European Economic Review 78 (2015) 120–128

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

The willingness to pay–willingness to accept gap: A failed replication of Plott and Zeiler

Dietmar Fehr [a], Rustamdjan Hakimov [a], Dorothea Kübler [a,b]

[a] WZB Social Science Center Berlin, Reichpietschufer 50, 10785 Berlin, Germany
[b] Technical University Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany

ARTICLE INFO

Article history:
Received 11 March 2015
Accepted 22 May 2015
Available online 6 June 2015

JEL classification:
C91
D03

Keywords:
WTP–WTA gap
Endowment effect
BDM mechanism
Misconceptions
Replication

ABSTRACT

The well-known willingness to pay–willingness to accept (WTP–WTA) gap refers to the observation that individuals attach a higher value to objects they own (WTA) than to objects they do not own (WTP). We report on experiments to re-investigate the possibility that the WTP–WTA gap arises from subject misconceptions due to experimental procedures as suggested by Plott and Zeiler (2005). The contribution of this paper is two-fold: first, we attempt to replicate the findings by Plott and Zeiler that the WTP–WTA gap disappears when using procedures that are aimed at reducing misconceptions, such as extensive training and practice rounds for the BDM mechanism. However, we fail to do so as the WTP–WTA gap persists in the main task where subjects state their WTA or WTP for a mug. Second, we use the paid practice rounds to identify subjects without apparent misconceptions and find that also for those subjects who never make dominated choices in the lottery tasks, the WTP–WTA gap in the mug task exists. Thus, we find no evidence of the idea that subject misconceptions are the main source of the WTP–WTA gap.

1. Introduction

The disparity between the willingness to accept (WTA) and willingness to pay (WTP) has long been considered as one of the most firmly established results in behavioral economics. It refers to the observation that a person's perceived value of an item is often higher when the person owns the item than when she does not own it. Thus, previous experiments suggest that being endowed with an object makes it harder to part with it (higher WTA) compared to the willingness to pay for an object one does not own (WTP). This effect is unsettling for economists, because it implies that trade is hampered by a status quo bias. The Coase Theorem breaks down and procedures matter for outcomes in a non-trivial way, which has important implications for the design of economic institutions. A large literature dating back to the 1960s has elaborated on the "endowment effect" or WTP–WTA gap (see Ericson and Fuster, 2014, for a recent survey).[1]

Two papers by Plott and Zeiler (2005, 2007) question the validity of much of the earlier discussions. They demonstrate how the experimental procedures that are typically used to show the "endowment effect" or WTP–WTA gap contribute to or

E-mail addresses: dietmar.fehr@wzb.eu (D. Fehr), rustamdjan.hakimov@wzb.eu (R. Hakimov), dorothea.kuebler@wzb.eu (D. Kübler).

[1] The term "endowment effect" was introduced by Thaler (1980) as a description of the tendency of individuals to underweight opportunity costs. In a seminal contribution, Kahneman et al. (1990) provided experimental evidence of the "endowment effect" in an environment which carefully controlled for alternative explanations. Subsequently, their contribution sparked not only a series of new experiments that investigate the robustness of the "endowment effect", but also theoretical contributions that try to reconcile the various findings (e.g., Köszegi and Rabin, 2006), provide evolutionary foundations (e.g., Huck et al., 2005) or use standard economic arguments to explain the gap (e.g., Hanemann, 1991).

http://dx.doi.org/10.1016/j.euroecorev.2015.05.006

even generate the effect. Plott and Zeiler (2005) [henceforth PZ] show that the WTP–WTA gap disappears when procedures are employed that are aimed at reducing the misconceptions of subjects.[2] In the absence of a theory or a clear definition of misconceptions, PZ propose that these procedures implicitly contain theories of misconceptions.[3] Most importantly, PZ use extensive training and elicit subjects' valuations for lotteries to familiarize subjects with the Becker–DeGroot–Marschak (BDM) mechanism (Becker et al., 1964). Only after this, subjects proceed to the task that is used to explore the WTP–WTA gap for mugs. While PZ argue that the data from these paid practice rounds should not be used to measure a WTP–WTA gap because they are contaminated by their training purpose, Isoni et al. (2011) [henceforth ILS] document the presence of a WTP–WTA gap in PZ's lottery tasks.[4] In an effort to add more control to the practice rounds than in PZ, ILS run additional experiments and document a valuation gap in their lottery task data as well. Consistent with PZ, though, they find no gap in their classic mug experiment. Although both PZ and ILS show that the WTP–WTA gap in the classic mug experiment disappears if subjects receive extensive instructions and training, the presence of a gap in the practice rounds in both studies raises the question of whether misconceptions are really responsible for the gap. In fact, the data for the paid practice rounds in ILS reveal substantial heterogeneity in the valuations of the lotteries, which suggests that misconceptions need not be present for all subjects.

We aim to shed more light on the important but puzzling findings of PZ and ILS and conduct a series of new experiments. Our main idea is that if misconceptions about the BDM procedure indeed cause a WTP–WTA gap then we should be able to see a difference in the gap between those subjects who understand the BDM procedure well and those who do not understand it. In particular, we expect to see no WTP–WTA gap for subjects who do not suffer from misconceptions about the BDM mechanism. To identify these subjects we add two new lotteries and exploit existing features of the 14 lottery tasks employed by PZ. This allows us to pin down irrational behavior in the valuation of an uncertain monetary outcome. Thereby, we are able to separate subjects with misconceptions about the BDM mechanism from subjects who are unlikely to lack a good understanding of the BDM mechanism.

In line with PZ and ILS, we find a strong WTP–WTA gap in the lottery tasks. In contrast to PZ and ILS, however, the gap does not disappear in the main task where subjects submit their WTA or WTP for a mug (or USB stick). We find substantial heterogeneity in the subjects' understanding of the elicitation mechanism. Although the procedures of PZ (and ILS) eliminate misconceptions about the BDM mechanism in at least some of our subjects, we find that a sizeable share of subjects still lack an adequate understanding of the mechanism even after extensive training and several paid practice rounds. Importantly though, we find a WTP–WTA gap for both groups of subjects, those who do not display any signs of misconceptions and those who clearly lack a full understanding of the BDM mechanism. More specifically, subjects whom we classify as rational, based on the lottery rounds, show an equally strong gap as subjects without a good understanding of the BDM. Thus, we make two contributions to the literature. First, we replicate the experiments by PZ and ILS, but observe a WTP–WTA gap in the valuations for a commodity. Second, we provide evidence that misconceptions about the BDM mechanism are unlikely to be the main driving force of the WTP–WTA asymmetry.

Not replicating a result is unsettling, especially since both PZ and ILS found that the WTP-WTA gap disappears in the mug task.[5] However, our findings are in line with Koh and Wong (2012), who attempt to replicate the experiments of PZ without using paid practice rounds but relying on a series of Yes/No questions to explain how to determine the WTP and WTA. They find a WTP–WTA gap that is insignificant in two small samples but becomes significant in the pooled sample. Moreover, they show that when strengthening the reference states of possession or non-possession while using the PZ procedures to clarify the properties of the BDM mechanism, the WTP–WTA gap remains significant. This indicates that the results by PZ are at least in part due to a weakening of the reference state and not only to subject misconceptions. In contrast, our findings reveal that the weaker reference state of the PZ procedures has no bearing on the existence of the WTP–WTA gap.

While writing up our results, we became aware of the contemporaneous and independent work of Bartling et al. (forthcoming). Similar to our paper, they find that subjects, for whom it is reasonable to assume that they understand the BDM mechanism, display a WTP–WTA gap for a box of chocolates. While we replicate the experiments of PZ and ILS as closely as possible and use features of their setup to identify subject misconceptions, Bartling et al. (forthcoming) focus on the claim in Cason and Plott (2014) that misconceptions about the BDM can be falsely interpreted as support for theories of framing, such as reference-dependent preferences. Our papers mainly differ in the tasks used to practice the BDM and to classify subjects. Bartling et al. (forthcoming) use a price-list version of the BDM to elicit the subjects' valuation for a card with a known value (as in Cason and Plott, 2014). After subjects made their choice for each possible price they had to correctly identify their earnings for each possible price, given their choice before they could proceed. In contrast, we train

[2] Other research, for example, shows that experience with trading alleviates or eliminates the WTP–WTA gap (see e.g., List, 2003, Engelmann and Hollard, 2010).

[3] To be more precise, PZ first identify procedures typically used by experimentalists to control for subject misconceptions and then apply them all simultaneously. This includes employing an incentive compatible mechanism to elicit valuations, explaining the optimal responses to subjects, performing unpaid and paid practice rounds, ensuring the anonymity of decisions and payouts and measuring the gap directly through valuations.

[4] PZ emphasize that the paid practice rounds were explicitly designed for training purposes. For example, there was no randomization of the order of WTA and WTP tasks and practice rounds were continuously used for mistake corrections, public answers to questions, etc. PZ also note that the lottery data is possibly contaminated by subjects' misconceptions, such as failing to understand the BDM mechanism or failing to understand statistical independence (see PZ, Fn 15, and Plott and Zeiler, 2011).

[5] Other studies that apply the PZ procedures, such as Brown and Cohen (2015) and Kovalchik et al. (2005), support the findings of PZ (and ILS) for the mug task.

subjects as in PZ and ILS by explaining the BDM mechanism at length, going through unpaid practice rounds with them, and using paid practice rounds involving lotteries to expose subjects to the consequences of their decisions.

2. Experimental procedures

The experiment consisted of two parts. In the first part subjects received intensive training of the BDM mechanism followed by several paid lottery tasks which aimed at providing subjects with sufficient experience with the BDM mechanism. The second part of the experiment elicited the subjects' valuations for a commodity with a subjective value, either a coffee mug or a USB stick (see Appendix E for pictures of the two objects). Importantly, the choices in the first part allow us to classify subjects according to their understanding of the BDM procedure and to subsequently test for the presence of a WTP–WTA gap within these subsamples. If subjects who appear to understand the BDM mechanism show a gap in the WTP–WTA valuation then misconceptions about the elicitation procedure are unlikely to play a role in our student sample. This would cast doubt on the explanation that the frequently observed WTP–WTA gap is a product of the experimental procedures.

We ran our experiments with the help of computers using z-Tree (Fischbacher, 2007) and we therefore closely followed the procedures of ILS, who also ran a computerized experiment. This involved all five phases of the experiment in PZ – general instructions, worked examples, unpaid training rounds, paid practice rounds, and payment – as detailed below. The experiment was conducted by a research assistant who received extensive training before the experiment. Moreover, he used a script in order to keep the sessions as similar as possible and to facilitate the replication for different experimenters or researchers (see Appendix C). On arrival, subjects were randomly assigned to computers, and they received printouts of the written instructions (see Appendix B). These instructions were taken from PZ and translated into German. The BDM mechanism was explained in detail in the instructions, and we gave a short presentation using the examples from PZ's instructions to publicly explain the BDM mechanism and tasks in the experiment (see Appendix D). After answering any remaining questions, the experimenter guided the subjects through two worked examples on the screen of their computer. In a final step, subjects went through two unpaid training rounds on their own as in ILS (see Appendix C for details).

For the paid lottery tasks we used the same lotteries as PZ except the outcomes of the lotteries were in Euro while PZ's lotteries were in dollars.[6] In addition, we introduced two extra lotteries as explained below. Table 1 shows the details of the lotteries used. Notice that the set of lotteries consists of six small-stake lotteries and eight high-stake lotteries. The lotteries in PZ were constructed such that for each WTA lottery $S=(x, p; y, 1-p)$ there is a corresponding WTP lottery $T=(x+c, p; y+c, 1-p)$ which is obtained by adding 10 cents to the small-stake lotteries $L4$—$L6$ and adding €1 to the high-stake lotteries $L12$–$L15$, respectively. Note that x indicates the low lottery outcome and y the high lottery outcome except for degenerate lotteries where $x=y$. Assuming constant absolute risk aversion, expected utility theory implies that $WTP(T)-WTA(S)=c$. As pointed out by ILS, this design feature is well-suited for making within-subject comparisons of WTP and WTA valuations. For example, $L7$ is a high-stake lottery (€0, 0.3; €7, 0.7) for which subjects had to indicate their WTA, i.e., they were endowed with the lottery. The corresponding WTP lottery is the high-stake lottery $L12$ (€1, 0.3; €8, 0.7) for which subjects were asked to state their maximum buying price. Thus, $WTP(L12)-WTA(L7)=$ €1.

In addition to using PZ's lotteries, we included the lottery (€1, 0.5; €1.5, 0.5) for both sellers ($L11$) and buyers ($L16$). Notice first that the difference between the two outcomes of this lottery is smaller than in all other non-degenerate lotteries ($L3$, $L6$–$L10$ and $L12$–$L15$) and that the two strictly positive outcomes allow subjects to under- or overvalue the lottery. Second, the two lotteries are the final WTA and WTP tasks in order to ensure that subjects have gained substantial experience with the BDM in the previous paid practice rounds. Therefore, these lotteries are well-suited for identifying subjects who likely do not suffer from misconceptions about the elicitation procedure.

In total, subjects had to indicate their WTA or WTP for 16 lotteries and all of them were payoff relevant. Subjects started out by indicating their WTA for the first three lotteries and then switched from the seller role to the buyer role to indicate their WTP for the next three lotteries. Four of the six small-stake lotteries ($L1$, $L2$, $L4$ and $L5$) are degenerate lotteries paying an amount $x=y$ with certainty. For the high-stake lotteries subjects again started indicating their WTA ($L7$–$L11$) and then switched to the buying task for which they had to state their WTP ($L12$–$L16$). The fixed offers for each lottery and the mug (USB stick) were randomly drawn at the session level before the start of a session. Also, subjects received an initial endowment of €5 to start with (show-up fee). After each lottery, subjects learned whether they had bought or sold the lottery. If they bought or did not sell the lottery, they also learned the lottery outcome.

In the second part, subjects were asked to state their WTA or WTP for a mug featuring the logo of the Technical University Berlin. At the time of the experiment, the retail price of the mug was €7.50 in the university shop, but subjects were not informed of the price. All subjects received a mug and had some time to inspect it, as in PZ.[7] Then we briefly explained that those who own the mug can sell it to the experimenter and those who do not own the mug can buy it from the experimenter. Subjects were informed of their role on their computer screen. That is, we randomly determined seller

[6] PZ used two series of lottery tasks (called A lotteries and B lotteries), which differed in the probabilities of the two outcomes x and y. In our mug sessions we used only PZ's A lotteries while in our USB-stick session we used a combination of PZ's A and B lotteries (for details see Table A6 in the appendix).

[7] In ILS subjects were not given the mug, but they saw a picture of the mug on their computer screen. Only in their replication of PZ where the mug task comes before the lottery tasks, ILS distributed the mugs.

Table 1
Overview of lotteries.

	Val. type		Mug sessions	Val. type		USB sessions
Small-stakes	WTA	*L*1	€0.20, 0.5; €0.20, 0.5	WTA	*L*1	€0.20, 0.5; €0.20, 0.5
		*L*2	€0.35, 0.5; €0.35, 0.5		*L*2	€0.35, 0.5; €0.35, 0.5
		*L*3	€-0.20, 0.7; €0.70, 0.3		*L*3	€-0.20, 0.7; €0.70, 0.3
	WTP	*L*4	€0.30, 0.5; €0.30, 0.5		*L*4	€0.30, 0.5; €0.30, 0.5
		*L*5	€0.45, 0.5; €0.45, 0.5		*L*5	€0.45, 0.5; €0.45, 0.5
		*L*6	€-0.10, 0.7; €0.80, 0.3		*L*6	€-0.10, 0.7; €0.80, 0.3
Large-stakes	WTA	*L*7	€0, 0.3; €7, 0.7	WTA	*L*7	€0, 0.3; €7, 0.7
		*L*8	€0, 0.6; €5, 0.4		*L*8	€1, 0.5; €1.5, 0.5
		*L*9	€-4, 0.5; €8, 0.5	WTP	*L*9	€0, 0.4; €5, 0.6
		*L*10	€0, 0.7; €10, 0.3		*L*10	€-4, 0.5; €8, 0.5
		*L*11	€1, 0.5; €1.5, 0.5		*L*11	€0, 0.3; €10, 0.7
	WTP	*L*12	€1, 0.3; €8, 0.7		*L*12	€1, 0.7; €8, 0.3
		*L*13	€1, 0.6; €6, 0.4		*L*13	€1, 0.4; €6, 0.6
		*L*14	€-3, 0.5; €9, 0.5		*L*14	€-3, 0.5; €9, 0.5
		*L*15	€1, 0.7; €11, 0.3		*L*15	€1, 0.3; €11, 0.7
		*L*16	€1, 0.5; €1.5, 0.5		*L*16	€1, 0.5; €1.5, 0.5

and buyer roles in each session and, accordingly, subjects were told whether they owned the mug in front of them and were able to sell it or whether they could buy the mug in front of them. Note that ILS introduced a second task after the mug task, which we did not implement in our experiments. In this task of ILS, subjects indicated their WTA or WTP for a lottery which offered a box of luxury chocolates with a probability of 0.25 and nothing otherwise.

After the second part, subjects completed a 10-question 2.5-minute version of the Wonderlic Personal Test, which provides a measure of general cognitive ability. Subjects received 25 cents for each correct answer in the test.[8] We also administered a brief questionnaire which included questions about gender, details about the subjects' field of study, and free-form questions about the experiment including the estimated retail price of the mug or USB stick, respectively. At the end of the experiment, subjects were informed of their earnings on their computer screens and were required to fill out their receipt. Subjects were identified through their computer number and received their final payments from a research assistant in a separate room in private.[9]

Before we set out to replicate PZ and ILS as closely as possible, as reported on above, we conducted an experiment where the main task was to sell or buy a USB stick with a retail price of €8.90. This experiment was intended as a baseline treatment for a different project.[10] Since in this project we were not genuinely interested in the lottery tasks, we adapted the first part of PZ and ILS slightly. More specifically, while we used the same instructions and presentation of the BDM mechanism as in our mug sessions, we did not implement the two worked examples and the two unpaid practice rounds. Moreover, we used a different experimenter, but he followed the same script as in the mug sessions (except for the worked examples). The most important difference is, however, that we changed the order of lotteries in the first part. That is, we implemented the first eight lotteries as seller tasks and the last eight lotteries as buyer tasks using lotteries from PZ (see Table 1). As in our mug sessions, we introduced the two extra lotteries as the final task for each role (lottery *L*8 and *L*16) to gain insight into the subjects' understanding of the BDM. Notice that due to the change in the order of lotteries it is not possible to conduct a within-subject comparison of matched lottery pairs as in PZ, ILS, and in our replication of them, except for our two extra lotteries.

We conducted the experiments in November 2013 (USB sessions) and November 2014 (mug sessions) in the experimental economics lab at Technical University Berlin. Overall, 191 students from various fields of study participated in the experiments, of them 95 in the mug treatment and 96 in the USB stick treatment. Subjects were recruited from a database where students can register to participate in economic experiments (ORSEE, Greiner, 2004). The experiment lasted 90 minutes on average, and participants earned on average €18.46 without a mug/USB stick and €16.05 with a mug/USB stick.

[8] The complete Wonderlic Personal Test lasts 12 minute for 50 questions. For more information about the test see www.wonderlic.com. In addition, we elicited subjects' attitudes toward losses using a hypothetical choice list, where subjects were presented with six risky lotteries that involve an equal chance of gaining €45 and losing €$x = \{5, 15, 25, 35, 45, 55\}$.

[9] PZ emphasize the importance of the anonymity of the payments. However, recent evidence by Brown and Cohen (2015) shows that anonymity is not necessary for the results of PZ.

[10] Since we expected to replicate the findings of PZ but failed to do so, we abandoned the planned project. Instead, we tried to replicate the PZ result using the exact same procedures as in PZ and ILS laid out above and decided to investigate the heterogeneity of subject misconceptions.

Table 2
Experimental results.

Panel A: mug sessions									
WTA valuation	*L*1	*L*2	*L*3	*L*7	*L*8	*L*9	*L*10	*L*11	Mug
N	95	95	95	95	95	95	95	95	48
EV	0.20	0.35	0.07	4.90	2.00	2.00	3.00	1.25	
Mean	0.22	0.35	0.20	4.48	2.32	2.60	3.51	1.12	3.39
Median	0.20	0.35	0.15	4.90	2.00	2.00	3.00	1.15	3.06
Standard deviation	0.19	0.05	0.18	1.64	1.01	1.93	1.97	0.43	2.56
Mean/EV	1.09	0.99	2.91	0.91	1.16	1.30	1.17	0.90	
Freq. dominated offers	25.3%	27.4%	1.1%	1.1%	0.0%	1.1%	0.0%	21.1%	
WTP valuation	*L*4	*L*5	*L*6	*L*12	*L*13	*L*14	*L*15	*L*16	Mug
N	95	95	95	95	95	95	95	95	47
EV	0.30	0.45	0.17	5.90	3.00	3.00	4.00	1.25	
Mean	0.31	0.43	0.35	4.91	2.91	2.68	4.11	1.14	1.34
Median	0.30	0.45	0.20	5.00	3.00	3.00	4.00	1.23	1.00
Standard deviation	0.19	0.08	0.57	2.26	1.18	1.64	2.46	0.25	1.21
Mean/EV	1.04	0.96	2.06	0.83	0.97	0.89	1.03	0.91	
Freq. dominated offers	40.0%	41.1%	4.2%	6.3%	4.2%	0.0%	3.2%	10.5%	
WTA/WTP[a]	*L*1/*L*4	*L*2/*L*5	*L*3/*L*6	*L*7/*L*12	*L*8/*L*13	*L*9/*L*14	*L*10/*L*15	*L*11/*L*16	Mug
Mean	1.15	1.09	1.65	2.09	6.57	9.49	6.11	1.50	2.53
Median	1.00	1.00	1.00	1.07	1.05	1.20	1.06	1.00	3.06
Significance[b]	n/a	n/a		**	***	***	*		***
Panel B: USB sessions									
WTA valuation	*L*1	*L*2	*L*3	*L*4	*L*5	*L*6	*L*7	*L*8	USB
N	96	96	96	96	96	96	96	96	48
EV	0.20	0.35	0.43	0.30	0.45	0.53	2.10	1.25	
Mean	0.33	0.42	0.47	0.36	0.48	0.54	2.97	1.23	5.60
Median	0.21	0.36	0.50	0.31	0.45	0.55	3.00	1.25	5.28
Standard Deviation	0.42	0.20	0.20	0.17	0.24	0.22	1.74	0.48	2.52
Mean/EV	1.64	1.20	1.10	1.19	1.07	1.02	1.41	0.98	
Freq. dominated offers	68.8%	69.8%	11.5%	76.1%	74.0%	8.3%	1.1%	34.3%	
WTP valuation	*L*9	*L*10	*L*11	*L*12	*L*13	*L*14	*L*15	*L*16	USB
N	96	96	96	96	96	96	96	96	48
EV	2.00	2.00	3.00	5.90	3.00	3.00	4.00	1.25	
Mean	2.44	3.08	3.79	5.01	3.14	3.60	4.01	1.16	3.54
Median	2.20	2.03	3.10	5.00	3.00	3.00	3.52	1.10	3.27
Standard deviation	1.20	2.26	2.19	1.73	1.51	2.31	2.40	0.47	2.41
Mean/EV	1.22	1.54	1.26	0.85	1.05	1.20	1.00	0.93	
Freq. dominated offers	0.0%	0.0%	0.0%	5.2%	8.3%	1.0%	4.2%	31.3%	
WTA/WTP[a]	n/a	n/a	n/a	n/a	n/a	n/a	n/a	*L*8/*L*16	USB
Mean	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1.27	1.58
Median	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1.02	1.62
Significance[b]	n/a	n/a	n/a	n/a	n/a	n/a	n/a		***

[a] The ratio is computed as (WTA + c)/WTP for the lotteries while for the mugs and USB sticks it is simply the ratio of means and medians. The constant c is €0.10 for small-stake lotteries (*L*1–*L*3), €1 for high-stake lotteries (*L*7–*L*10) and €0 for the extra lotteries *L*11 (mug sessions) and *L*8 (USB stick sessions). Lotteries in the USB stick session cannot be matched, except for lotteries *L*8 and *L*16.

[b] Tests based on paired *t*-tests for lotteries and on *t*-tests corrected for unequal variances for mugs and USB sticks. Tests are not reported for the degenerate lotteries *L*1, *L*2, *L*4 and *L*5.

*** Significant at the 1% level.

** Significant at the 5% level.

* Significant at the 10% level.

3. Results

3.1. Is there a WTP–WTA gap?

Table 2 presents an overview of WTP–WTA valuations for lotteries in both the mug sessions (Panel A) and the USB-stick sessions (Panel B). More specifically, we report in each column of Table 2 the results from a WTA lottery and the matched WTP lottery in the mug sessions. For each lottery, the table shows the expected value of the lottery (EV), the mean, median, and standard deviation of subjects' offers, the ratio of the mean reported valuation to the EV, and the number of observations. Following ILS, we additionally show the mean and median of the WTA/WTP ratio for each pair of lotteries, which is calculated on an individual basis.

The last column in Table 2 displays the WTP–WTA valuations of subjects for mugs and USB sticks. The resulting statistics are based on a between-subject comparison as subjects acted either as a buyer or seller in these tasks. The last row in each of the two panels shows statistical tests for the hypothesis that WTA and WTP valuations for lotteries are the same using paired *t*-tests (within-subject comparison) and that WTA and WTP valuations for mugs (USB sticks) are the same using *t*-tests corrected for unequal variances across groups (between-subjects comparison).[11] As our alternative hypothesis is that WTA is higher than WTP in both tasks the tests are one-tailed. To allow for direct comparisons, we display the results of PZ and ILS in Table A1 in the appendix.

In line with ILS and the PZ data reported therein, we find pronounced gaps between the WTA and the WTP for the lotteries in the mug sessions. The WTA values significantly exceed the WTP values in four out of the five pairs of non-degenerate lotteries. We find no gap for our extra lottery pair ($L11$ and $L16$) for which the mean valuations of the WTA and the WTP are 1.12 and 1.14 respectively. In the sessions with USB sticks where it is only possible to compare the two extra lotteries, the average WTA for lottery $L8$ is 1.23 and the average WTP for lottery $L16$ is 1.16. There is no significant difference between the two valuations.

In sharp contrast to the results for the mug tasks of PZ and ILS, we find a prominent gap in the valuations for mugs and for USB sticks. On average, subjects are willing to pay €1.34 for a mug, but the subjects' minimum selling price is, on average, €3.39. Similarly, we find that subjects ask for a higher amount to part with their USB stick (€5.60) than they are willing to pay for it (€3.54). We can clearly reject the hypothesis of equal valuations for selling and buying mugs (*t*-test, $t=5.01$, $p<0.01$) and USB sticks (t-test, $t=4.08$, $p<0.01$).[12] Thus, we are not able to replicate the results on the valuations of mugs by PZ and ILS.

3.2. Do rational subjects show a WTP–WTA gap for mugs?

The lottery data provides important insights about subjects' understanding of the BDM mechanism. An indicator of subjects' understanding is that they follow a dominant bidding strategy in a lottery task. As bids and fixed offers were discrete (1 cent increments), an offer z is consistent with a weakly dominant bidding strategy if $x \le z \le y+0.01$ in WTA tasks and $x-0.01 \le z \le y$ in WTP tasks, where x and y denote the low and high lottery outcome, respectively. In Table 2, the last row for each lottery provides the share of bids for lotteries that are dominated.

Table 2 reveals only few violations of dominance in the non-degenerate lotteries (except for our two extra lotteries). The share of subjects who made an offer outside the boundaries of a non-degenerate lottery varies between 0 and 6 percent in the mug sessions and between 0 and 12 percent in the USB sessions. As hypothesized, the share of dominated bids for the two extra lotteries is higher than in the other non-degenerate lotteries in both the mug and the USB sessions because of the smaller range of weakly dominant bids and the possibility of underbidding, which is impossible in most other non-degenerate lotteries. In total, 26 percent of subjects made a bid outside the bounds of the two extra lotteries in the mug sessions, whereas 47 percent submitted a dominated bid in the USB sessions.

The degenerate lotteries ($L1$, $L2$. $L4$ and $L5$) offer the possibility to precisely measure the subjects' true valuations, because the outcome of the lottery is certain and is known to the subjects.[13] A substantial share of subjects made bids which are not consistent with weakly dominant bidding for the degenerate lotteries. About 25 percent of subjects made dominated offers for $L1$ and $L2$ (as a seller) and about 40 percent made dominated bids for $L4$ and $L5$ (as a buyer). Only about 47 percent of subjects made weakly dominant offers for all four lotteries, while about 15 percent always made dominated offers.[14] Compared to the mug sessions, the share of violations of weak dominance is substantially higher in the USB sessions. Between 69 and 76 percent of subjects failed to obey weak dominance in $L1$, $L2$, $L4$ and $L5$ (all as a seller), and only 11 out of 96 subjects (11 percent) made undominated offers for the four degenerate lotteries. That the share of dominated bids is consistently higher in the USB sessions shows the effectiveness of the additional worked examples and unpaid training rounds in the mug sessions.

We now use the lottery data to identify subjects who show a good understanding of the BDM mechanism (and call them "rational") and to explore whether these subjects display a WTP–WTA gap. If there is no gap for these subjects, there is support for the argument by PZ that the gap is generated by confusion about the elicitation procedure. More specifically, we identify two subgroups of rational subjects for the mug sessions: (i) subjects who make undominated bids for the extra lotteries, and (ii) subjects who make undominated bids for all 16 lotteries.[15] As a robustness check, we use these two criteria and require in addition that subjects submit the same bid for both extra lotteries. That is, we not only focus on subjects for

[11] All test results that we report are robust to using Wilcoxon signed-rank tests for lotteries and Mann–Whitney tests for mugs (USB sticks).

[12] We also find little evidence that WTP or WTA valuations are affected by individual characteristics or accumulated profits from the lottery tasks (for details see Table A2 in the appendix).

[13] As subjects likely differ in their risk attitudes, the non-degenerate lotteries are not well suited for an exact measurement of whether subjects' stated valuations equal their true valuations.

[14] In comparison, 60 percent of subjects in ILS made weakly dominant offers for all four degenerate lotteries and only 6 percent made dominated bids throughout (see Table A1 in the appendix). The share of violations of weak dominance in PZ is not directly comparable because PZ intervened during the lottery tasks in an effort to reduce misconceptions about the elicitation mechanism. For the sake of completeness panels B and C of Table A1 in the appendix report the shares from PZ.

[15] Notice that the set of subjects who avoid dominated bids in all lotteries exactly coincides with the set of subjects identified through undominated choices in the two extra lotteries and the four degenerate lotteries.

whom we have evidence that they understand the BDM mechanism, but also on subjects who do not display a WTP–WTA gap in the final pair of lotteries.[16]

First we consider those subjects who made undominated offers for the extra lotteries $L11$ and $L16$. Remember that subjects have already gained considerable experience with the BDM mechanism (the same amount of paid practice as PZ subjects) when making offers for these lotteries as these are the final WTA and WTP lotteries, respectively. By restricting the sample to this subgroup, we allow for dominated offers in earlier lottery rounds, which can then be viewed as practice rounds along the lines of PZ. This rationality criterion excludes 25 subjects (or 26 percent) from the sample. The average offer in the WTA task in this subgroup was 3.72, and the average offer in the WTP task was 1.29 (see left column of Panel A in Table 3). We can reject the hypothesis of equal WTP and WTA ($p < 0.01$). The WTA/WTP ratio is 2.88, which is higher than the ratio of 2.53 of the entire sample. Interestingly, the ratio is also smaller for non-rational subjects, i.e., subjects who submitted offers outside the bounds of the extra lotteries, albeit the difference in WTA or WTP valuations between rational and non-rational subjects is not statistically significant (for details see Table A3 in the appendix). In the right column of Panel A in Table 3, we go one step further and require not only that bids are undominated but also that they are consistent, i.e., that WTA valuations are the same as WTP valuations in $L11$ and $L16$. In essence this implies no WTP–WTA gap for these subjects as one would expect for money and lotteries over monetary outcomes. In total, 34 out of 95 subjects (36 percent) fulfill this additional criterion. The gap is smaller than in the larger sample because of the lower WTA (on average, 2.64), but the difference remains significant ($p < 0.01$).

Panel B of Table 3 presents the results for our second criterion, which requires that subjects use a weakly dominant bidding strategy in all 16 lotteries. As this is a stronger requirement than the previous criterion, only about half of our entire sample (42 out of 95 subjects) uses such strategies throughout all lottery tasks. The average WTP offer was 1.48 and the average WTA was 3.02, yielding a WTA/WTP ratio of 2.04. Again, we can reject the hypothesis of equal WTP and WTA ($p < 0.015$). Note that, the WTA/WTP ratio (3.03) is higher for non-rational subjects who made at least one dominated bid in any lottery (see Table A3 in the appendix). In the right column of Panel B we consider only subjects who made rational bids for all lotteries and in addition consistent bids for the extra lotteries, i.e., WTA($L11$)=WTP($L16$). While in total only 27 subjects fall into this category, the valuations of sellers are still significantly higher than the valuations of buyers ($p < 0.048$).

For all criteria, we find evidence that subjects who fulfill the criteria have, on average, higher scores in the Wonderlic test than subjects who made at least one mistake in the respective sets of lotteries (see Table A4 in the Appendix). This provides further support that subjects who met our rationality criteria understood the elicitation mechanism correctly (see Bartling et al., forthcoming, for a related finding).

Finally, Panel C of Table 3 reports the results of the USB sessions. Here we only look at subjects with undominated offers for the two extra lotteries $L8$ and $L16$ (last WTA and WTP tasks respectively) as the second criterion does not result in a sufficient number of observations to run statistical tests.[17] Again, we find a substantial WTP–WTA gap. The average offer of rational subjects in the WTP task was 3.02 and in the WTA task 5.30, the difference being statistically significant ($p < 0.01$).

Together, these findings show a persistent WTP–WTA gap even for subjects for whom we have evidence that they understood the elicitation procedure. If misconceptions about the BDM mechanism are responsible for the frequently observed WTP–WTA gap, then we should not observe a gap for those who understand the mechanism. Indeed, there is no evidence of a gap for those subjects in ILS who made weakly dominant bids for all 14 lotteries, whereas those who made at least one mistake display a significant gap (for the details see Table A5 in the appendix).[18] However, the ILS data also reveal that subjects who seemingly suffered from misconceptions about the BDM mechanism, display no gap in their second task, the chocolate gamble. In contrast, subjects without misconceptions have, on average, a significantly higher WTA than WTP for the chocolate box. This suggests that misconceptions can both contribute to or decrease the WTP–WTA gap. Given our evidence it is, however, unlikely that misconceptions about the elicitation method are the main source for the observation of a WTP–WTA gap.

4. Conclusion

PZ proposed procedures aimed at eliminating subject misconceptions about the BDM mechanism and demonstrated that the WTP–WTA gap for an object can be turned on and off. In spite of utilizing the experimental procedures of PZ (and ILS), we observe a WTP–WTA gap for lotteries over monetary payoffs and for mugs or USB sticks. In particular, we show that the gap does not disappear when considering only subjects who show a good understanding of the BDM mechanism, i.e., subjects who made undominated offers for a certain subset of lotteries or for all lotteries. That the proportion of subjects making dominated bids is lower in the mug sessions where subjects received more training before the lottery tasks, suggests that the PZ procedures succeed in eliminating misconceptions about the BDM mechanism for some subjects. At the same time, however, there is still a substantial share of subjects in our experiments who have difficulties with the BDM mechanism even after extensive instructions and training. Notwithstanding these differences in understanding the BDM,

[16] Since bids are in 1 cent increments, we also allow the WTA($L11$) to be 1 cent higher than the WTP($L16$).

[17] Requiring undominated bids in all lotteries reduces the sample to eight subjects (one buyer and seven sellers).

[18] The same is true for the (weaker) classification according to number of weakly dominant bids for the degenerate lotteries (all versus at least one mistake). Recall that it is not possible to identify subjects with and without misconceptions in PZ because the experimenter intervened during the lottery rounds to correct misconceptions.

Table 3
WTP–WTA gap for subsamples of rational subjects.

Panel A				
Mug sessions: extra lotteries L11 and L16	$1 \leq offer \leq 1.50$[a]		& WTA(L11)=WTP(L16)	
	# Obs.	Avg. offer	# Obs.	Avg. offer
WTP mug	$N=37$	1.29	$N=17$	1.23
WTA mug	$N=33$	3.72	$N=17$	2.64
WTA/WTP		2.88		2.14
p-value		0.00		0.01
Panel B				
Mug sessions: all lotteries	$x \leq offer \leq y$[a]		& WTA(L11)=WTP(L16)	
	# Obs.	Avg. offer	# Obs.	Avg. offer
WTP mug	$N=23$	1.48	$N=13$	1.47
WTA mug	$N=19$	3.02	$N=14$	2.67
WTA/WTP		2.04		1.81
p-value		0.013		0.048
Panel C				
USB sessions: extra lotteries L8 and L16	$1 \leq offer \leq 1.50$[a]			
	# Obs.	Avg. offer		
WTP USB stick	$N=24$	3.02		
WTA USB stick	$N=27$	5.30		
WTA/WTP		1.75		
p-value		0.00		

Notes:
[a] WTA offers are consistent with a weakly dominant bidding strategy if they are in the interval $[x, y+0.01]$ and WTP offers are consistent if they are in the interval $[x-0.01, y]$, where x denotes the low and y the high lottery outcome.

both groups clearly show a WTP–WTA gap for mugs and USB sticks. Therefore, we find no evidence of the idea that misconceptions about the BDM mechanism are the main source of the WTP–WTA gap.

Acknowledgements

We thank two anonymous referees, the editor, and seminar participants at the WZB Berlin for helpful suggestions, as well as our colleagues in the project “Cultural Framing Effects in Experimental Game Theory,” Michael Hutter and Juliane Böhme, for many stimulating discussions. We also thank Kathryn Zeiler and Charlie Plott as well as Andrea Isoni, Graham Loomes, and John Sugden for sharing their data and material, and Björn Bartling and Andrea Isoni for valuable comments. We are grateful to Nina Bonge for programming, Jennifer Rontganger for copy editing, Sharwin Rezagholi and Juliane Böhme for assistance in running the sessions as well as Renke Schmacker for research assistance.

Appendix. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.euroecorev.2015.05.006.

References

Bartling, Björn, Engl, Florian, Roberto, A. Weber, 2015. Game form misconceptions are not necessary for a willingness-to-pay vs. willingness-to-accept gap. J. Econ. Sci. Assoc. (forthcoming).

Becker, Gordon M., DeGroot, Morris H., Marschak, Jacob, 1964. Measuring utility by a single-response sequential method. Behav. Sci. 9 (3), 226–232.

Brown, Alexander L., Cohen, Gregory, 2015. Does anonymity affect the willingness to accept and willingness to pay gap? A generalization of Plott and Zeiler. Experimental Economics 18 (2), 173–184.

Cason, Timothy N., Plott, Charles R., 2014. Misconceptions and game form recognition: challenges to theories of revealed preferences and framing. J. Polit. Econ. 122, 1235–1270.

Engelmann, Dirk, Hollard, Guillaume, 2010. Reconsidering the effect of market experience on the “Endowment Effect”. Econometrica 78 (6), 2005–2019.

Ericson, Keith M., Fuster, Andreas, 2014. The endowment effect. Annu. Rev. Econ. 6, 555–579.

Fischbacher, Urs., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.

Greiner, Ben, 2004. An online recruitment system for economic experiments. In: Kurt, Kremer, Volker, Macho (Eds.), Forschung und wissenschaftliches Rechnen 2003 GWDG Bericht 63, Ges. fuer Wiss. Datenverarbeitung, Göttingen, pp. 79–93.

Hanemann, W. Michael, 1991. Willingness to pay and willingness to accept: how much can they differ. Am. Econ. Rev. 81 (3), 635–647.

Huck, Steffen, Kirchsteiger, Georg, Oechssler, Joerg, 2005. Learning to like what you have: explaining the endowment effect. Econ. J. 115, 689–702.

Isoni, Andrea, Loomes, Graham, Sugden, Robert, 2011. The willingness to pay–willingness to accept gap, the “Endowment Effect,” subject misconceptions, and experimental procedures for eliciting valuations: comment. Am. Econ. Rev. 101, 991–1011.

Kahneman, Daniel, Knetsch, Jack L., Thaler, Richard H., 1990. Experimental tests of the endowment effect and the Coase theorem. J. Polit. Econ. 98, 1325–1348.
Koh, Weining, Wong, Wei-Kang, 2012. The WTA-WTP gap: do possession and framing still matter despite important controls for subject misconceptions. Mimeo.
Kovalchik, Stephanie, Camerer, Colin F., Grether, David M., Plott, Charles R., Allman, John M., 2005. Aging and decision making: a comparison between neurologically healthy elderly and young individuals. J. Econ. Behav. Organ. 58 (1), 79–94.
Köszegi, Botond, Rabin, Matthew, 2006. A model of reference-dependent preferences. Q. J. Econ. 121 (4), 1133–1165.
List, John A., 2003. Does market experience eliminate market anomalies. Q. J. Econ. 118, 41–71.
Plott, Charles R., Zeiler, Kathryn, 2005. The willingness to pay–willingness to accept gap, the "Endowment Effect," subject misconceptions, and experimental procedures for eliciting valuations. Am. Econ. Rev. 95, 530–545.
Plott, Charles R., Zeiler, Kathryn, 2007. Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory. Am. Econ. Rev. 97, 1449–1466.
Plott, Charles R., Zeiler, Kathryn, 2011. The willingness to pay–willingness to accept gap, the "Endowment Effect," subject misconceptions, and experimental procedures for eliciting valuations: reply. Am. Econ. Rev. 101 (2), 1012–1028.
Thaler, Richard, 1980. Toward a positive theory of consumer choice. J. Econ. Behav. Organ. 1, 39–60.