_{DC} = WTP_{DM}$

H2: $WTP_{DC} = WTP_{TWPL\ upper\ end}$

H3^{I}: $WTP_{DM} = WTP_{TWPL\ lower\ end}$

H3^{II}: $WTP_{DM} = WTP_{TWPL\ upper\ end}$

Additionally, the mitigation scenario is varied between a unilateral reduction scenario (UNILAT) and a multilateral reduction scenario (MULTILAT). The UNILAT scenario states that other countries do not change their behaviour, and the MULTILAT scenario states that simultaneously, other countries would also increase efforts to reduce emissions. WTP is expected to be larger for the MULTILAT scenario, because a fraction of participants might be conditional contributors (Fischbacher *et al.* 2001) who only vote yes in the presence of contributions by others and because the utility of an international emission reduction is larger than that of a unilateral reduction. Therefore, we also test for the difference between UNILAT and MULTILAT:

H4: $WTP_{MULTILAT} > WTP_{UNILAT}$ in all formats

4. Sample and study design

The Internet survey was carried out between 19 March 2014 and 9 April 2014 by Panelbiz, one of the largest online fieldwork service providers in Germany. Panel members were invited by the provider to participate in a survey via email. The sample was drawn from a population of approximately 250,000 panel members and consists of 5021 individuals that were randomly assigned to the treatment variations. We excluded 367 straightliners who completed the questionnaire in less than 4 min or gave inconsistent answers on two attitude scales. This yields roughly 900 participants for each referendum format and 475 participants for each payment ladder format. As the referendum formats are statistically less efficient than the payment ladder, more observations were assigned to those treatments. The sample is

representative for the German population between 18 and 69 years by gender and age. The share of respondents with a secondary education is a little lower than in the population, while the share of respondents with a tertiary education is larger. Nonetheless, the sample is representative by professional education if one ignores subjects that are still in education. Thus, the sample is a good representation of the German population between 18 and 69 years.

The study employed six split treatments to elicit WTP under the DC, DM and TWPL formats, which were presented as unilateral (UNILAT) and multilateral (MULTILAT) mitigation scenarios. The CV scenarios and the information screen are designed to closely resemble real climate change mitigation policy options, namely the strengthening of the 2020 national CO_2 emission target from 30 per cent to 40 per cent CO_2 reduction compared to emission levels of 1990. The Federal Ministry of Environment acts on the assumption that with current GHG mitigation policies, a reduction of 33 per cent will be achieved by 2020 (BMUB 2014). To close the gap of seven percentage points, present mitigation policies must be tightened and additional measures have to be introduced, especially in the sectors of energy, industry housing and traffic.

Each respondent received only one treatment combination, yielding six independent WTP estimators. Before subjects saw the CV scenario, they were informed about the policy goal and the policy mix to reach the new emission target. The information reflects the political discussion in Germany and is reproduced in Appendix I. After a short cheap talk script (see Appendix II), subjects saw the CV scenario. Figure 1, for example, presents the UNILAT scenario with the DC referendum format. The bid vector for all treatments comprised 14 bid levels from very low values to very high values: {48; 72; 84; 108; 156; 192; 252; 324; 432; 540; 720; 960; 1200; 1440} € per household per year. The corresponding costs per month were also stated. This design was chosen to cover a large distribution of WTP, while maintaining a reasonable number of respondents per bid point.

The MULTILAT scenario differed from the UNILAT treatment only via the text in the box, which specified that the EU and other industrialised countries would simultaneously also increase reduction targets. An example for MULTILAT in combination with the DM referendum is given in Appendix III.

The DM referendum provides further categories to the standard DC referendum, which allow the respondent to express favour for the referendum without having to agree to pay the posted price. The TWPL gives subjects the opportunity to express their WTP as an interval. On one side, they can select an amount they would definitely be willing to pay, and on the other side, an amount they would definitely be no more willing to pay. To compare the outcomes of the different formats, the payment ladder format wording adapted a majority rule where subjects stated the amount they would agree to in a possible referendum to avoid the individual donation perspective of the conventional payment ladder format. The TWPL screen is given in Appendix IV.

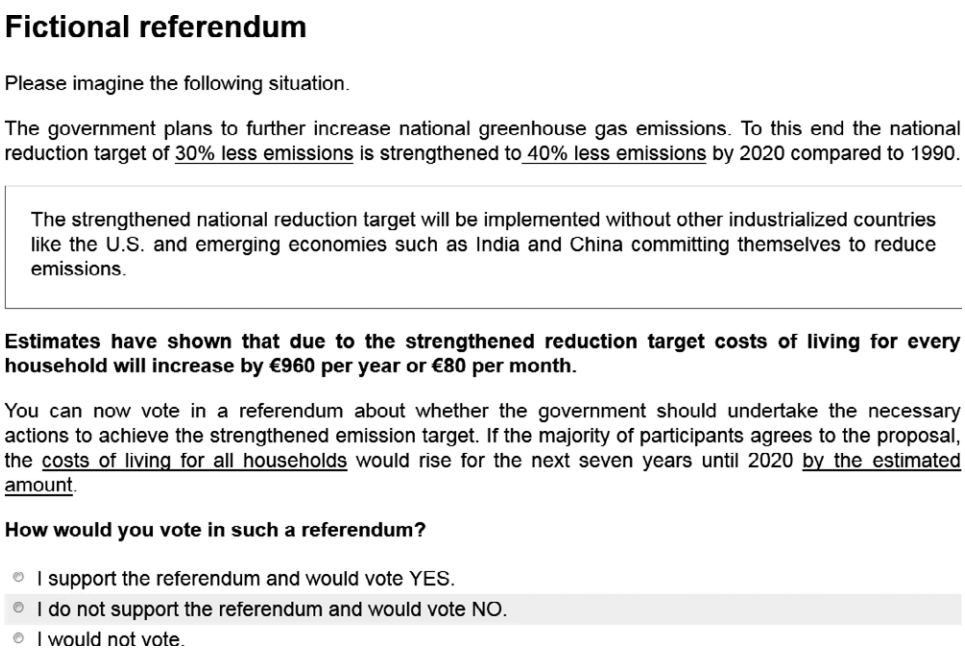


Figure 1 CV scenario for the DC referendum (UNILAT).

A theoretically incentive compatible CV protocol needs a coercive and fixed (Carson and Groves 2007), respectively, decoupled payment vehicle (Green *et al.* 1998). To decouple an agent's potential future costs from his response, the payment vehicle in all treatments specifies that costs will be distributed across all consumers by an equal increase in the cost of living of all households.[1]

5 Results

5.1 Analyses of the referendum formats

Table 1 shows the proportion of answers to the referendum formats by scenario and the number of observation in each treatment. In the DC referendum, 37 per cent (UNILAT) and 41 per cent (MULTILAT) support the proposal at given bid levels and a large share of about 15 per cent refrains from voting. The DM referendum gives much lower shares of yes votes. This is because the DM format includes the additional possibility to express

[1] Other distribution rules would have been possible, for example an income tax surcharge. We chose the equal cost for all households rule, because it is straightforward and easy to understand. Progressive income tax increases might be viewed as fairer and more realistic but can introduce other form of biases, for example a percentage bias, and require much more effort for a meaningful response. We do not preclude that the distribution rule has no effect on WTP and suggest interpreting absolute WTP values with care.

Table 1 Percentage share of votes in the referendum treatments

	DC_Unilat	DC_Multilat	DM_Unilat	DM_Multilat
Yes	37.27	40.69	20.47	25.81
No	47.54	43.70	15.11	13.50
Yes, but I am not able to pay	—	—	19.19	16.74
Yes, but I am not willing to pay	—	—	31.62	33.59
Yes, but would not pay, because …	—	—	11.04	7.56
Abstention	15.19	15.61	2.57	2.81
N	915	929	933	926

support for the program, but to reject the proposed bid level. Here, the majority supports the proposal, but only 20 per cent or 26 per cent are willing to pay the given amount. The share of abstentions is three per cent, and therefore, abstentions are included as no votes in all formats to make the results comparable. Otherwise, 15 per cent of observations would be excluded only in the DC format, which could introduce a missing value bias.

Figure 2 shows the distribution of yes responses to the DC and DM referendum formats along with the smoothed demand curves. As expected, the propensity to say yes decreases with increasing price. The probability for a yes response is larger in the MULTILAT treatment only at low costs. Additionally, curves overlap for many bidding points, and thus the distributions of yes responses do not indicate support H4 of a larger mean WTP in the MULTILAT treatment in any of the referendum formats.

To derive a welfare estimate from the DC referendum data, the nonparametric Turnbull estimator (Turnbull 1976) provides a good starting point, as it derives WTP from the empirical distribution without further distributional assumptions and requires only that the proportion of yes responses decreases monotonically with bid levels. Because shares of yes responses do not decrease monotonically, a monotonicity restriction (Haab and McConnell 1997) is imposed. The idea is to pool cost levels for which the share of yes responses increases together with the next lowest bid level and weight each pooled share with the respective lower bid. Table 2 summarises the resulting WTP measures. The top panel gives the WTP for the DC referendum and the lower panel the WTP for DM referendum. In the DC referendum, WTP_{UNILAT} is €369 per household per year. By comparison, $WTP_{MULTILAT}$ is €293 and, therefore, actually lower than WTP_{UNILAT} although the overall share of yes responses in the former is larger. This contradiction can be explained by the cost-level pooling: The pooling affects the last bid level of the MULTILAT data, and therefore, $WTP_{MULTILAT}$ is biased downwards. This is illustrated by recalculating WTP after the exclusion of four observations that stated yes to the last bidding point. This exclusion decreases the share of yes responses, which should decrease WTP but instead WTP is increased by almost €100 to €380 simply because the last category is then not pooled down. Therefore, the blank evidence for a larger

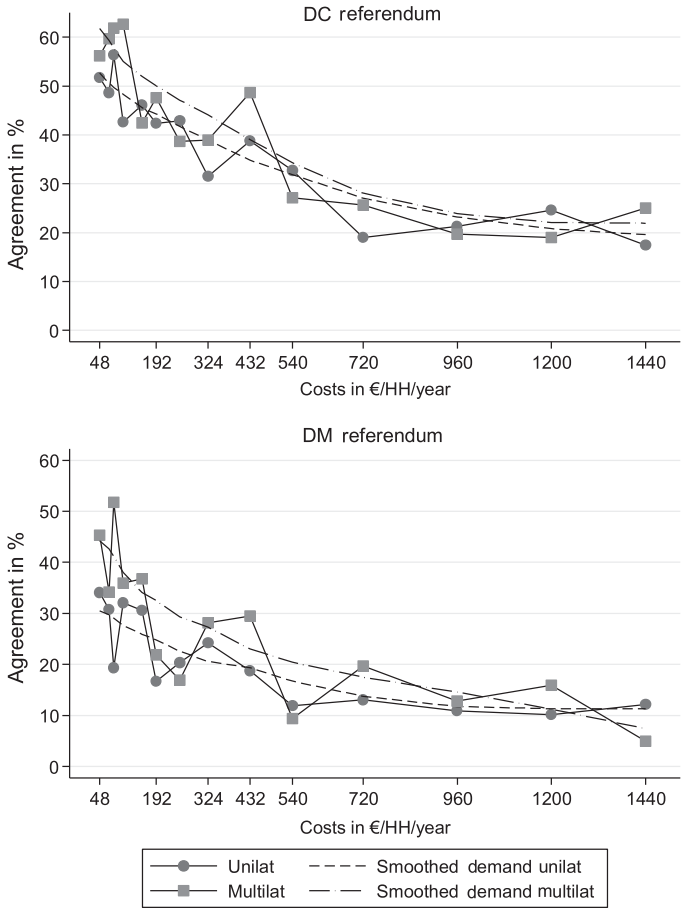


Figure 2 Distribution of yes responses in the referendum formats.

WTP in UNILAT given by the monotonicity restricted Turnbull should not be given too much weight, because the Turnbull estimator for the MULTILAT treatment is sensitive to small changes in the data set.

In the DM referendum, WTP_{UNILAT} is €154 per household per year, about €50 lower than $WTP_{MULTILAT}$ (t-test, $P < 0.1$).[2] But this result is again caused by the monotonicity restriction, which affects the last category in UNILAT only. Excluding one observation which stated yes to the highest bid-amount inconsistently increases WTP for UNILAT to €179, and the difference to WTP MULTILAT is no longer significant.

To avoid this kind of bias due to asymmetric cost-level pooling, we propose a different procedure to satisfy the monotonicity restriction. Demand curves are smoothed by employing locally weighted regressions and the Turnbull

[2] The 95 per cent confidence interval is $\overline{WTP} \pm 1.96 \times \sqrt{\sigma^2}$. The test statistic for the difference in average WTP constitutes $t = \frac{\overline{WTP_1} - \overline{WTP_2}}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ (Haab and McConnell 2003).

Table 2 WTP/HH/year in the DC referendum format

	DC referendum				
	UNILAT	MULTILAT	*t*-value	Sig. level	Pooled
Turnbull WTP [95% CI]	369 [297; 441]	293 [254; 332]	1.84	$P < 0.1$	403 [359; 446]
Median WTP	48–108	108–156	—	—	108–156
Turnbull WTP on locally smoothed proportions [95% CI]	419 [366; 472]	460 [407; 513]	1.07	$P > 0.1$	437 [400; 474]
	DM referendum				
	UNILAT	MULTILAT	*t*-value	Sig. level	Pooled
Turnbull WTP [95% CI]	154 [123; 185]	200 [157; 243]	1.69	$P < 0.1$	205 [172; 238]
Median WTP	0–48	0–48	—	—	0–48
Turnbull WTP on locally smoothed proportions [95% CI]	228 [188; 268]	269 [228; 310]	1.40	$P > 0.1$	249 [221; 278]

Note: *t*-values refer to the difference between UNILAT and MUILTILAT.

analysis is performed on the smoothed yes proportions to each bid level. The demand curves are depicted in Figure 2, and the resulting WTP is given in Table 2. It turns out that the WTP estimators from the smoothed demand curves are somewhat higher than with cost pooling. The reason is that yes votes are not pooled down into lower categories, but instead departure from monotonicity is treated as part of the randomness in the data generating process. This way the original bid vector can be retained. Therefore, this approach is more appropriate to compare WTPs across groups, especially if many categories have to be pooled to ensure monotonicity.[3]

Summing up, the statistically weak differences between WTP in UNILAT and MULTILAT are generated by asymmetries in the cost pooling procedure and are very sensitive to small changes in the data set. A thorough Turnbull analysis fails to reject equality of WTP across scenarios, which we therefore pool together. The smoothed Turnbull WTP is then €437 in the DC referendum and €249 in the DM referendum. This difference is highly significant (*t*-value = 7.83), and we reject hypothesis H1 of equal WTP in the DC and DM referenda.

5.2 WTP in the payment ladder format

The comparison of a payment ladder WTP with the Turnbull estimator is straightforward, because the mean of the payment ladder values corresponds

[3] We also performed a sensitivity analysis on the pooled data by reducing the bandwidth for the calculation of the smoothed values as much as possible, so that monotonicity is still reached. The results are robust, because with the according bandwidth of 0.4, WTP is only about €10 smaller than with a bandwidth of 0.8, which is used for the results in Table 2.

Table 3 WTP in the payment ladder

	L	*U*	Interval midpoint	Linear index†	Uncertainty adjusted‡	Interval regression
Mean WTP	90	257	174	146	133	131
Median WTP	48	108	78	68	75	—

Notes: The number of observations is 951. †Linear index according to Hanley *et al.* (2009): $2/3L + 1/3U$. ‡Uncertainty adjusted WTP according to Voltaire *et al.* (2013): $U-(U-L) \times (U-L)/U$.

with the lower bound Turnbull estimator of the DC referendum (Haab and McConnell 2003; p. 127). The TWPL collects a range of WTP, namely the amount that one is *definitely willing to pay* and the amount one is *definitely no more willing to pay*. The mean upper (U) and lower (L) ends of the interval are not different between scenarios according to a Wilcoxon rank-sum test, and we continue to analyse the pooled data. Table 3 shows the interval WTP along with alternative indices for point WTP.

As expected, WTP from the payment ladder is much smaller than WTP_{DC}. Second, the lower end payment ladder WTP (L) and the point measures for WTP are much smaller than WTP_{DM} of €205 while the upper end WTP (U) of €257 comes close to the smoothed Turnbull WTP of €249. This finding suggests that WTP_{DM} is similar to the maximum WTP from a payment ladder. As our upper end of the interval WTP is the amount that one is *definitely no more willing to pay*, the point estimator for maximum WTP lies between the amount that one would no longer be willing to pay and the next lowest category. The resulting mean maximum WTP is then in the interval [202; 257].

To construct a test statistic that consistently compares the estimators for maximum WTP with the DM referendum format, we create artificial referendum data from the payment ladder.[4] We randomly assign a cost level to each observation in the payment ladder group, check whether the assigned bid level is smaller or larger than the chosen WTP and create an indicator that is zero if chosen WTP is smaller than the bid level and unity otherwise. We then calculate the Turnbull WTP along with its variance. This process is repeated 400 times, and the means of the resulting Turnbull WTPs and variances are derived. This way, the dichotomous character from the DC and DM formats is imitated, and a t-test for the differences in Turnbull WTP can be applied. Because we do not have a point estimator, we will perform this analysis for both ends of the maximum WTP interval. If WTP_{DM} is not statistically different from the ends of the payment ladder interval, it will not be different from the point estimator as well.

Table 4 summarises the resulting WTP estimators for the maximum WTP interval from the payment ladder, the artificial Turnbull WTP for both ends

[4] Note that interval midpoints would neglect uncertainty about the exact values within each interval and interval regression assumes normality for the ML estimator and the confidence intervals to be consistent (Cook and McDonald 2013).

Table 4 Maximum WTP from payment ladder and artificial Turnbull WTP

	Lower end	Upper end
Observed max. Payment Ladder WTP	202	257
Mean artificial Turnbull WTP [95% CI]	179 (1.12)† [147; 211]	228 (0.86)† [189; 267]
Artificial Turnbull WTP on mean shares [95% CI]	202 (2.23)‡ [171; 232]	257 (0.38)‡ [221; 294]

Notes: 95% confidence intervals in squared brackets. †t-values for the difference against mean WTP_{DM} of 205 €. ‡t-values for the difference against mean smoothed WTP_{DM} of 249 €.

of the interval together with their 95 per cent confidence intervals and the t-values for difference to WTP_{DM}. The mean artificial Turnbull WTP in the second row of Table 4 is smaller than the original observed ladder WTP. This can again be attributed to the cost-level pooling, which has a systematic downward effect on WTP. The t-values confirm that we cannot reject equality between both ends of the maximum WTP interval from the payment ladder and WTP_{DM}. Therefore, hypothesis $H3^{II}$ cannot be rejected and we conclude that WTP_{DM} equals the WTP from a payment ladder that asks for the maximum amount someone is willing to pay.

The downward bias in the mean artificial Turnbull WTP justifies a closer inspection. We could alternatively mimic referendum data from the payment ladder by first averaging over the shares of yes at the randomly assigned cost levels and then calculate the Turnbull WTP. The resulting artificial Turnbull WTP is exactly the same as the WTP observed in the payment ladder. This procedure eliminates the randomness in responses and yields a monotonous decreasing bid function, which can be compared to the smoothed WTP_{DM} from Table 1. With the t-values given in the third column of Table 4, we still cannot reject hypothesis $H3^{II}$ at the one per cent confidence level.[5]

6 Discussion

This study determines the WTP for strengthening Germany's emission target from 30 per cent to 40 per cent less emissions by 2020 under three different CV question formats and varying international cooperation. International cooperation had no influence on WTP in all question formats. This coincides with the results by Cai *et al.* (2010) who also find a general lack of sensitivity of WTP to international cost shares and that an increase of cost shares for only the US and Japan has an effect on WTP and only if respondents believe that these countries bear a responsibility for climate change. Thus our simple method of switching from a unilateral to a multilateral mitigation policy might be too crude to capture this kind of subtle nuances of response behaviour.

[5] T-tests were performed for different bandwidths of smoothed WTP_{DM} (0.4; 0.5; 0.6; 0.7; 0.9; 1). For all bandwidths, we cannot reject equality between WTP_{DM} and the upper end of the maximum WTP of the payment ladder at the 10 per cent level. For the lower end, we cannot reject equality at the one per cent level.

The second finding is that WTP under a DM format equals maximum WTP of a TWPL. This result proves that two conceptually very different question formats, which nevertheless induce qualitatively the same behavioural and strategic response incentives, yield equal WTP estimates.

The equivalence of WTP across these two formats has important implications. First, mean WTP from a DM referendum represents an upper limit of a WTP range subjects consider to be willing to pay, because it equals mean maximum WTP from a TWPL. A similar result was also found by a comparison of the open-ended format and an open-ended interval question format for WTP for changing a stream flow in northern Sweden (Belyaev and Kriström 2010). Second, the presentation of a list of possible WTP values in a payment ladder does not bias WTP differently from a referendum style question in which the problem of yea-saying is specifically addressed. Third, the article demonstrates a possible approach to overcome the different statistical characteristics of these question formats and to perform a consistent analysis for differences in mean WTP. Additionally, the important consequence for any CV project is that if one believes that behavioural responses might exert more influence on WTP results than possible strategic underbidding, the decision of which alternative to the DC format to adopt is not that important for absolute WTP as long as strategic and behavioural responses of the formats are similar. This increases confidence in the payment ladder, which should be favoured over the DM format, because it collects WTP information more efficiently.

Furthermore, our overview of the experimental evidence on preference revelation (Cummings and Taylor 1998; Morrison and Brown 2009; Vossler and Evans 2009) suggests that a low consequentiality is not sufficient for the DC format to reveal true WTP for public goods associated with a mix of large value uncertainty, payment and provision uncertainty and low marginal utility of the group contribution. Under these conditions, other formats such as the payment ladder could be more appropriate for WTP elicitation. However, further systematic experimental evidence on this issue is needed to derive more distinguished guidance on when to use which question format.

References

Achtnicht, M. (2011). Do environmental benefits matter? Evidence from a choice experiment among house owners in Germany, *Ecological Economics* 70, 2191–2200.

Achtnicht, M. (2012). German car buyers' willingness to pay to reduce CO_2 emissions, *Climatic Change* 113, 679–697.

Akter, S. and Bennett, J. (2011). Household perceptions of climate change and preferences for mitigation action: the case of the Carbon Pollution Reduction Scheme in Australia, *Climatic Change* 109, 417–436.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Özdemiroğlu, E., Pearce, D.W., Sugden, R. and Swanson, J. (2002). *Economic Valuation with Stated Preference Techniques: A Manual*. Edward Elgar, Cheltenham/Northampton.

Belyaev, Y. and Kriström, B. (2010). *Approach to Analysis of Self-Selected Interval Data*. CERE Working Papers 2010:2. Centre for Environmental and Resource Economics, Umeå Universitet, Umeå, Sweden.

Berrens, R.P., Bohara, A.K., Jenkins-Smith, H.C., Silva, C.L. and Weimer, D.L. (2004). Information and effort in contingent valuation surveys: application to global climate change using national internet samples, *Journal of Environmental Economics and Management* 47, 331–363.

Blamey, R.K., Bennett, J.W. and Morrison, M.D. (1999). Yea-saying in contingent valuation surveys, *Land Economics* 75, 126–141.

BMUB (2014). Aktionsprogramm Klimaschutz 2020 http://www.bmub.bund.de/themen/klima-energie/klimaschutz/klima-klimaschutz-download/artikel/aktionsprogramm-klimaschutz-2020-eckpunkte-des-bmub/?tx_ttnews[backPid]=289 [accessed 18 September 2014].

Boyle, K.J., Johnson, F.R., McCollum, D.W., Desvousges, W.H., Dunford, R.W. and Hudson, S.P. (1996). Valuing public goods: discrete versus continuous contingent-valuation responses, *Land Economics* 72 (3), 381–396.

Brown, T.C., Champ, P.A., Bishop, R.C. and McCollum, D.W. (1996). Which response format reveals the truth about donations to a public good?, *Land Economics* 72, 152–166.

Cai, B., Cameron, T.A. and Gerdes, G.R. (2010). Distributional preferences and the incidence of costs and benefits in climate change policy, *Environmental and Resource Economics* 46, 429–458.

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

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

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

Champ, P.A. and Bishop, R.C. (2006). Is willingness to pay for a public good sensitive to the elicitation format?, *Land Economics* 82, 162–173.

Cook, J. and McDonald, J. (2013). Partially adaptive estimation of interval censored regression models, *Computational Economics* 42, 119–131.

Cummings, R.G. and Taylor, L.O. (1998). Does realism matter in contingent valuation surveys?, *Land Economics* 74, 203–215.

Diederich, J. and Goeschl, T. (2014). Willingness to pay for voluntary climate action and its determinants: field-experimental evidence, *Environmental and Resource Economics* 57, 405–429.

Fischbacher, U., Gachter, S. and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment, *Economics Letters* 71, 397–404.

Flachaire, E. and Hollard, G. (2007). Starting point bias and respondent uncertainty in dichotomous choice contingent valuation surveys, *Resource and Energy Economics* 29, 183–194.

Green, D., Jacowitz, K.E., Kahneman, D. and McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods, *Resource and Energy Economics* 20, 85–116.

Haab, T.C. and McConnell, K.E. (1997). Referendum models and negative willingness to pay: alternative solutions, *Journal of Environmental Economics and Management* 32, 251–270.

Haab, T.C. and McConnell, K.E. (2003). *Valuing Environmental and Natural Resources. The Econometrics of Non-market Valuation*. E. Elgar Pub, Cheltenham/Northampton.

Hanley, N., Kriström, B. and Shogren, J.F. (2009). Coherent arbitrariness: on value uncertainty for environmental goods, *Land Economics* 85, 41–50.

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

Interis, M.G. and Petrolia, D.R. (2014). The effects of consequentiality in binary- and multinomial-choice surveys, *Journal of Agricultural and Resource Economics* 39, 201–216.

Johnston, R.J. (2006). Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum, *Journal of Environmental Economics and Management* 52, 469–481.

Kanninen, B.J. (1995). Bias in discrete response contingent valuation, *Journal of Environmental Economics and Management* 28, 114–125.

Kotchen, M.J., Boyle, K.J. and Leiserowitz, A.A. (2013). Willingness-to-pay and policy-instrument choice for climate-change policy in the United States, *Energy Policy* 55, 617–625.

Li, H., Berrens, R.P., Bohara, A.K., Jenkins-Smith, H.C., Silva, C.L. and Weimer, D.L. (2004). Would developing country commitments affect US households' support for a modified Kyoto Protocol?, *Ecological Economics* 48, 329–343.

Loomis, J. (2014). 2013 WAEA Keynote address: strategies for overcoming hypothetical bias in stated preference surveys, *Journal of Agricultural and Resource Economics* 39, 34–46.

Loomis, J., Traynor, K. and Brown, T. (1996). Trichotomous choice: a possible solution to dual response objectives in dichotomous choice contingent valuation questions, *Journal of Agricultural and Resource Economics* 24, 572–583.

Löschel, A., Sturm, B. and Vogt, C. (2013). The demand for climate protection – empirical evidence from Germany, *Economics Letters* 118, 415–418.

Martínez-Espiñeira, R. and Lyssenko, N. (2012). Alternative approaches to dealing with respondent uncertainty in contingent valuation: a comparative analysis, *Journal of Environmental Management* 93, 130–139.

Moore, R., Bishop, B.P. and Champ, P.A. (2010). Accounting for respondent uncertainty to improve willingness-to-pay estimates, *Canadian Journal of Agricultural Economics* 58, 381–401.

Morrison, M. and Brown, T.C. (2009). Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies, *Environmental and Resource Economics* 44, 307–326.

Ready, R.C., Buzby, J.C. and Hu, D. (1996). Differences between continuous and discrete contingent value estimates, *Land Economics* 72, 397–411.

Ready, R.C., Navrud, S. and Dobourg, R. (2001). How do respondents with uncertain willingness to pay answer contingent valuation questions?, *Land Economics* 77, 315–326.

Reaves, D.W., Kramer, R.A. and Holmes, T.P. (1999). Does question format matter? Valuing an endangered species, *Environmental and Resource Economics* 14, 365–383.

Turnbull, B.W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society Series B* 38, 290–295.

Voltaire, L., Pirrone, C. and Bailly, D. (2013). Dealing with preference uncertainty in contingent willingness to pay for a nature protection program: a new approach, *Ecological Economics* 88, 76–85.

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

Vossler, C.A. and Kerkvliet, J. (2003). A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum, *Journal of Environmental Economics and Management* 45, 631–649.

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

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

Appendix I
Introduction to referendum

Fictional referendum

Now we come to the most important part of the survey. We would like to ask you to participate in a thought experiment about climate policy in Germany. For this, please read the following information on climate policy in Germany and Europe:

The international community acts on the assumption, that global greenhouse gas emissions must be halved by the year 2050 compared to 1990 emission levels to ensure that the increase in global average temperature will not exceed 2°C by 2050. The member states of the European Union therefore set binding targets to gradually reduce greenhouse gas emissions by 2050.

To achieve the emissions targets by 2020, already implemented and new mitigation measures must be carried out consistently. These include amongst others:

- the energy-related renovation and modernisation of the housing stock
- the further expansion of renewable energies
- the corresponding modification of the power supply structures
- the expansion of emissions trading
- the further increase of energy efficiency

The mitigation measures in Germany have already increased the cost of living and will further increase costs for households in Germany.

Appendix II
Cheap talk script

Before you vote in the referendum, we would like you to think about your current household income and your household expenses. Please note that expenses for climate change mitigation can no longer be used for other things, for example consumption or saving. In addition, your household is already paying for the existing level of climate protection. As electricity consumers, for instance, you pay for the extension of renewable energies and existing gasoline, natural gas and heating oil prices already include parts of the costs of reducing greenhouse gas emissions.

Appendix III
DM question format (MULTILAT)

Fictional referendum

Please imagine the following situation.

The government plans to further increase national greenhouse gas emissions. To this end the national reduction target of 30% less emissions is strengthened to 40% less emissions by 2020 compared to 1990.

> Simultaneously with strengthening the emission target in Germany, the EU would strengthen its emission target from 20% to 30% reduction over the same time period and other industrialized countries like the U.S. and emerging economies such as India and China would arrange to meet mandatory reduction targets in their part.

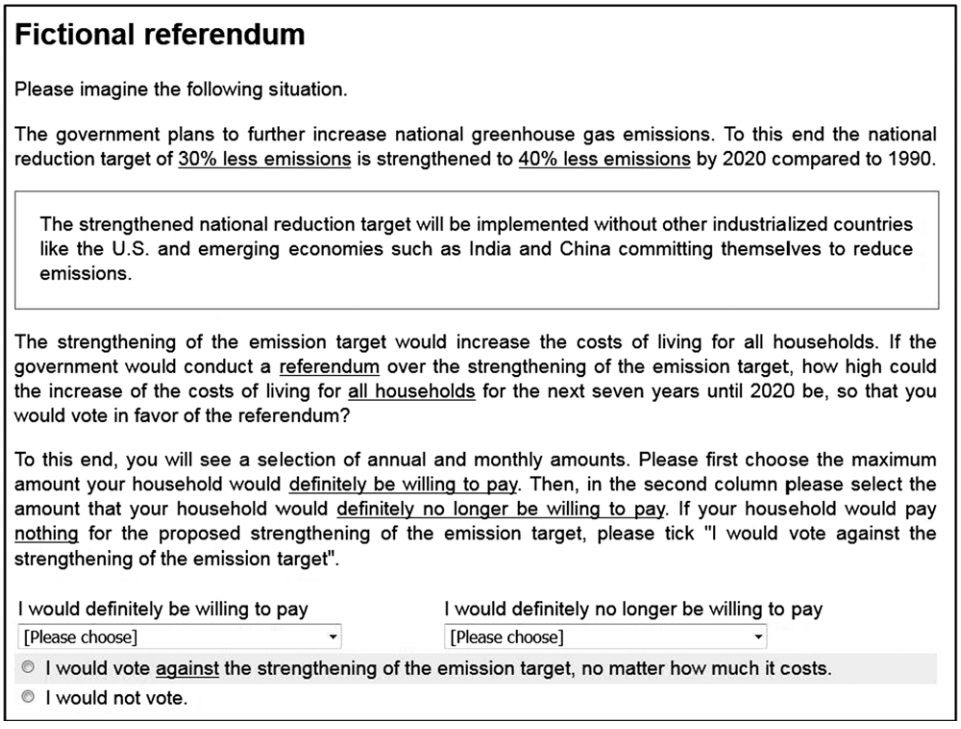
Estimates have shown that due to the strengthened reduction target costs of living for every household will increase by €960 per year or €80 per month.

You can now vote in a referendum about whether the government should undertake the necessary actions to achieve the strengthened emission target. If the majority of participants agrees to the proposal, the costs of living for all households would rise for the next seven years until 2020 by the estimated amount.

How would you vote in such a referendum?

- ○ I support the referendum and would vote YES.
- ○ I support the referendum, but I am not willing to pay this amount and would therefore vote NO.
- ○ I support the referendum, but I cannot afford this amount and would therefore vote NO.
- ○ I support the referendum, but I would vote NO because _____.
- ○ I do NOT support the referendum and would vote NO.
- ○ I would not vote.

Appendix IV
Payment ladder format (UNILAT)



Fictional referendum

Please imagine the following situation.

The government plans to further increase national greenhouse gas emissions. To this end the national reduction target of 30% less emissions is strengthened to 40% less emissions by 2020 compared to 1990.

The strengthened national reduction target will be implemented without other industrialized countries like the U.S. and emerging economies such as India and China committing themselves to reduce emissions.

The strengthening of the emission target would increase the costs of living for all households. If the government would conduct a referendum over the strengthening of the emission target, how high could the increase of the costs of living for all households for the next seven years until 2020 be, so that you would vote in favor of the referendum?

To this end, you will see a selection of annual and monthly amounts. Please first choose the maximum amount your household would definitely be willing to pay. Then, in the second column please select the amount that your household would definitely no longer be willing to pay. If your household would pay nothing for the proposed strengthening of the emission target, please tick "I would vote against the strengthening of the emission target".

I would definitely be willing to pay [Please choose]

I would definitely no longer be willing to pay [Please choose]

I would vote against the strengthening of the emission target, no matter how much it costs.

I would not vote.