CERE Working Paper, 2015:13

Is there a valuation gap? The case of interval valuations

Oben K. Bayrak and Bengt Kriström[1]

November, 2015

The **Centre for Environmental and Resource Economics** (CERE) is an inter-disciplinary and inter-university research centre at the Umeå Campus: Umeå University and the Swedish University of Agricultural Sciences. The main objectives with the Centre are to tie together research groups at the different departments and universities; provide seminars and workshops within the field of environmental & resource economics and management; and constitute a platform for a creative and strong research environment within the field.

[1] Oben K. Bayrak (Corresponding Author), PhD Candidate, Department of Forest Economics, SLU and CERE (Center for Resource and Environmental Economics), Skogsmarksgränd, Umeå, 901 83, +46 727469175, oben.bayrak@slu.se

Bengt Kriström, Professor, Department of Forest Economics, SLU and CERE (Center for Resource and Environmental Economics), bengt.kristrom@slu.se

Department of Economics, Umeå University, S-901 87, Umeå, Sweden

www.cere.se

Is there a valuation gap? The case of interval valuations

Oben K. Bayrak and Bengt Kriström[1]

November, 2015

Abstract

We extend the literature on the willingness-to-pay/willingness-to-accept (WTP/WTA) disparity by testing two hypotheses, distilled from the literature. We also introduce a modified mechanism for eliciting the subjective valuation range if the individual cannot articulate the subjective value as a precise amount confidently. Our key finding is that the disparity disappears under the intervals treatment, suggesting that response format is important, given that earlier experimental studies invariably uses point values (i.e. open ended questions about WTP/WTA). Moreover, for the risky prospect we observe that from their admissible range the buyers state the lower bound as their WTP whereas sellers state the upper bound as their WTA. We conclude that this type of behavior can to some extent explain the observed disparity at least for the risky prospects.

Keywords: Valuation Gap; Imprecise Preferences; Interval Valuation; Willingness to Pay and Accept Disparity, Endowment Effect.

JEL-codes: Q5; C9

[1] Oben K. Bayrak (Corresponding Author), PhD Candidate, Department of Forest Economics, SLU and CERE (Center for Resource and Environmental Economics), Skogsmarksgränd, Umeå, 901 83, +46 727469175, oben.bayrak@slu.se

Bengt Kriström, Professor, Department of Forest Economics, SLU and CERE (Center for Resource and Environmental Economics), bengt.kristrom@slu.se

1. Introduction

The "valuation gap" refers to the empirically found disparity between WTP and WTA. It remains one of the most prominent anomalies in standard economic theory, because we expect that WTP and WTA should be similar if the goods in question have close substitutes and if the income effects are small (Hanemann, 1991). The gap was first documented by mathematical psychologists Coombs et al. (1967) and by Hammack and Brown (1974) in an early contingent valuation study. Knetsch and Sinden (1984) brought the issue into the laboratory using real monetary incentives and found a significant difference between WTP and WTA. Since then, the disparity has been found in an array of studies, including contingent valuation surveys and in field and laboratory experiments for a wide range of goods: e.g. mugs, pens, movie tickets, hunting permits, nuclear waste repositories, foul-tasting liquids, and pathogen-contaminated sandwiches (Horowitz and McConnell, 2002).

The gap has many implications for the application of economic theory, but also for theory proper. For example, if a cost-benefit analysis is conducted on a proposed policy that generates both winners and losers, estimated net benefits will then depend on whether WTA or WTP was used in the assessment. At a more fundamental level, the gap raises questions about the power of standard preference models to explain economic behavior (Braga and Starmer, 2005).

Explanations of what may drive the disparity include the endowment effect which suggests that preferences are reference dependent and losses loom larger than gains. Thus sellers perceive giving away the good as a loss and ask for more as a compensation for their loss (Thaler, 1980). Theorists have also developed alternative models of economic behavior that address the disparity and several other anomalies[2].

Yet, emerging evidence suggests that, under certain types of procedures and settings, the WTP-WTA disparity is smaller than previously observed. Shogren et al. (1994) find that the size of the gap depends on the type of good that is used in the experiments (e.g., mugs, candies, lottery tickets, and tokens). Other researchers find that the disparity declines with trading experience (List 2003; List 2004; Shogren et al. 2001; Loomes et al., 2003). Sayman and Onculer (2005) conducted a meta-analysis of 39 studies and found that studies that employ iterative bidding

[2] Briefly, Prospect Theory (Kahneman and Tversky 1979), Cumulative Prospect Theory (Tversky and Kahneman, 1992), Third Generation Prospect Theory (Schmidt et al., 2008), Rank Dependent Utility Theory (Quiggin, 1982), and Regret Theory (Bell, 1982; Loomes and Sugden, 1982).

exhibit smaller disparities. These findings suggest that experimental design features are critically important. Indeed, in the most recent meta-study, Tuncel and Hammit (2014) find that studies that were published after the first meta-study (Horowitz and McConnell (2002)), exhibit lower WTP-WTA ratios and interpret this as the improvements in study design characteristics. This begs the question of what an "improvement" entails. We suggest two criteria that can be used to assess an experiment:

i. the experimental instructions and procedures are clear to the subject,
ii. the response format is close to the "natural way" that people think about their valuations.

The first item has been covered by Plott and Zeiler (2005), who conducted experiments to control for subject misconceptions about experimental mechanisms, such as the Becker–DeGroot–Marschak mechanism (BDM). Their design employs numerical examples, paid and unpaid training rounds, anonymity of the subjects' identities, and verbal explanations of how to obtain the optimal response. The disparity is not observed for ordinary market goods when procedures to eliminate subjects' misunderstandings about the experimental mechanism are employed: Their result weakened the loss aversion explanation of the disparity. However, Isoni et al. (2011) pointed out that the disparity persists when using lottery tickets, so the issue extends beyond subject misconceptions.

Our second criterion has not yet been sufficiently explored in valuation gap studies in an experimental setting. In the contingent valuation literature, a substantial number of papers have been published on the subject of elicitation mechanisms. One strand of this literature compares open-ended and dichotomous choice formats (Reaves et al., 1999; Loomis et al., 1997). In the open-ended format, subjects are simply asked how much they are willing to pay, whereas in the latter, subjects are asked to accept or reject a pre-selected price. More recent elicitation mechanism allows for respondent uncertainty in various ways; see Mahieu et al. (2014) for a recent survey. In short, experimental studies that find a disparity, invariably uses an open-ended valuation question. This format is not currently much used in contingent valuation, the most important reason being that the response rates are typically low.

The contingent valuation literature rather converged on finding a response format that is allegedly closer to the way that individuals think about their valuations (Brown et al., 1996). For

most individuals, valuation (of the maximum/minimum buying/selling price) is not a routine task. Therefore, asking individuals for precise estimates of their subjective valuations can be cognitively challenging (Mitchell and Carson, 1989), especially for complex and unfamiliar goods (Gregory et al., 1995; Ready et al., 1995). We also know from the psychology literature that when individuals are faced with difficult tasks, they have a tendency to employ heuristics to facilitate them (Shah and Oppenheimer, 2008). For example, McCollum and Miller (1994) found that 44% of the respondents reported $0 due to their inability to provide a precise WTP; even when they indicated a positive attitude towards the good.

If the same behavioral pattern is also present in experiments on disparity, then it might, for example, cause buyers to understate their subjective valuations and cause the observed disparity. A particularly useful alternative mechanism caters for imprecision, without compromising the possibility for an individual to state a precise amount. In this variation, individuals are asked to state interval valuations, in case they are unable to come up with a point[3].

In a related literature which focuses on imprecise preferences, subjects are assumed to have an admissible range of subjective valuations from which they cannot state a precise amount confidently (See Cohen et al. 1987; Butler and Loomes 1988, 2007, 2011; and Morrison). Butler and Loomes (2011) claims that preference imprecision could explain anomalies within Expected Utility Theory. For example, Butler and Loomes (2007)[4] explore imprecision as a way to understand preference reversals. They argue that many individuals' choices and valuations involve a degree of uncertainty or imprecision, and their findings suggest that imprecision explains a significant portion of the preference reversal phenomenon[5].

Interval valuation as a response format is yet to be tested thoroughly in an experimental setting. To the best of our knowledge, Benarjee and Shogren (2014) is the only study that has

[3] Some researchers in the contingent valuation related literature have suggested the use of self-selected intervals in surveys. The basic idea behind self-selected intervals dates to at least Morgan and Small (1992), who suggested them as a way of overcoming "overconfidence" in surveys and to address the anchoring problem. There is also a connection to symbolic data analysis (Billard and Diday, 2007), in which intervals play an important role. Detailed statistical theory for handling this unusual kind of interval censoring has been developed by Belyaev & Kriström (2015).

[4] See Gal (2006) and Neilson (2008) for a theoretical approach to imprecision and empirical studies that can be classified similarly but that used non-incentivized elicitation methods for strength of preference (Dubourg et. al, 1994; Loomes et al., 1997).

[5] They asked the subjects to state their preferences in a series of binary choices in which one option (A) was held constant and the other (B) was adjusted upwards or downwards by $1, depending on the starting point. (In one treatment, they started from $1 and increased, whereas in another treatment they started from a positive payoff of the first lottery and gradually decreased.) In each binary choice problem, the subjects stated which option they chose and selected one of the following phrases that reflected the strength of their decision: definitely prefer A, prefer A but not sure, prefer B but not sure, and definitely prefer B. However, "preference strength elicitation" is not incentivized under their design.

examined bidding behavior in a second-price auction. They found that under the interval valuation scheme subjects behave rationally.

As a simple remedy for the problem observed by McCollum and Miller (1994), we frame the response format as intervals of which the bounds are determined by subjects: if they cannot provide a precise estimate, they are allowed to state an interval for their WTP and WTA, and we test whether the disparity survives under this framing. If individuals are stating some amount lower than they would pay for the good merely because they cannot provide a precise amount, but the experimental design asks them to do so, framing the response format as intervals can make the format closer to the natural way that they think. It can decrease the cognitive burden and make them think more carefully about their valuations, their limits and flexibility of their admissible range (This hypothesis about this is called *Response Format Framing Hypothesis;* RFFH). It is called *framing*, because only the buyer's upper bound and the seller's lower bound are incentivized; the trade is determined by comparing only the incentivized bound with the randomly selected market price. To see this, consider a buyer who states a range and buys the good, if the market price is within or below the stated range. For the seller, trade occurs if the market price is within or above the stated range.

Gregory et al. (1995) found that individuals display a surprisingly large WTP range, and when they are asked to state a single amount as buyers, they are likely to state an amount closer to the middle of their range. As sellers, subjects tend state a point close to the upper bound of their admissible range. This behavior might produce the observed disparity. This hypothesis is here called the Preference *Cloud Hypothesis* (PCH). The Preference Cloud Hypothesis posits that individuals cannot intrinsically determine precise single points, but can identify a range of values for their personal valuation of the good. If the experiment forces them to state a point, they employ a heuristic: buyers state the lower bound while sellers state the upper bound in their admissible range[6]. Therefore, this type of behavior might be the underlying reason of the observed disparity.

[6] Providing a theoretical discussion of the issue is beyond the scope of this paper, see Bayrak and Hey (2015) for a recent theory which asserts that individuals end-up having a range of expected utilities due to their vague understanding of the numerical objective probabilities, and when they are asked to state a single amount as their valuation, they withdraw a single amount from the range depending on their pessimism level. The values withdrawn depends on the role that they are assigned such as buyer and seller, because they determine which bound is worst and best case in a reference dependent way. Therefore in a buying task the worst case is the lower bound of their admissible EU range whereas for a seller the upper bound is the worst case. Since the worst thing that can happen to a buyer is the good that is bought have a low utility, whereas for a seller the converse is true.

2. Experimental Design

We conducted a between-subjects experiment with two treatments: *Points* and *Intervals*. The only difference between two treatments is that in the *Intervals* treatment subjects were allowed to state their valuations in terms of ranges. Subjects are allowed to state single amounts if they prefer. In *Points* only single amounts are allowed, there were the usual two groups: Buyers and Sellers denoted B_p and S_p, respectively. In *Intervals*, we use three groups: Buyers, Sellers, and Buyer-Seller Uncertainty (B_{int}, S_{int}, and BS_{int}, respectively).

In the *Points* treatment: subjects state their offers, and then a market price is determined randomly. If the market price equals or is below the stated offer, a buyer pays the market price and buys the good. For sellers, if the market price equals or is above the stated offer, the seller gets the amount of money equal to the market price and gives away the good. In the B_{int} and S_{int} groups of *Intervals* treatment as we discuss in the introduction, it can be seen as a new type of framing, because only the buyer's upper bound and the seller's lower bound are incentivized; the trade is determined by comparing the incentivized bounds with the randomly selected market price. For the buyer only the upper bound of the stated range is binding. For the seller, trade occurs if the market price is within or above the stated range. Therefore only the lower bound of the range is binding. The only difference between *Points* and *Intervals* is the response format; thus, any difference in the results was because of this feature. We compare the values elicited by B_p, S_p, B_{int}, S_{int} to test the RFFH, (WTP_p, WTA_p, WTP_{int}, WTA_{int}; respectively). If we observe a statistically significant difference between WTP_p and WTA_p but not between WTP_{int} and WTA_{int} we can confirm RFFH.

Testing PCH is not straightforward, we need to compare the point offers with ranges that are elicited in an incentive compatible way. Remember that PCH claims that buyers state the lower bound, whereas sellers state the upper bound of their admissible range, which is the underlying reason for observing the disparity. We elicit the usual point offers in B_p and S_p however we cannot use the ranges elicited in B_{int} and S_{int} because only one bound of those ranges are incentivized. They are only appropriate to test RFFH which is a hypothesis focusing on the framing of the response format.

To accomplish this we developed BS_{int} which is a modified version of BDM, since the subjects learned whether they were buyers or sellers after stating their offers, both the lower and upper bounds were incentivized (See the appendix for details): At the end of the experiment, roles were

determined randomly; the probability of being designated as a buyer is ½ (likewise, the probability of being a seller is ½). If subjects overstate their valuations, there is a 50% chance of being a buyer and a risk of paying an undesirably high amount. If they understate their values, they might be a seller and would have to sell the good for an undesirably low amount.

Table 1: Summary of the Experimental Design

1.Anonymity	Assigning subjects an ID number randomly
2.Instruc-tions	Also read aloud
	Numerical examples to explain optimal response
	Hypothetical Training Round
3.Goods	Four goods with real incentives
´ Good 1	Premium bitter chocolate
Good 2	Created their own package of three cans, from five different flavours of a beverage brand
Good 3	Select one of the ten different flavours of a chocolate brand
Good 4	Lottery ticket: winning 30 SEK with a probability of 0.5, zero otherwise
4.Incentives	Show-up fee of 100 SEK≈$12
	One of the four goods and a market price selected randomly
	Only in BS_{int} group, subject role (buyer, seller) is also selected randomly after value elicitation

The roles were determined after the four tasks were completed using the following procedure: The experimenter wrote "*buyer*" and "*seller*" on two separate pieces of paper, placed them in two separate envelopes, one of them is picked from an opaque bag. In addition, the procedure was explained to the subjects in detail when the instructions were provided.

We recruited the subjects by announcement (flyers and posters) from Umeå University and the Swedish University of Agricultural Sciences (SLU)[7]. In total, 38 subjects participated in points, and 54 subjects participated in intervals, most of whom were master's degree students from various fields of study. The sessions lasted approximately 40 minutes, and the average earnings were 108 SEK[8] (including a 100 SEK show-up fee). Each subject chose an envelope marked with an ID number upon entering the room. We told the subjects to keep these ID cards and to use them to retrieve their earnings after the experiment. The instructions were read aloud, and the

[7] These two universities are very close to each other and can considered the same campus area. Umeå University has over 20,000 students, whereas SLU is a much smaller university.

[8] 1 SEK is approximately 0.15 US Dollars.

participants were instructed not to communicate with each other or react verbally to any events that occurred during the experiment.

In both experiments, following Plott and Zeiler (2005), certain training procedures were employed to minimize or prevent subject misconceptions, i.e., anonymity was ensured and numerical examples were used to explain the mechanism[9] together with examples to show the subjects why stating their true value is the dominant strategy. In addition, the participants were provided with an unpaid training round in which the good was a candy. As indicated in Plott and Zeiler (2005), the provision of paid practice rounds is not an essential procedure: In both treatments; the one with paid practice and with unpaid practice rounds, no disparity is observed in their experiments.

After the training round, the subjects were encouraged to ask questions. They wrote their questions on pieces of paper and raised their hands; the experimenter silently read the questions and answered them by writing on the same piece of paper.

The practice round was followed by four tasks (goods), and the subjects were told that these four tasks had an equal chance of being selected and the payoffs will be determined according to the randomly selected task.

In task 1, the good was a premium bitter chocolate. In task 2, the subjects were given a list of five different flavors (regular, light, zero, vanilla, and cherry) of a nonalcoholic beverage brand. They were asked to create any package of three cans; thus, they were allowed to mix and match among the five types. Then, they stated offers for their created package. The good in task 3 was similar: In that case, 10 different flavors of the same brand of chocolate were provided, and we asked the subjects to select one of the flavors.

Goods 2 and 3 are homogenous for all subjects, since prices in local shop does not vary with the flavors and these two goods can be considered as vouchers providing the right to choose a favorite flavor. We included these to contribute the literature by re-examining the disparity with a new type of goods and to have stronger results because the endowment effect might be stronger for these goods since the subjects picked their favorite flavors; thus, they might have felt more attached to these goods.

[9] The numbers that are used in the examples are completely unrelated to the possible range of prices in the experiment to avoid any anchoring effects (e.g., 1000–1020 SEK, whereas the experiment market price can be between 1 and 30 SEK). The numerical examples were part of the written instructions provided, and they were explained on a board.

The participants were not provided with any information about market prices during the experiment. The prices of the goods in tasks 1, 2, and 3 were 19 SEK, 24 SEK, and 22 SEK, respectively, at a local shop.

Finally, the fourth good was a lottery ticket with the following prospects: winning 30 SEK with a probability of 0.5 and winning nothing with a probability of 0.5. The lottery outcome was determined by using one hundred ping-pong balls that were numbered from 1 to 100 and placed in an opaque bag. At the end of the experiment, a ball was selected from the bag. If the number on the ball was 50 or below, the lottery paid 30 SEK; otherwise zero.

After a task had been completed, the response sheets for that task were collected, and the next response sheet was handed out to prevent cheating. The subjects were given the goods and told to examine them before recording their offers. The sellers were told that they owned the good; the buyers were told that they could inspect the good but that they did not own it.

When all four tasks were completed, one task was chosen as "real," and the market price was drawn for that task. In all of the tasks, including the unpaid training round, the subjects were told that the market price would be randomly selected from a range of 1 to 30 SEK using the ping-pong balls. The market price was determined by picking one ball out of 30, each with a single price written on it. To avoid any bias that might result from the potential market price range, we used the 1–30 SEK range as a potential market price range for all of the tasks (see Bohm et al. (1997) for a comprehensive discussion of this issue).

At the end of the experiments, the subjects were given both a questionnaire requesting demographic information and test of their understanding of the instructions. Only the subjects who answered all quiz questions correctly were included in the analysis.

3. Results

Summary statistics are reported in Table 2. The second column indicates the percentage of subjects that preferred to state intervals. Except for the BS_{int} group for good 3, a majority preferred intervals.

Overall, statistical tests confirm both PCH and RFFH. To test the PCH, we should look at the first set of results (Table 3): For good 2, a ratio of 1.20 is significant with a p-value of 0.0449; the W statistic is 86.0 according to the Wilcoxon-Mann-Whitney rank sum test. A test of medians resulted in a Pearson $\chi 2$ test statistic of 0.6461 ($p = 0.421$). For good 4, the ratio is 1.82, which is

significant with a p-value of 0.0014 and a W statistic of 51.0. The median test results are also in line with this result. Good 2 exhibits significance (p-value=0.0449) but just barely. Therefore we focus on good 4 (p-value=0.0014) and compare the point bids with the bounds that were elicited in the BS_{int}. The second set of results presents these comparisons, showing that the Wilcoxon-Mann-Whitney tests support our hypothesis: We cannot reject the hypothesis that the mean *points* WTP in *and* the mean lower bound of BS bids were drawn from the same distributions as the mean WTA in *points* and the upper bound of BS bids.

Table 2: Summary Statistics

	Treatment	Mean $_L$	Mean $_U$	Median $_L$	Median $_U$	σ_L	σ_U
Good 1							
(Premium bitter	B_{int} (71%)	17.3	20.2	18.0	22.0	6.6	5.4
chocolate)	S_{int} (83%)	18.9	21.8	20.0	22.0	4.5	4.7
	BS_{int} (62%)	14.9	17.8	15.0	17.8	3.2	3.9
	B_p	14.3		15.0		6.3	
	S_p	13.5		14.0		5.1	
Good 2							
(3 cans of Coke)	B_{int} (65%)	15.0	18.4	15.0	18.0	9.4	9.5
	S_{int} (56%)	18.6	20.5	17.5	20.0	7.2	7.9
	BS_{int} (62%)	14.6	17.5	15.0	18.0	6.7	6.6
	B_p	13.9		15.0		8.0	
	S_p	19.1		18.0		6.8	
Good 3							
(Chocolate)	B_{int} (75%)	13.3	16.6	13.5	15.0	5.6	4.5
	S_{int} (50%)	19.0	20.7	17.0	20.5	6.5	5.7
	BS_{int} (46%)	14.3	16.2	15.0	16.0	5.3	5.7
	B_p	16.4		15.0		6.4	
	S_p	19.3		19.5		6.8	
Good 4							
(Lottery ticket)	B_{int} (59%)	11.8	14.5	10.0	15.0	8.7	9.3
	S_{int} (61%)	14.3	17.2	15.5	18.0	6.9	7.4
	BS_{int} (54%)	12.5	18.2	14.0	15.0	6.2	6.7
	B_p	12.5		11.0		5.6	
	S_p	20.2		20.0		7.2	

Notes: The subscripts *L* and *U* denote the *lower* and *upper* bound, respectively. σ denotes the standard deviation. The values within the squares are the incentivized ones. In the treatment column, the percentages in parentheses denote the portion of subjects stating a range of values for the specific task. Sample size for each treatment: B_{int}=17, S_{int}=18, BS_{int}=13, B_p=19, S_p=14.

To examine the support for RFF, we look for the existence of the WTA-WTP disparity in *points* and its absence for *intervals*. For the *Points,* we observed a significant disparity for good 2 (p=0.0449) and good 4 (p=0.0014) (beverage and lottery ticket respectively). For the *intervals*, test results comparing the incentivized bounds ($3.S_{int}^{L}/B_{int}^{U}$) suggests that the difference is not statistically different.

Table 3: Statistical Test Results

		Wilcoxon-Mann-Whitney rank sum test (Null hypothesis: identical distributions)			Median test (Null hypothesis: identical medians)		
	Ratio[a]	W	*p*-value	Conclusion (α = .05)	Pearson χ^2	*p*-value	Conclusion (α = .05)
1. S_p / B_p							
Good 1	0.93	141.5	0.6299	Cannot reject null	0.7599	0.383	Cannot reject null
Good 2	1.20	86.0	0.0449	Reject null	0.6461	0.421	Cannot reject null
Good 3	1.30	97.5	0.1005	Cannot reject null	0.6461	0.421	Cannot reject null
Good 4	1.82	51.0	0.0014	Reject null	6.2307	0.013	Reject null
2. Good 4[b]							
B_p / BS_{int}^{L}	0,79	115.5	0.7699	Cannot reject null	0.0000	1.000	Cannot reject null
B_p / BS_{int}^{U}	0,73	64.0	0.0211	Reject null	1.0796	0.299	Cannot reject null
S_p / BS_{int}^{L}	1,43	145.0	0.0087	Reject null	4.8048	0.028	Reject null
S_p / BS_{int}^{U}	1,33	111.5	0.3233	Cannot reject null	0.3426	0.558	Cannot reject null
3. $S_{int}^{L} / B_{int}^{U}$							
Good 1	0.91	176.5	0.7871	Cannot reject null	0.2447	0.621	Cannot reject null
Good 2	0.97	168.0	0.6959	Cannot reject null	0.0340	0.854	Cannot reject null
Good 3	1.13	122.5	0.1594	Cannot reject null	1.4138	0.234	Cannot reject null
Good 4	1.03	148.0	0.4407	Cannot reject null	0.0340	0.854	Cannot reject null

[a] Median ratios. [b] Two sided

In order to explore the power of our statistical tests we used the method of Plott and Zeiler (2005).We test the null hypothesis of WTA=2·WTP for the results obtained in the *intervals* treatment (See Table 4).

Table 4: Power of the Tests

Goods	T-test (Unequal Variances)		Wilcoxon-Mann-Whitney rank-sum test	
	T	p-value	z	p-value
1	-7.5642	0.0000	4.810	0.0000
2	-3.6934	0.0007	3.255	0.0011
3	-5.4225	0.0000	4.437	0.0000
4	-3.0386	0.0032	2.267	0.0234

The reason for multiplication by two is the same that Plott and Zeiler (2005) suggested. In the previous literature several authors claim that WTA is twice the WTP (e.g. Dubourg et al., 1994 and Knetsch et al., 2001). A t-test assuming unequal variances led to a rejection of the null in favor of the alternative, WTA < 2·WTP for all goods (See Table 4 first two columns). A two-sample Wilcoxon-Mann-Whitney rank-sum test gives the same result (See Table 4 last two columns). It also should be noted that, although the ratio of WTA to WTP is not exactly one, the difference in *Intervals* is not statistically significant.

4. Concluding Remarks

The unusually large difference between WTA and WTP, in cases where theory predicts little or no difference, remains one of the most studied issues in experimental economics. It has also attracted significant interest from theorists who have tried various different approaches to explain the disparity. Empirical tests have invariably relied on eliciting points, following the lead from standard consumer theory. As we show in this paper, allowing subjects to state their sentiments using any interval on the line (of which a point is a special case) essentially removes the disparity in several cases. This has several implications for the design of experiments, but also for the development of new theory that can cater for imprecision in a more substantive way compared to standard consumer theory. Indeed, introspection and evidence from a large number of areas reveals that intervals are used in daily communication to a very significant extent. Our brains are not necessarily configured to make quantitative assessments in terms of points; ordinarily we tend to use intervals in many estimation tasks in daily life (Zimmer, 1984). Consequently, using intervals in experimental settings follows rather naturally from salient observations of human behavior. Furthermore, while intervals are used routinely in many surveys today, they are based on bracketing, which is altogether a different elicitation mechanism that inter alia introduces a risk for anchoring. Thus, our research touches upon fundamental issues in survey research even though we explore the use of incentivized mechanisms, which typically would infeasible in a consumer survey.

Morrison (1998) observed a large gap between the two ranges; lower bound of WTA being more than one and a half times the upper bound of WTP. However, he did not use any procedures to minimize the subject misconceptions raising doubts about the reliability of his results. Our

results suggest that Plot and Zeiler procedures (especially the part which explains the subjects how to find their optimal offers, see instructions for details), allowing subjects to state intervals lead them think more carefully about the task and their valuations, decrease tendency of biases and heuristics. Sayman and Onculer (2005) found that the disparity is lower in an iterative setting; the sequential process helps subjects to discover their optimal responses. In short, our results suggests that preference imprecision should not be discarded as a potential explanation of the observed anomalies.

Banerjee & Shogren (2014) is the only published paper we are aware of that uses a similar set-up. Consequently, many questions are left to be explored in more detail. For example, why do we observe a disparity for lottery tickets but not for ordinary market goods when we ask for single amounts? How do individuals form admissible ranges? Why do buyers/sellers state different bounds? Thus, the area is fertile ground for development of new theory and additional testing. This could lead to an improved understanding of a long-standing controversy regarding the WTA-WTP disparity and potentially to the development of novel designs of survey instruments. Because the bulk of empirical research in e.g. social science is based on surveys, we do believe that there are good reasons to further explore the elicitation mechanisms studied in this paper.

REFERENCES

Adamowicz, W. L., Bhardwaj V., Macnab B., 1993. Experiments on the Difference between Willingness to Pay and Willingness to Accept. Land Economics 69 (4): 416–27.

Ariely, D., Loewenstein G., Prelec, D., 2003. 'Coherent Arbitrariness': Stable Demand Curves Without Stable Preferences. The Quarterly Journal of Economics 118 (1): 73–106. doi:10.1162/00335530360535153.

Banerjee, P., Shogren, J.F., 2014. Bidding Behavior given Point and Interval Values in a Second-Price Auction. Journal of Economic Behavior & Organization 97: 126–37.

Bell, D.E., 1982. Regret in Decision Making under Uncertainty. Operations Research 30 (5): 961–81.

Belyaev, Y., Kriström, B., Approach to statistical analysis of self-selected interval data. Technical report, CERE, 2010. CERE Working Paper, 2010:2.

Belyaev, Y., Kriström, B., Two-step approach to self-selected interval data in elicitation surveys. Technical report, CERE, 2012. CERE Working Paper, 2012:10.

Belyaev, Y., Kriström, B., 2015. Analysis of survey data containing rounded censoring intervals. Informatics and Applications, 9, 3: 2-16

Billard, L., Diday, E., 2007. Symbolic Data Analysis: Conceptual Statistics and Data Mining (Wiley Series in Computational Statistics). John Wiley & Sons.

Bohm, P., Lindén J., Sonnegård, J., 1997. Eliciting Reservation Prices: Becker-DeGroot-Marschak Mechanisms vs. Markets. The Economic Journal 107 (443): 1079–89.

Boyce, R.R., Brown, T.C., McClelland, G.H., Peterson, G.L., Schulze, W.D., 1992. An Experimental Examination of Intrinsic Values as a Source of the WTA-WTP Disparity. The American Economic Review 82 (5): 1366–73.

Boyle, K. J., Johnson, F. R., McCollum, D.W., 1997. Anchoring and Adjustment in Single-Bounded, Contingent-Valuation Questions. American Journal of Agricultural Economics 79 (5): 1495–1500. doi:10.2307/1244370.

Braga, J., and Starmer, C., 2005. Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis. Environmental and Resource Economics 32 (1): 55–89.

Brookshire, D.S., Coursey, D.L., 1987. Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures. The American Economic Review, 554–66.

Brown, Thomas C., Patricia A. Champ, Richard C. Bishop, and Daniel W. McCollum. 1996. "Which Response Format Reveals the Truth about Donations to a Public Good?" Land Economics 72 (2): 152–66. doi:10.2307/3146963.

Butler, D. J., & Loomes, G. (1988). Decision difficulty and imprecise preferences. Acta Psychologica, 68, 183–196.

Butler, D.J., Loomes, G.C., 2007. Imprecision as an Account of the Preference Reversal Phenomenon. The American Economic Review 97 (1): 277–97.

Butler, David, and Graham Loomes. 2011. "Imprecision as an Account of Violations of Independence and Betweenness." Journal of Economic Behavior & Organization 80 (3): 511–22. doi:10.1016/j.jebo.2011.05.008.

Cohen, M., Jaffray, J.-Y., & Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. Organizational Behavior and Human Decision Processes, 39, 1–22

Coombs, C.H., Bezembinder, T.G., Goode, F. M., 1967. Testing Expectation Theories of Decision Making without Measuring Utility or Subjective Probability. Journal of Mathematical Psychology 4 (1): 72–103.

Coursey, D.L., Hovis, J.L., Schulze, W.D., 1987. The Disparity between Willingness to Accept and Willingness to Pay Measures of Value. The Quarterly Journal of Economics, 679–90.

Dubourg, W.R., Jones-Lee, M.W., Loomes, G., 1994. Imprecise Preferences and the WTP-WTA Disparity. Journal of Risk and Uncertainty 9 (2): 115–33.

Dubourg, W. R., Jones-Lee, M. W., & Loomes, G. (1997). Imprecise preferences and survey design in contingent valuation. Economica, 64, 681–702.

Gal, D., 2006. A Psychological Law of Inertia and the Illusion of Loss Aversion. Judgment and Decision Making 1 (1): 23–32.

Gregory, Robin, Sarah Lichtenstein, Thomas C. Brown, George L. Peterson, and Paul Slovic. 1995. "How Precise Are Monetary Representations of Environmental Improvements?" Land Economics 71 (4): 462–73. doi:10.2307/3146711.

Håkansson, C., 2008. A New Valuation Question: Analysis of and Insights from Interval Open-Ended Data in Contingent Valuation. Environmental and Resource Economics 39 (2): 175–88.

Hammack, J., Brown G.M., 1974. Waterfowl and Wetlands: Toward Bioeconomic Analysis. Resources for the Future Washington, DC.

Hanemann, W.M., 1991. Willingness to Pay and Willingness to Accept: How Much Can They Differ? The American Economic Review 81 (3): 635–47.

Horowitz, J.K., McConnell, K.E., 2002. A Review of WTA/WTP Studies. Journal of Environmental Economics and Management 44 (3): 426–47.

Irwin, J.R., 1994. Buying/selling Price Preference Reversals: Preference for Environmental Changes in Buying versus Selling Modes. Organizational Behavior and Human Decision Processes 60 (3): 431–57.

Isoni, A., Loomes, G., Sugden, R., 2011. The Willingness to Pay—willingness to Accept Gap, the 'endowment Effect,' Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Comment. The American Economic Review 101 (2): 991–1011.

Kahneman, D., Knetsch, J.L., Thaler, R.H., 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. Journal of Political Economy, 1325–48.

Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica: Journal of the Econometric Society, 263–91.

Knetsch, J.L., Sinden, J.A., 1984. Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. The Quarterly Journal of Economics 99 (3): 507–21. doi:10.2307/1885962.

Knetsch, J.L., Tang, F., Thaler, R.H., 2001. The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing? Experimental Economics 4 (3): 257–69.

List, J.A., 2003. Does Market Experience Eliminate Market Anomalies? The Quarterly Journal of Economics 118 (1): 41–71.

———. 2004. Neoclassical Theory versus Prospect Theory: Evidence from the Marketplace. Econometrica 72 (2): 615–25.

Loomes, G., and Dubourg, W.R., Jones-Lee, M.W., 1997. Imprecise Preferences and Survey Design in Contingent Valuation. Economica 64 (256): 681–702.

Loomes, G., Starmer, C., Sugden, R., 2003. Do Anomalies Disappear in Repeated Markets? The Economic Journal 113 (486): C153–66.

Loomes, G., Sugden, R., 1982. Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. The Economic Journal, 805–24.

Loomis, John, Thomas Brown, Beatrice Lucero, and George Peterson. 1996. "Evaluating the Validity of the Dichotomous Choice Question Format in Contingent Valuation." Environmental

and Resource Economics 10 (2): 109–23. doi:10.1023/A:1026403916622.

Mahieu, Pierre-Alexandre, Pere Riera, Bengt Kriström, Runar Brännlund, and Marek Giergiczny. “Exploring the Determinants of Uncertainty in Contingent Valuation Surveys.” Journal of Environmental Economics and Policy. Vol. 3.2014, 2, P. 186-200 2, (7). Routledge, 2014.

Manski, C.F. 2003. Partial Identification of Probability Distributions. Springer. http://link.springer.com/content/pdf/10.1007/b97478.pdf.

———. 2009. Identification for Prediction and Decision. Harvard University Press.

Manski, C.F. and Molinari, F 2010 Rounding Probabilistic Expectations in Surveys. Journal of Business and Economic Statistics, bf 28, 219-231.

McCollum, D. W., and S. M. Miller. "Alaska Voters: Their Wildlife Viewing Trip Characteristics and Economics." Alaska Department of Fish and Game, Division of Wildlife Conservation, Anchorage (1994).

Mitchell, Robert Cameron, and Richard T. Carson. Using surveys to value public goods: the contingent valuation method. Resources for the Future, 1989.

Morgan, M.G., Small, M., 1992. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press.

Morrison, G. C. (1998). Understanding the disparity between WTP and WTA: Endowment effect, substitutability, or imprecise preferences? Economics Letters, 59, 189–194.

Neilson, W., McKee, M., Berrens, R.P., 2008. Value and Outcome Uncertainty as Explanations for the WTA vs WTP Disparity: Theory and Experimental Evidence. Working Paper 08-07. Department of Economics, Appalachian State University. https://ideas.repec.org/p/apl/wpaper/08-07.html.

Neilson, W.S., 2008. Impulsive Actions and Agonizing Decisions. Citeseer. http://web.utk.edu/~wneilson/ImpulsiveActions.pdf.

Peters, E., Slovic, P., Gregory, R., 2003. The Role of Affect in the WTA/WTP Disparity. Journal of Behavioral Decision Making 16 (4): 309–30.

Plott, C.R., Zeiler, K., 2005. The Willingness to Pay-Willingness to Accept Gap, the ‘Endowment Effect,’ Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. The American Economic Review 95 (3): 530–45.

Johansson, P.O., Kriström, B.,2012. The economics of Evaluating Water Projects: Hydroelectricity vs. other uses. Springer Verlag.

Johansson, P.O., Kriström, B., 2013. Evaluating Water Projects: Cost-Benefit Analysis Versus Win-Win Approach. Springer Verlag.

Morgan, M.G., Henrion, M. and Small, M 1992 Uncwertainty, a Guie to Dealing with Uncertainty and in Quantitative Risk and Policy Analysis, Cambridge University Press, Cambridge

Price, C.R., Sheremeta, R.M., 2011. Endowment Effects in Contests. Economics Letters 111 (3): 217–19. doi:10.1016/j.econlet.2011.02.003.

Quiggin, J., 1982. A Theory of Anticipated Utility. Journal of Economic Behavior & Organization 3 (4): 323–43. doi:10.1016/0167-2681(82)90008-7.

Ready, Richard C., John C. Whitehead, and Glenn C. Blomquist. 1995. "Contingent Valuation When Respondents Are Ambivalent." Journal of Environmental Economics and Management 29 (2): 181–96. doi:10.1006/jeem.1995.1040.

Reaves, Dixie Watts, Randall A. Kramer, and Thomas P. Holmes. 1999. "Does Question Format Matter? Valuing an Endangered Species." Environmental and Resource Economics 14 (3): 365–83. doi:10.1023/A:1008320621720.

Sayman, S., Öncüler, A., 2005. Effects of Study Design Characteristics on the WTA–WTP Disparity: A Meta Analytical Framework. Journal of Economic Psychology 26 (2): 289–312. doi:10.1016/j.joep.2004.07.002.

Schmidt, U., Starmer, C., Sugden, R., 2008. Third-Generation Prospect Theory. Journal of Risk and Uncertainty 36 (3): 203–23. doi:10.1007/s11166-008-9040-2.

Shah AK, Oppenheimer DM. 2008. Heuristics made easy: an effort-reduction framework. Psychol. Bull. 137:207–22

Shogren, J.F., Cho, S., Koo, C., List, J., Park, C., Polo, P., Wilhelmi, R., 2001. Auction Mechanisms and the Measurement of WTP and WTA. Resource and Energy Economics 23 (2): 97–109.

Shogren, J.F., Shin, S.Y., Hayes, D.J., Kliebenstein, J.B., 1994. Resolving Differences in Willingness to Pay and Willingness to Accept. The American Economic Review, 255–70.

Thaler, R., 1980. Toward a Positive Theory of Consumer Choice. Journal of Economic Behavior & Organization 1 (1): 39–60.

Tunçel, Tuba, and James K. Hammitt. 2014. "A New Meta-Analysis on the WTP/WTA Disparity." Journal of Environmental Economics and Management 68 (1): 175–87.

doi:10.1016/j.jeem.2014.06.001.

Tversky, A., Kahneman, D., 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty 5 (4): 297–323.

Zimmer, Alf C. 1984. “A Model for the Interpretation of Verbal Predictions.” International Journal of Man-Machine Studies 20 (1): 121–34.

APPENDIX

The treat used for the BS_{int} group in *intervals* is a modified BDM, in which the lower and upper bounds were incentivized since the subjects learned whether they were buyers or sellers after stating their offers. The roles were determined after the four tasks were completed by using the following procedure: The experimenter wrote "*buyer*" and "*seller*" on two separate pieces of paper, placed them in two separate envelopes, and one of them was picked from an opaque bag. In addition, the procedure was explained to the subjects in detail when the instructions were provided.

The probability of being designated as a buyer is ½ (likewise, the probability of being a seller is ½). If subjects overstate their valuations, there is a 50% chance of being a buyer and, thus, of paying an undesirably high amount. If they understate their values, they would have to sell the good for an undesirably low amount (if seller). We wish to be careful about the assumptions made about this treatment. We do not argue that this procedure eliminates buyer-seller bias; however, it represents a first step toward developing such a mechanism because, to the best of our knowledge, no other study has introduced a method to minimize buyer-seller bias. We tested the PCH by comparing the bounds that were elicited in BS_{int} with the point offers that were elicited in *Points* treatment.

Consider three possible cases or groups of people to understand how BS mechanism works and what their optimal strategy under BS:

i. Individuals who have a precise estimate of their WTA and WTP and they exhibit no endowment effect therefore behave according to the Standard Economic Theory: The optimal response for them is obviously to state the precise estimate as a single point and they are allowed to do so in BS. Note that for this type the WTP equals WTA.

ii. Individuals who have a precise estimate of WTA and WTP but they exhibit loss aversion, therefore their WTA is higher than their WTP. For these type of subjects, the best strategy for them is to state the mid-point of the two: To see this more formally consider the following: Selling utility of good X due to loss aversion is $u(\lambda X)$, where λ is the loss aversion parameter and buying utility is $u(X)$. Therefore WTP equals $u(X)$, whereas WTA equals $u(\lambda X)$. So far is stand-

ard in studies which explain endowment effect (WTA-WTP disparity) with loss aversion concept. However, under BS mechanism, individual does not know whether his or her role is buyer or seller in advance (both is equally likely, determined by a random mechanism). Optimal offer $u(offer^*)$ under this setting is given by:

$$u\left(offer^*\right) = \tfrac{1}{2}u\left(\lambda X\right) + \tfrac{1}{2}u\left(X\right) = \tfrac{1}{2}u\left(WTA\right) + \tfrac{1}{2}u\left(WTP\right)$$

Note that for type ii individuals, an optimal offer does not guarantee a positive payoff in all cases: Consider an individual who has a WTA of 10 and WTP of 5, thus states 7.5. Now suppose the randomly selected market price is 8 and the individual is designated as seller, randomly. Thus, trade occurs: the individual sells the good for 8 which is lower than 10 (WTA). However stating the mid-point is still optimal: optimal does not mean that the payoff will be positive in all states of the world; it means it is the best strategy among the possible ones.

iii. Now consider the case in which individual cannot come up with a precise estimate of his or her subjective value but a range and individual cannot confidently state a single amount from this range (See Luce 1956 for a theoretical discussion).

Suppose individual comes up with a range of values between 5 and 10 dollars but cannot state one of them confidently. Now consider the subcases in which individual misstates his/her true subjective valuation range:

a. Overstating the lower bound: Suppose the individual states a range between 7 and 10 dollars; if the individual is randomly assigned as being a seller at the end of the experiment, the individual loses the chance to sell the good for prices between 5 and 7 dollars and remember that these are inside the true subjective valuation range (5-10 dollars).

b. Understating the lower bound: Suppose the individual states a range between 3 and 10 dollars. If the individual is randomly assigned as being a seller at the end of the experiment and the market price is randomly determined as some amount between 3 and 5 dollars, then the individual sells the good for an undesirably low price. Note that the true range is between 5 and 10 dollars.

c. Overstating the upper bound: Suppose the individual states a range between 5 and 12 dollars; if the individual is assigned as a buyer at the end of the experiment and the market price is between 10 and 12 dollars, individual has to buy the good for an undesirably high price.

d. Understating the upper bound: Suppose the individual states a range between 5 and 8 dollars; if the individual is assigned as a buyer at the end of the experiment and the market price is between 8 and 10 dollars, individual misses the chance to buy the good for these prices which are inside the individual's acceptable range.

Looking at i and ii, we can say that the individuals who stated a single point belongs to these two groups thus having precise preferences which can be compatible with standard economic theory or the non-standard ones which incorporate loss aversion. Therefore the responses in the form of the intervals belong to the individuals which can be categorized as the third group explained above. Because of this we conduct the test of PCH by eliminating the point responses too, but the results still confirm PCH.