

Game form misconceptions are not necessary for a willingness-to-pay vs. willingness-to-accept gap

Björn Bartling¹ · Florian Engl¹ · Roberto A. Weber¹

Received: 15 November 2014 / Accepted: 9 April 2015 / Published online: 27 May 2015
© Economic Science Association 2015

Abstract Cason and Plott (J Polit Econ, 122(6):1235–1270, 2014) show that subjects' misconception about the incentive properties of the Becker-DeGroot-Marschak (BDM) value elicitation procedure can generate data patterns that look like—and might thus be misinterpreted as evidence for—preferences constructed from endowments or reference points. We test whether game form misconceptions are necessary to produce willingness-to-pay (WTP) vs. willingness-to-accept (WTA) gaps in a valuation experiment in which subjects are randomly assigned to the role of either buyer or seller. We employ a design that allows us to identify whether a subject understood the incentive properties of a price-list version of the BDM mechanism. We find a robust WTP-WTA gap, even among subjects whose elicited valuations for a good of induced and known monetary value and whose ability to identify the payoffs resulting from their choices indicate an understanding of the incentive properties of the BDM mechanism. We conclude that game form misconceptions are not a necessary condition for the emergence of WTP-WTA gaps.

Keywords WTP-WTA gap · Game form misconception · BDM mechanism · Endowment effect · Experimental methods · Highly replicable laboratory environment

JEL Classification C91 · D03

Electronic supplementary material The online version of this article (doi:[10.1007/s40881-015-0008-0](https://doi.org/10.1007/s40881-015-0008-0)) contains supplementary material, which is available to authorized users.

✉ Roberto A. Weber
roberto.weber@econ.uzh.ch

¹ Department of Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zurich, Switzerland

1 Introduction

A recent article by Cason and Plott (2014) argues for caution in drawing inferences about non-standard preferences from choice data using experimental value elicitation procedures, such as the mechanism introduced by Becker et al. (1964, henceforth, BDM). This argument raises challenges to parts of the behavioral literature, especially the literature on willingness-to-pay (WTP) vs. willingness-to-accept (WTA) gaps—i.e., the tendency for subjects endowed with an object to state higher valuation for the object than subjects not endowed with it—which is one of the most widely studied phenomena in behavioral economics (see, e.g., Kahneman et al. 1991; Ericson et al. 2014).

Cason and Plott support their argument with evidence that a systematic bias in subjects' stated valuations for an item with a known dollar value is reflective of subjects' misunderstanding of the game form—that is, subjects systematically fail to understand the incentive properties of the BDM mechanism and therefore over-report their WTA. Specifically, Cason and Plott conduct a classroom experiment in which they endow subjects with a card that is redeemable from the experimenter for \$2, and allow subjects to state sales price offers through the BDM procedure. A random posted price between \$0 and an upper price limit ranging from \$4 to \$8 is generated, and a subject sells his card to the experimenter only if this posted price exceeds the subjects' stated offer price. Using this design, Cason and Plott find that a large majority of subjects (83 %) do not select offers within 5 cents of the dominant strategy of \$2. This frequency of apparently confused subjects decreases, to 69 %, with feedback and repetition. Moreover, subjects' offers are biased upward, and are also influenced by the upper range of the distribution from which the posted prices are drawn. From this evidence, Cason and Plott argue that studies in which researchers attempt to make inferences about preferences from experiments using the BDM elicitation procedure should be taken with caution.

The possibility that game form misconceptions may account for behavior inconsistent with standard models of choice in some experiments is intriguing—indeed, this seems to be the case in Cason and Plott's experiment. In this paper, we ask if game form misconceptions are *necessary* to produce such data, or whether WTP-WTA gaps are observed even among subjects with no apparent misconceptions about the incentive properties of the BDM.

To test whether subjects' misunderstanding of the incentive properties of the BDM mechanism is necessary in order to observe WTP-WTA gaps, we follow the recommendation that “experimental procedures should be designed to avoid ‘subject misconceptions’” (Plott and Zeiler 2005). But, we also go further and focus the analysis on those subjects for whom we have evidence that there is no misconception. If subject misconceptions about the BDM procedure are largely absent, but subjects' choices nevertheless exhibit significant WTP-WTA gaps, then we can exclude that the observed behavior is entirely driven by mistakes in understanding the incentive properties of the BDM.

To address this question, we conduct a valuation experiment in which subjects are randomly assigned to the role of either buyer or seller and where we elicit, respectively, subjects' WTP or WTA values. Our experiment consists of two parts.

First, in Part I, we elicit buying and selling prices for a card worth precisely 8.50 Swiss francs (CHF). We employ a price-list representation of the BDM procedure, in which subjects are presented with a series of prices and asked, for each price, whether they are willing to trade at that price (Kahneman et al. 1990; Murphy et al. 2010). Part II of the study is identical to Part I, except for exactly one difference: in Part II subjects do not trade an object of induced value (a card worth 8.50 CHF) but instead an item with private and heterogeneous valuation—a box of chocolates.

Furthermore, after subjects indicate their choices in each part, but before they can finalize their choices, we also require them to identify the actual payoff consequences for each choice and, in case they make a mistake, we provide feedback and ask them to correct their mistake. Hence, all subjects, at the time of implementing a choice, have correctly entered all of the payoff consequences from their choices. Moreover, we can identify those subjects who entered their payoffs correctly on the very first attempt.

These procedures allow us to identify a subset of subjects who satisfy three conditions that are likely to mean they understand the BDM mechanism and its payoff consequences. First, we consider only subjects who provide payoff-maximizing valuations in Part I, where values are common and known. Second, we further eliminate any subjects who do not enter their payoffs entirely correctly on their first attempt in Part I. Third, we also eliminate subjects who do not enter their payoffs correctly on their first attempt in Part II, when bidding or offering prices for the box of chocolates. Hence, what remains are subjects who (1) make optimal choices in the BDM with known value, (2) identify all payoff consequences correctly in the BDM with known value on their first attempt, and (3) identify all payoff consequences correctly in the BDM for the box of chocolates on their first attempt. These “sophisticated” subjects are unlikely to have misconceptions about the BDM game form.

We test whether a WTP-WTA gap for the object of unknown value—the box of chocolates—still obtains within this subset of our population. Our main result is that, looking only at the subjects with no apparent game form misconception, we nevertheless replicate a strong WTP-WTA gap in Part II. Hence, WTP-WTA gaps obtain even among a population for which misconception of the incentive properties of the BDM mechanism is likely not an issue.

However, while we observe that a majority of subjects, 70 %, provide the payoff-maximizing valuation in Part I of the experiment, we also observe that 30 % do not. Hence, we confirm the limitations of the BDM as a valuation mechanism observed by Cason and Plott (2014), providing support for their claim that caution is warranted when relying on valuations elicited with this mechanism due to subject miscomprehension of the incentive properties.

Before proceeding to our design and results, we note that we do not attempt to provide evidence supporting any particular explanation for WTP-WTA gaps. Thus, we make no claims that the WTP-WTA gap in our data is driven by loss aversion, or whether it results from other behavioral phenomena, such as misperception of the value of a commodity or alternative sources of utility or disutility (cf. Plott and Zeiler 2005, 2007; Weaver and Frederick 2012). Instead, our simpler goal is to test

whether game form misconceptions are a necessary condition to produce WTP-WTA gaps, or whether such gaps obtain even when such misconceptions are absent.

2 Experimental design

Subjects participated in two decision tasks, presented as Parts I and II of the experiment. Part I consisted of a BDM value elicitation choice task for an object with fixed and known value. Part II consisted of an identical choice task for a good with heterogeneous and unobservable value. Subjects were informed that either Part I or Part II would count for their earnings in the study, but not both, and that the decisions in one part had no consequences for the respective other part. No feedback was given between Part I and Part II.

At the beginning of the study, subjects were randomized to be in the role of either buyer or seller. Randomization took place on the session level; that is, in a given session all subjects were either buyers or sellers. At the outset of the study, subjects were informed that they received an initial endowment of CHF 25, as a show-up fee.

2.1 Part I

Subjects in a *buyer* session could buy a “card” from the experimenter. The card was a small piece of paper located at the subject’s computer terminal, which stated, “This card is worth 8.50 CHF.” If a subject bought the card, he or she could redeem it for 8.50 CHF from the experimenter at the end of the study. The BDM mechanism was used to elicit the highest price a subject was willing to pay to buy the card. At the end of the study, the computer drew an individually randomized price between 0 and 20 CHF, in increments of 1 CHF. Each subject decided, for all possible 21 prices, whether or not to buy the card at that price. If a subject bought the card, he or she could redeem it for 8.50 CHF, but had to pay the price out of the initial endowment of 25 CHF. If the subject did not buy the card, he or she could not redeem it, but did not pay the randomly drawn price. Subjects entered their choices in a table in which each row corresponded to a possible price, by clicking either on a box stating “I buy at this price” or on a box stating “I do not buy at this price.”¹ Prices were ordered from 0 CHF (top row) to 20 CHF (bottom row). Subjects were informed that they should indicate the maximum price they were willing to pay for the card, and that they should therefore only have one price at which they switched from buying to not buying, as their decisions would otherwise be inconsistent.² For each possible price, we also displayed the payoff consequences of the two possible choices on the corresponding row.

Subjects in a *seller* session were endowed with the card—the instructions explicitly stated, “The card is yours, you own it”—but had the option to sell it to the experimenter. They had to decide, for each possible price, whether or not to sell the

¹ Instructions and choice interfaces are provided in the Online Supplementary Material.

² If a subject indicated more than one switching point, the computer would not allow the subject to proceed.

card at that price. If they sold the card, they could not redeem it for 8.50 CHF, but received the randomly drawn price, in addition to their initial 25 CHF. If they did not sell the card, they could redeem it for 8.50 CHF, in addition to their initial 25 CHF, but did not receive the price. Each seller indicated his or her choice in a table that was equivalent to the one presented to buyers. For each of the 21 possible prices, a seller had to click either on a box stating, “I sell at this price,” or on a box stating, “I do not sell at this price.” As with buyers, the software enforced unique switching points and also displayed payoff information for the two possible choices in each row.

Clearly, with an induced value of 8.50 CHF, payoff-maximizing subjects should buy the card for prices of at most 8 CHF and sell it for prices of 9 CHF or higher. Hence, a buyer who switches from buying the card at a price of 8 CHF to not buying at a price of 9 CHF or a seller who switches from not selling at a price of 8 CHF to selling at a price of 9 CHF is less likely to suffer from game form misconception than if they behaved otherwise.

We also included an additional test of whether subjects understood the incentive properties of the BDM mechanism in Part I. After they made their choices for each price and clicked an “OK” button to proceed, subjects had to identify the exact payoffs that would result for each of the possible 21 prices, given their choice (buy/not buy or sell/not sell) at that price. Panel A of Fig. 1 provides part of this interface for buyers. Subjects could always alter their choices for any of the 21 possible prices by returning from the payoff entry screen to the decision screen before confirming their payoff entries. Once subjects were ready to proceed, the computer verified whether they had entered all of the resulting payoffs correctly. If a subject entered all 21 payoffs correctly on the first attempt, given his or her final choices, the subject earned an additional 2 CHF. If there was a mistake in any of the payoff entries, the subject was asked to identify the resulting payoffs once more. The computer allowed a subject to proceed only once he or she entered all payoffs correctly.

2.2 Part II

Part II of the study was identical to Part I except for exactly one difference: subjects did not trade an object of induced value (a card worth 8.50 CHF) but a box of chocolates.³ The box of chocolates was purchased from a well-known upmarket Swiss confectionary at a retail price of 17 CHF, but subjects were not informed about this price.

As in Part I, subjects in a *buyer* session could buy the box of chocolates from the experimenter. A box was located at each subject’s computer terminal. We used exactly the same BDM mechanism as in Part I to elicit a subject’s maximum willingness to pay, with random prices between 0 and 20 CHF (individually drawn at the end of the study), in increments of 1 CHF. The screen

³ We did not counterbalance the order of the two tasks. Cason and Plott (2014) find that subject misconception about the BDM decreases with experience. Therefore, by having the task that we use to measure misconception first, our classification of subject comprehension in the second task is conservative.

Final payoff (25 CHF - price + value of card)	Buy at this price	Price (one row randomly selected to count)	Do not buy at this price	Final payoff (25 CHF)	Your final payoff (in CHF)
33.50 CHF	<input checked="" type="checkbox"/>	0 CHF	<input type="checkbox"/>	25 CHF	
32.50 CHF	<input checked="" type="checkbox"/>	1 CHF	<input type="checkbox"/>	25 CHF	
31.50 CHF	<input checked="" type="checkbox"/>	2 CHF	<input type="checkbox"/>	25 CHF	
30.50 CHF	<input checked="" type="checkbox"/>	3 CHF	<input type="checkbox"/>	25 CHF	

Panel A – Part I

Final payoff (25 CHF - price + chocolate)	Buy at this price	Price (one row randomly selected to count)	Do not buy at this price	Final payoff (25 CHF)	Your final payoff (in CHF)	Your final payoff (chocolate)
25 CHF + chocolate	<input checked="" type="checkbox"/>	0 CHF	<input type="checkbox"/>	25 CHF		no <input type="radio"/> yes <input type="radio"/>
24 CHF + chocolate	<input checked="" type="checkbox"/>	1 CHF	<input type="checkbox"/>	25 CHF		no <input type="radio"/> yes <input type="radio"/>
23 CHF + chocolate	<input checked="" type="checkbox"/>	2 CHF	<input type="checkbox"/>	25 CHF		no <input type="radio"/> yes <input type="radio"/>
22 CHF + chocolate	<input checked="" type="checkbox"/>	3 CHF	<input type="checkbox"/>	25 CHF		no <input type="radio"/> yes <input type="radio"/>

Panel B – Part II

Fig. 1 Computer interfaces of the payoff entry screens. Screen shots of parts of the payoff entry stages for buyers in Part I (a) and Part II (b). Subjects’ choices for each price are displayed in grey (in the above example, the subject bought at all shown prices). The full screens contain all 21 rows for prices from 0 CHF to 20 CHF. The screens for sellers are equivalent. All screens are included in full in the Online Supplementary Material. In Part II, the payoff entry screen also required subjects to indicate whether or not the final payoff additionally included the box of chocolates

that was employed to elicit buying decisions for each possible price corresponded exactly to the screen used in Part I. If a subject bought the box, the subject could take it home, but had to pay the price out of the initial endowment of 25 CHF. If a subject did not buy the box, the subject could not take it home, but did not have to pay the drawn price. A unique switching point was enforced by the software, as in Part I.

Subjects in a *seller* session were endowed with the box of chocolates but had the option to sell it to the experimenter. As in Part I subjects had to decide, for each possible price between 0 and 20 CHF, whether or not to sell the box of chocolate at that price. If a subject sold the box, he or she could not take it home, but received the randomly drawn price, in addition to the initial 25 CHF. If a subject did not sell the box, he or she could take it home, in addition to the initial 25 CHF, but did not receive the price. The interface that was employed to elicit selling decisions for each possible price corresponded exactly to the screen used in Part I. The software enforced a unique switching point.

As with Part I, subjects again had to identify all possible realized payoffs based on their specific choices, before proceeding. Subjects now had to enter their resulting monetary earnings and whether they would own or not own the box of chocolates, for each of the possible 21 prices, given their choices at that price. Panel B of Fig. 1 provides a screen shot of part of the payoff entry screen. At that point, subjects had the option to change any of their previous buying or selling decisions, and they earned an additional 2 CHF only if they entered all 21 potentially resulting payoffs correctly on the first attempt. The computer allowed a subject to proceed only once he or she entered all payoffs correctly.

An important feature of the experimental design is that using an item of induced value in Part I, together with the incentivized calculation of the resulting payoffs in Parts I and II, allows us to identify which subjects understand the incentive properties of the BDM mechanism and thus to restrict our analysis of Part II behavior to subjects who revealed such understanding of the mechanism. Subjects who make payoff-maximizing choices in Part I and identify all of their payoffs correctly in Parts I and II, on the first attempt, presumably hold very little misconception about the BDM procedure. Hence, for these subjects, we assume that elicited valuations in Part II for the box of chocolate are unlikely to be biased by game form misconception.

2.3 Additional measurements

After subjects completed Part II, we administered two measures of cognitive ability: a cognitive reflection test (Frederick 2005) and a computerized 12-item Raven progressive matrices test (Raven et al. 1998). Subjects received additional 0.50 CHF for each correctly solved puzzle in the Raven test.

2.4 Methodological note: highly replicable laboratory environment

Our design also employs a novel procedure that allows high degrees of replicability across experimental sessions and, in principle, across experimental laboratories. Specifically, subjects received instructions both through printed instructions at their desk and through an audio recording played via loudspeaker in the laboratory.⁴ Hence, the key elements of our experimental environment can be reliably replicated across sessions in our laboratory or elsewhere. We consider this enhanced replicability important for at least two reasons. First, the use of pre-recorded audio files to deliver instructions ensures that these are delivered in exactly the same manner across multiple sessions and by a speaker unaware of the experimental hypotheses. Second, it lowers the barriers to direct replication, since a researcher at any point in the future has a clearer understanding of the precise laboratory environment and the ability to replicate it more closely. Indeed, a researcher need only distribute paper instructions, play the audio files, and start the experimental software to conduct a session virtually identical to the ones in this study.

2.5 Information on sessions, subjects, and earnings

We conducted four sessions on two consecutive days in November 2014, with 140 subjects in total. On each day, we conducted one buyer session and one seller session. The order of the buyer and seller sessions was reversed on the second day. Overall, 69 subjects participated in the role of buyer and 71 subjects in the role of seller.

⁴ Prior to conducting the experiment, we contracted a professional voice artist (hired through fiverr.com) to record audio files of the instructions. The audio files, along with other experimental program files, are provided in the Online Supplementary Material.

All sessions took place at the Laboratory for Behavioral and Experimental Economics of the Department of Economics at the University of Zurich. The experiments were run with the software “z-Tree” (Fischbacher 2007). Subjects were students from the University of Zurich and the Swiss Federal Institute of Technology in Zurich. We used the software “hroot” (Bock et al. 2014) for recruitment.⁵ Each subject participated only once, either as seller or buyer.⁶ Sessions lasted about 75 min. On average, subjects earned CHF 36.22, which includes a show-up fee of CHF 25. In addition, 19 of our subjects also took home a box of chocolates.

3 Results

We start by analyzing the decisions of buyers and sellers in Part I. Buyers and sellers indicated, respectively, their maximum WTP and minimum WTA for a card worth 8.50 CHF. If subjects simply seek to maximize monetary earnings, then subjects’ true WTP and WTA for this card should both be 8.50 CHF. Our price-list design does not allow us to observe valuations with high precision. What we can identify is that, for instance, a buyer who is willing to pay X but not $X + 1$ must have a valuation in the interval, $WTP \in [X, X + 1)$. Similarly, a seller who is willing to accept $X + 1$ but not X must have a valuation in the interval, $WTA \in (X, X + 1]$. Therefore, in our analysis, we use the mid-point of the two values between which a buyer switches from buying to not buying or a seller switches from not selling to selling, $(X + X + 1)/2$, as our estimate of the WTA or WTP value for that subject.⁷

Overall, 98 of 139 subjects (70.5 %) made choices consistent with WTA and WTP values of 8.50 CHF. More precisely, 53 out of 69 subjects (76.8 %) in the buyer condition switched from buying to not buying between 8 and 9 CHF and 45 out of 70 subjects (63.4 %) in the seller condition switched from not selling to selling in the same interval.

The mean estimates of WTP and WTA for the card are 8.43 CHF and 9.01 CHF, respectively. These values are also provided in the first column of Table 1. The left panel of Fig. 2 presents the cumulative frequencies of estimated WTP and WTA values. The two frequencies overlap considerably, indicating little difference between the estimated valuations for buyers and sellers for the card with known value of 8.50 CHF. Consistent with the differences in means, there are slightly more

⁵ The study was conducted in English. One subject in the role of seller had to be excused from the study due to a lack of fluency in English. The analysis thus comprises 139 subjects in total.

⁶ Our subjects generally had little prior experience with laboratory experiments. On average, they had participated in 0.94 experiments in our lab prior to our study, with a median of 0. Of those who had prior experience with laboratory experiments, only nine subjects had participated in experiments that were labeled (in our recruiting software) as “simple value-based choice tasks.” Thus, it is unlikely that many of our subjects had prior exposure to the BDM mechanism or WTP-WTA type experiments.

⁷ If a subject chose a corner solution, i.e., chose to buy or sell at any or at no price in the list, we use the respective boundary of the price list as WTP or WTA. For example, if a subject chose not to sell the box of chocolate at any price, then we assign a WTA of 20 for that subject. In total, two subjects made such choices.

Table 1 WTP-WTA gaps

	Card (all subjects)	Chocolate (all subjects)	Chocolate (sophisticated)	Chocolate (confused)	p-value (sophisticated vs. confused)
WTP (buyer)	8.43 (0.191, N = 69)	4.78 (0.377, N = 69)	4.63 (0.468, N = 41)	5.00 (0.635, N = 28)	0.711
WTA (seller)	9.01 (0.276, N = 70)	8.62 (0.612, N = 70)	7.65 (0.765, N = 40)	9.91 (0.962, N = 30)	0.054
p-value (WTP vs. WTA)	0.219	0.000	0.005	0.000	

Mean estimated WTP and WTA in CHF. Numbers in brackets are standard errors and number of subjects. P-values are from Wilcoxon rank-sum tests

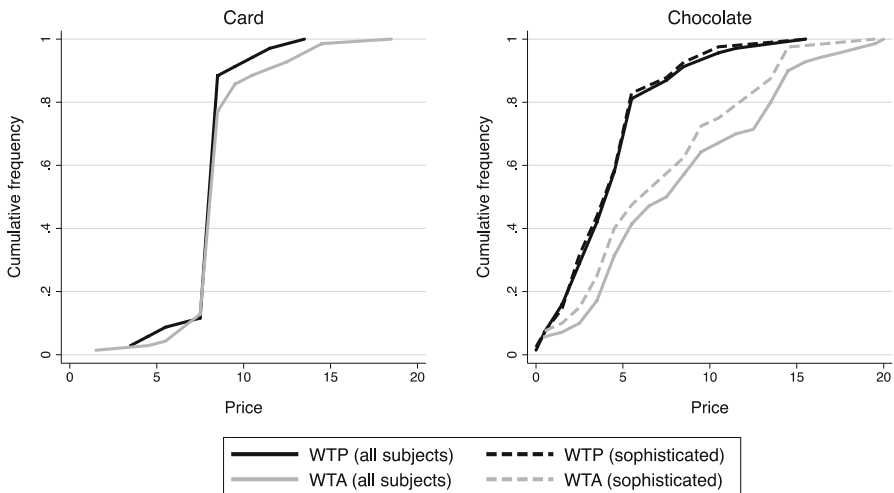


Fig. 2 Cumulative distribution functions of WTP and WTA estimates in Part I (Card) and Part II (Chocolate)

high value estimates for the sellers than for buyers. However, the two means and the two distributions do not differ significantly (Wilcoxon rank-sum test, $p = 0.219$; Kolmogorov–Smirnov test, $p = 0.698$).

In both conditions, we observe underpaying as well as overpaying. Of the 16 buyers with estimated WTP other than 8.50 CHF, exactly eight underpaid and eight overpaid for the card. In the seller condition, nine subjects yielded WTA estimate values below 8.50 CHF and 16 gave values greater than 8.50 CHF. However, neither of these proportions differs significantly from 50 % (two-sided binomial test: $p = 1$ and $p = 0.230$, respectively).

These findings demonstrate, firstly, that our price-list version of the BDM procedure for eliciting WTP and WTA resulted in fewer deviations from the optimum than the stated-price version in Cason and Plott (2014). They find that only

41 out of 245 (16.7 %) of their subjects chose the payoff maximizing WTA.⁸ Secondly, we find that deviations are relatively balanced and thus do not indicate that our implementation of the BDM mechanism leads subjects to exhibit a clear systematic bias in the sense of stating either too high or too low WTP or WTA values, respectively.

Result 1: *Overall, 98 out of 139 subjects provided choices that maximized their monetary payoffs in the valuation task for the object with known value. For those subjects who did not do so, we find little evidence of systematic mistakes in one direction or of differences between buyers and sellers.*

After subjects indicated their WTP or WTA in Part I, we asked them to identify, for every possible random price realization, their final payoff given their choice at that price. Overall, 54 out of 69 buyers (78.3 %) and 58 out of 70 sellers (82.9 %) correctly provided their final payoffs for every possible price on the first attempt. Hence, this second measurement tool shows that on aggregate, 80.6 % of subjects indicated good understanding of the payoff implications of their choices.⁹

In Part II, subjects indicated their WTP and WTA for a box of chocolates. We again obtain WTP and WTA estimates for each subject by taking the midpoint of the prices at which a subject switched from buying to not buying or from not selling to selling. Taking all subjects into account, the mean WTA for the box of chocolates was significantly higher than the WTP (4.78 CHF vs. 8.62 CHF; Wilcoxon rank-sum test, $p < 0.01$). Hence, taking into account all subjects, our study replicates the common finding of a substantial WTP-WTA gap.

However, our main research question is whether the gap remains if we consider only those subjects whose choices indicate that they understood the incentive properties of the BDM mechanism. To see whether heterogeneity in subject misconceptions can account for the WTP-WTA gap in Part II, we separately identify “sophisticated” subjects who appear to understand the incentive properties of the BDM mechanism. This group consists of those subjects who chose the payoff maximizing switching point in Part I (between prices of 8 and 9 CHF) and who

⁸ Given that we use a price-list version of the BDM with increments of 1 CHF, we cannot rule out that we classify a subject who switched between 8.00 and 9.00 CHF as behaving consistently with a valuation of 8.50 CHF even though this subject would have stated a switching point other than exactly 8.50 CHF in a stated-price version of the BDM mechanism. Strictly speaking, we thus interpret choices within 0.50 CHF from the real price of 8.50 CHF as consistent with payoff-maximizing behavior. (Note, however, that subjects whom we later classify as “sophisticated” also had to identify the payoff consequences of their choices and we did not allow for any error margin in this respect.) Cason and Plott (2014) employ a narrower tolerance interval, of plus or minus \$0.05 from the real value of \$2, for their classification of payoff-maximizing behavior. However, this difference in size of the intervals does not account for the higher fraction of “confused” subjects in their study. Applying a \$0.50 interval to Cason and Plott’s data (available online) only increases conformance with payoff maximization to 23.27 %.

⁹ Out of those 53 buyers with estimated WTP of 8.50 CHF, 43 correctly identified all the final payoffs on the first try and so did 11 out of the 16 buyers with estimated WTP different from 8.50 CHF. Out of those 45 sellers with estimated WTA of 8.50 CHF, 40 correctly identified all the final payoffs on the first try and so did 18 out of 25 with estimated WTA different from 8.50 CHF. Therefore, for both buyers and sellers, the proportion of those who correctly entered their final payoff on the first try is higher for subjects who made the payoff maximizing choice (81.1 % vs. 68.8 % and 88.9 % vs. 72 %, respectively).

Table 2 Probit regressions

Dependent variable: “sophistication”	(1)	(2)	(3)
Raven test score	0.072*** (0.023)		0.065*** (0.025)
CRT score		0.085** (0.039)	0.028 (0.043)
Observations	139	139	139
Pseudo R ²	0.059	0.026	0.061

The dependent variable is a binary variable taking on value 1 if a subject is categorized as “sophisticated” and 0 otherwise. *Raven test score* denotes the number of correctly solved puzzles (out of 12). *CRT score* measures the number of correct answers in the cognitive reflection test (out of 3). The coefficients report marginal effects from a probit regression calculated at the mean. Robust standard errors are reported in parenthesis

*** Denote significance at 1 %, ** at 5 %, and * at 10 %

correctly identified the final payoffs for all possible prices on their first attempts in both Part I and Part II. Hence, this group, which consists of 41 buyers and 40 sellers,¹⁰ or 58.3 % of our total sample, satisfies fairly strong requirements for being classified as understanding the incentive properties of the BDM mechanism. This group did not only make the payoff-maximizing choice in the Part I valuation task—which is Cason and Plott’s test for BDM game-form sophistication—but also provided two sets of 21 correct payoff entries, each on their first attempt.

The remaining group, which we label “confused,” comprises subjects who failed at least one of the three conditions. Hence, this group either made dominated choices in Part I, or made at least one mistake in one of the two sets of payoff entries.

A closer look at the composition of the two groups reveals that those subjects who score higher on two measures of cognitive ability—the Raven Progressive Matrices test and the Cognitive Reflection Test—are significantly more likely to be “sophisticated;” see Models (1) and (2) in Table 2, respectively. Model (3) reveals that only the Raven Progressive Matrices test remains significant when we consider both tests simultaneously. Hence, by at least one measure, subjects in the “sophisticated” group are smarter on average, which lends further support to the notion that those subjects understood the BDM mechanism correctly.

As Table 1 reveals, there is no difference in WTP estimates and a marginally significant difference in WTA estimates for the box of chocolate between the sophisticated and confused subjects [WTP (sophisticated) = 4.63 CHF vs. WTP (confused) = 5.00 CHF, Wilcoxon rank-sum test, $p = 0.711$; WTA (sophisticated) = 7.65 CHF vs. WTA (confused) = 9.91 CHF, Wilcoxon rank-sum test, $p = 0.054$].¹¹ Furthermore, the distributions of estimated WTP and WTA values are

¹⁰ Only 1 out of 70 sellers and 3 out of 69 buyers incorrectly entered their final payoffs in Part II on their first attempt.

¹¹ We find a similar pattern in Part I where there is no difference in WTP estimates and a marginally significant difference in WTA estimates between sophisticated and confused subjects (Wilcoxon rank-sum test: $p = 1.000$ and $p = 0.052$, respectively).

similar for the two groups and do not differ significantly (Kolmogorov–Smirnov test, $p = 1.000$ and $p = 0.451$, respectively).¹²

Finally, and most importantly for the research question addressed in this paper, WTA estimates are significantly greater than WTP estimates, even when we only consider sophisticated subjects (4.63 CHF vs. 7.64 CHF, Wilcoxon rank-sum test, $p = 0.005$). This difference is also evident when comparing the distributions in the right panel of Fig. 1 (Kolmogorov–Smirnov test, $p = 0.010$). Hence, we have our main result: WTP–WTA gaps persist even in the absence of game form misconceptions.

Result 2: *Estimated WTA for a box of chocolates is significantly higher than estimated WTP for the same box of chocolates. This result holds even for subjects who revealed a high level of understanding of the incentive properties of the BDM mechanism.*

4 Conclusion

We follow up on the claim by Cason and Plott (2014) that “the failure of game form recognition can masquerade as support for the theory of framing, such as preferences constructed from reference points.” They support this claim by showing that a BDM value elicitation experiment in which there is a lot of misconception can produce data that could be misinterpreted as reflecting support for a non-standard preference theory of framing or endowments, while the behavioral patterns in the data merely reflect subjects’ game form misconception. While this claim might be true in some data sets, such as the one collected by Cason and Plott, they leave unanswered the important question whether such behavioral patterns persist for subjects who do not suffer from game form misconceptions. That is, they leave open the question of whether subject misconception is *necessary* to produce such behavioral effects.

In our paper, we employ a design that allows us to identify whether subjects comprehend the incentive properties of a price-list version of the BDM mechanism. Our main result is that the WTP–WTA gap in a standard valuation task persists for subjects who revealed their understanding of the game form of the BDM mechanism. We can thus conclude that game form misconceptions are not a necessary condition for the emergence of a WTP–WTA gap.¹³

¹² We find no significant relation when we regress, separately and jointly, the WTA of sellers or the WTP of buyers in Part II on the subjects’ Raven test or CRT scores. However, the coefficients for the effects on the WTA are both negative and not close to zero, which suggests that subjects who score higher on these tests tend to state somewhat lower WTA values. The coefficients for the effects on the WTP show no such relationship; they are very close to zero and differ in signs.

¹³ In related research, Fehr et al. (2015) fail to replicate Plott and Zeiler’s (2005) findings that the WTP–WTA gap disappears when procedures, such as training rounds, are employed to reduce misunderstanding.

Finally, our study also provides some evidence that supports Cason and Plott's argument that researchers must be careful when using the BDM mechanism. We, too, found a non-trivial proportion of subjects, about 30 %, that make sub-optimal choices when stating valuations for a good with known value—even though we used procedures intended to minimize confusion regarding the mechanism. Furthermore, we find some evidence that supports Cason and Plott's claim that confusion can lead subjects to display behavior that appears consistent with framing effects. First, while the proportion of overpaying to underpaying buyers in Part I is perfectly balanced, it is not so for sellers—more sellers overpaid than underpaid in Part I. This latter difference could be misinterpreted as evidence of framing effects, though it is not statistically significant in our data. Second, sellers who we classified as “confused” stated, on average, a marginally significantly higher WTA than “sophisticated” sellers in both Parts I and II. This biases the data in the direction of what is commonly interpreted as a framing or endowment effect, by increasing the WTP-WTA gap. However, for buyers in Part II the opposite holds: “Confused” buyers stated, on average, a higher WTP than “sophisticated” buyers. This goes in the opposite direction of misconceptions masquerading as typical framing effects, though the difference is not statistically significant.

Overall, our results do not contradict Cason and Plott's argument that misconceptions can lead to data that looks consistent with and might thus be misinterpreted as evidence for non-standard preferences. However, we demonstrate that a large and significant WTP-WTA gap remains even absent any evidence of subject misconceptions.

Acknowledgments We thank the editor, Nikos Nikiforakis, and two anonymous referees for insightful and constructive comments and participants in the Reading Group in Experimental Economics at the University of Zurich for valuable discussion.

References

- Becker, Gordon M., DeGroot, Morris H., & Marschak, Jacob. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 226–232.
- Bock, Olaf, Baetge, Ingmar, & Nicklisch, Andreas. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71, 117.
- Cason, Timothy N., & Plott, Charles R. (2014). Misconceptions and game form recognition: challenges to theories of revealed preferences and framing. *Journal of Political Economy*, 122(6), 1235–1270.
- Ericson, Keith M., Marzilli, E., & Fuster, A. (2014). The endowment effect. *Annual Review of Economics*, 6, 555–579.
- Fehr, D., Hakimov, R., Küber, D. (2015). The willingness to pay-willingness to accept gap: A failed replication of plott and zeiler, WZB Discussion Paper SP II 2015-204.
- Fischbacher, Urs. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10, 171–178.
- Frederick, Shane. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Kahneman, Daniel, Knetsch, Jack L., & Thaler, Richard H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, 98(6), 1325–1348.

- Kahneman, Daniel, Knetsch, Jack L., & Thaler, Richard H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1), 193–206.
- Murphy, James J., Stevens, Thomas H., & Yadav, Lava. (2010). A comparison of induced value and home-grown value experiments to test for hypothetical bias in contingent valuation. *Environmental and Resource Economics*, 47(1), 111–123.
- 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. *American Economic Review*, 95(3), 530–545.
- Plott, Charles R., & Zeiler, Kathryn. (2007). Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory? *American Economic Review*, 97(4), 1449–1466.
- Raven, John, Raven, John C., & Court, John. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. San Antonio: Harcourt Assessment.
- Weaver, Ray, & Frederick, Shane. (2012). A reference price theory of the endowment effect. *Journal of Marketing Research*, 49(5), 696–707.