WTP and WTA in repeated trial experiments: Learning or leading?

Gwendolyn C. Morrison *

School of Economics, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

Received 2 April 1998; received in revised form 20 April 1999; accepted 31 May 1999

Abstract

Repeated trials are often employed in economics experiments. The rationale is that by allowing respondents to ‘practice’ in the (often unusual) market situations in which experiments place them, they are given the opportunity to refine their responses to more accurately reflect their preferences. This paper presents the results of a pilot experiment in which repeated trials are used to elicit and compare willingness-to-pay (WTP) and willingness-to-accept (WTA) measures of value while controlling for other sources of disparity that have been proposed in the literature. Other studies have found that WTP and WTA do not converge over repeated trials, but this study finds the more disturbing result that the disparity between the two does not even decrease over repeated trials. However, individual respondents show evidence of learning over the series of trials and so there appears to be value in the practice of incorporating repeated trials in experiments. © 2000 Elsevier Science B.V. All rights reserved.

PsycINFO classification: 2260; 2340; 2343

JEL classification: D1; C9

Keywords: Experimental design; Perception; Willingness to pay; Willingness to accept; Learning effects

* Tel.: +44-115-951-4768; fax: +44-115-951-4159.
E-mail address: wendy.morrison@nottingham.ac.uk (G.C. Morrison).
1. Introduction

Repeated trials are often employed in economics experiments (e.g., Brookshire & Coursey, 1987; Shogren, Shin, Hayes & Kliebenstein, 1994; Morrison, 1997a). The rationale is that such experiments typically ask respondents to perform tasks that are out of the ordinary in that either they are valuing a good for which there is not normally a private market, or the response mode by which respondents are asked to express their value for the good is unfamiliar to them. Allowing respondents to ‘practice’ in this unusual market situation through repeated trials, gives them the opportunity to refine their response to more accurately reflect the value they place on the good.

Since Hicksian consumer theory predicts that willingness to pay (WTP) and willingness to accept (WTA) measures of value are equal, when economics experiments revealed a persistent disparity between WTP and WTA measures of value (e.g., Knetsch & Sinden, 1984; Knetsch, 1989; Kahneman, Knetsch & Thaler, 1990), experimentalists turned their attention to determining the source of this disparity. Hanemann (1991) argued that the disparity stemmed from WTP and WTA questions being posed such that the two related to movements along different indifference curves and, thus, different marginal rates of substitution applied. Thaler (1980) considered the possibility that people’s preferences are dependent upon their initial endowment looking at alternatives in terms of a gain or loss relative to that reference point – this is termed the endowment effect. Coursey, Hovis and Schulze (1987) (CHS) argued the importance of allowing respondents to learn, showing that in some cases, repeated trials led to large reductions in the disparity between the two measures. Dubourg, Jones-Lee and Loomes (1994) considered that the disparity may be a symptom of respondents having imprecise preferences and therefore being unable to pinpoint their true valuation of a good.

The experiment discussed here was designed to examine WTP and WTA responses while controlling for substitutability, learning, and imprecise preferences. The experiment allows us to examine changes in the WTP/WTA disparity over repeated trials. Means of testing for evidence of convergence

---

1 This is rooted in Kahneman and Tversky’s (1979) ‘prospect theory’.
between WTP and WTA are described, and evidence of leading over the course of the repeated trials is presented.

2. Background

Different studies have tried to control for different explanations. Morrison (1997a) notes that WTA exceeds WTP even when controlling for Haneemann’s argument of differing marginal rates of substitution and allowing learning through the use of repeated trials. Dubourg et al. (1994) found that the WTP/WTA disparity persists even when accommodating imprecise preferences by allowing individuals to give a response range and comparing the upper-bound WTP and lower-bound WTA responses. And Morrison (1998) found that WTA exceeded WTP even controlling for Hanemann’s argument, imprecise preferences, and learning effects.

In the context of the WTP/WTA disparity, it was anticipated that accommodating learning by asking respondents to complete repeated trials would lead to the convergence of the two values. But, given that the disparity remains even after repeated trials, there are two important questions that must be asked. Firstly, is there any evidence that the disparity between two measures of value does get smaller after repeated trials as found by CHS? If the difference between the two is smaller for the final than for the first trial, then perhaps the values would have converged had there been more trials. If, on the other hand, responses remain pretty much constant from the first to the last trial, then such repetition in experiments is unnecessary. The costs of repeated trials are time and respondent fatigue – as a result, fewer experimental questions can be asked. Secondly, even if individuals change their responses over the different trials, is there any evidence that they are learning? If respondents are changing their responses randomly rather than honing in on their actual value, then the additional costs of using repeated trials are not justified since there are no benefits. Or worse, what if there is a pattern to changes in stated values between trials, but it is the process of repeating trials itself that influences results? As suggested by Knetsch and Sinden (1987), “the procedure may shape, as well as reveal, preferences.” If the outcome of one trial leads the respondent to change their response, not as a result of learning, but because some aspect of the experimental market is leading their response in a particular direction (not necessarily pulling WTP closer to WTA), then repeated trials can introduce bias into the evaluation process rather than refining results as intended.
3. Experimental design

An experiment was run in November 1995 using a sample of 20 students taking a course in experimental economics at the University of St Andrews, Scotland. Each participant was randomly allocated to one of two experimental groups. The subjects were informed that they would be paid £4 for participating in the experiment, plus or minus any money that they traded during the experiment. For WTP questions, respondents were not restricted to spending only the £4 participation payment, but could also spend their own money.  

The experiment was comprised of two parts, the first endowing WTA respondents with a chocolate bar (which everyone was informed could be purchased from a near by shop for 33 pence) and elicited WTA and WTP values for the chocolate. This part of the experiment corresponds to Hane-mann’s (1991) explanation of the WTP/WTA disparity in that WTA respondents’ decisions related to movements along higher indifference curves than the WTP respondents’. Respondents were asked to indicate three key values on the answer sheets 3 in each of five trials: the minimum (maximum) values that they were sure they would be willing to accept (pay), the maximum (minimum) amount that they were sure they would not accept (pay), and the amount that made them feel indifferent between accepting the money and keeping the chocolate. 4 This approach is intended to help the respondent hone in on their value for the good. The amount that makes them feel indifferent between the money and the chocolate bar is taken to be their ‘best estimate’ of the value they place on the good. At the end of each trial, answers were collected and the price of exchange (chosen at random from a range of values prior to the experiment) was announced. At the end of the final trial, one of the five trials was randomly chosen to be the binding trial. That is, if the number four was chosen then anyone that stated a WTA less than or equal to the price of exchange in the forth trial would, at the end of the experiment, give up the chocolate bar and be paid the amount that they indicated would make them indifferent between accepting the money and

---

2 When invited to participate, it was suggested to students that they might want to bring some money with them to the experiment.
3 Respondents were told that they could use the values on the answer sheets or write in different values if they wished.
4 This response procedure is taken from Jones-Lee, Loomes and Philips (1995). The instructions given to respondents are presented in Appendix A.
keeping the chocolate. 5 Or, in the WTP section, anyone that stated a WTP greater than or equal to the price of exchange in the randomly chosen trial, would be given a chocolate bar in exchange for their stated WTP.

The second part of the experiment followed a similar format but those that had answered WTP questions with respect to the chocolate bars (i.e., group 2) were given University of St Andrews coffee mugs 6 and the others were given a cash redeemable voucher. Following the same routine as in the first part of the experiment, WTA responses were obtained, in five trials, from those endowed with a mug. The mean stated WTA from the fifth trial was calculated and this value was assigned to the cash redeemable voucher held by the WTP respondents. Before answering WTP questions they were told that the voucher could be ‘cashed in’ at the end of the experiment. The intention here was to compensate WTP respondents sufficiently for both WTA and WTP questions to approximate equal and opposite movements along the same indifference curve.

4. Results

4.1. Magnitude of disparity over repeated trials

First let us address the question of whether or not the disparity between the two measures of value converges after repeated trails. The basic results of the experiment, presented in Morrison (1997a) showed that for all five trials eliciting a value for the chocolate bar the hypothesis that $WTA = WTP$ cannot be rejected at the 10% level of significance, whereas for all five trials valuing a mug the null that $WTA = WTP$ must be rejected at $z = 0.01$. Bateman, Munro, Rhodes, Starmer and Sugden (1997) suggest that WTP/WTA disparities might be expected to be smaller in the case of regularly consumed goods – “subjects might be expected to be surer of their own preferences”. And Morrison (1997b) showed how the endowment effect, if present, would be expected to lead to a smaller WTP/WTA disparity for

---

5 In fact, when both parts of the experiment had been completed and payments were made to those stating a WTA less than or equal to the price of exchange, the chocolate bars were not taken back from them. Since the experiment was over at that stage, this oversight will not have affected the results.

6 Everyone was told that the same kind of mug could be purchased for £1.90 at the students’ union shop.
goods with closer substitutes. A common chocolate bar such as that used in the experiment could be expected to be both more regularly consumed and to have closer substitutes than the University of St Andrews mug.

Regardless of the reasons for the discrepancy between the findings of the two parts of the experiment, for the chocolate the two measures of value were not significantly different in the first trial, so it would seem that nothing was gained by having four more trials. However, for the mug, not only is there a substantial difference between the two measures (WTA was approximately double WTP), but WTA and WTP did not converge over repeated trials. It could be that five trials were not enough for the two to converge. If the difference between WTA and WTP declined over the five trials, then perhaps additional trials would have lead to convergence. To consider this proposition, it is necessary to examine whether a pattern emerged over the course of the five trials. Table 1 and Fig. 1 show the relationship between mean WTA and WTP responses for all five trials of both parts (chocolate bar and mug) of the experiment. Consider the ratio of and difference between WTA and WTP in the experiment eliciting values for the chocolate bar. Whereas in the first trial WTA is roughly equal to WTP, the disparity between WTA and WTP actually increased over each of the five trials. This is true even when normalising by considering (WTA–WTP)/WTA as illustrated in Fig. 1. So, contrary to CHS, the two values actually diverged over each of the trials.

The second part of the experiment, which elicited values for a University of St Andrews coffee mug, follows a different pattern. Mean WTA is initially about twice as big as WTP. The disparity grows over the next two trials and, although it begins to decline toward its original value over the remaining two trials, it remains larger in the final than in the first trial. Furthermore, examining the normalised difference between WTA and WTP in Table 1 and Fig. 1 we see that there is really very little change in that disparity over all five trials.

So, contrary to expectations (e.g., Shogren, Shin, Hayes & Kliebenstein, 1994) and the findings of CHS, this study found that WTA and WTP esti-

---

7 Morrison (1997b) uses a pivot of an indifference curve from a reference point to illustrate that, once the individual is endowed with a consumption bundle, that bundle (which becomes the reference point) is on a higher utility curve than before the individual was endowed. The utility function might take the following form: \( U = f(S, X, \text{loss}) \), where utility is positively related to the quantity of money and of the good \( X \), but where a loss of money (as in a WTP question) or a loss of a quantity of \( X \) (as in a WTA question) will yield a reduction in utility. Shogren and Hayes (1997) counter-argument does not hold unless we allow utility functions to undergo some additional transformation (i.e., unless the indifference curves are bent or stretched in addition to pivoting from the reference point).
mates of value not only do not converge over repeated trials, but we found no
evidence to suggest that the disparity even decreases. Indeed, in the first part
of the experiment the two measures of value actually diverged. Therefore, if
this hypothesised convergence is the only argument in favour of using re-
peated trials in experiments, then the argument is not supported and the
additional cost (in time and respondent fatigue if not financial cost) of re-
peating trials is not justified.

4.2. Learning?

The next question to address is whether or not individual respondents
exhibit signs of learning over the course of several trials. The evidence just

<table>
<thead>
<tr>
<th>Chocolate</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTA:WTP</td>
<td>0.990:1 (0.993:1)(^a)</td>
<td>1.092:1</td>
<td>1.086:1</td>
<td>1.125:1</td>
<td>1.131:1</td>
</tr>
<tr>
<td>WTA–WTP</td>
<td>−0.003 (−0.002)(^a)</td>
<td>0.026</td>
<td>0.026</td>
<td>0.035</td>
<td>0.037</td>
</tr>
<tr>
<td>(WTA – WTP)/WTA</td>
<td>−0.010 (−0.007)(^a)</td>
<td>0.084</td>
<td>0.079</td>
<td>0.111</td>
<td>0.116</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mug</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WTA:WTP</td>
<td>2.009:1</td>
<td>2.219:1</td>
<td>2.420:1</td>
<td>2.293:1</td>
<td>2.188</td>
</tr>
<tr>
<td>WTA–WTP</td>
<td>1.100</td>
<td>1.225</td>
<td>1.285</td>
<td>1.235</td>
<td>1.200</td>
</tr>
<tr>
<td>(WTA – WTP)/WTA</td>
<td>0.502</td>
<td>0.549</td>
<td>0.587</td>
<td>0.564</td>
<td>0.543</td>
</tr>
</tbody>
</table>

\(^a\) One WTA respondent in the first trial valuing the chocolate bar seemingly reversed their ‘best estimate’
and ‘lower bound’ WTA (i.e., best estimate < lower bound). The responses in brackets are changed around
for these calculations.

Fig. 1. Normalised difference between mean WTA and WTP.
presented in Section 4.1 shows that WTA and WTP measures of value diverged over repeated trials when valuing chocolate and that when valuing a mug the (significant) difference between the measures remained fairly constant. With Hicksian consumer theory dictating the equality of these values (at least with respect to the mugs), neither finding would seem to suggest that practice gained through repeating trials served as a learning experience for respondents. But, in order to assess whether respondents learned by going through repeated trials, it is necessary to examine the data at the individual level.

If respondents do not modify their answers with experience, then it is reasonable to conclude that they have not gained anything from that experience. A response being changed in every trial and yet not following any recognisable pattern might be interpreted in this way as well. However, if an individual does modify their answer after the first trial, and they eventually hone in on another answer, then it would appear that they have refined their response in the light of experience – that is, they have learned. Conversely, if subjects change their responses in such a way that they are following the (randomly selected) price of exchange from one trial to the next, then rather than learning over the repeated trials they are simply being led by a random element of the experimental design.

First let us consider what might constitute evidence in support of or against the hypothesis that individuals learn over the course of several trials. If a person is converging on their value for a good, then one would expect that the (absolute value of the) difference between their responses in the first and second trial would be larger than that between the fourth and fifth trial. This should hold regardless of whether WTA or WTP is used, so the two types of responses are combined \( n = 20 \) in testing this hypothesis. Wilcoxon matched-pairs signed-ranks tests and paired \( t \)-tests were used to test the following hypothesis for both the chocolate and mugs: \( H_0: |v_5 - v_4| = |v_2 - v_1| \), and \( H_1: |v_5 - v_4| < |v_2 - v_1| \), where \( v \) denotes the stated WTA or WTP value \( (n = 20) \) and the number indicates the trial number. \(^8\) With respect to the experiment eliciting values for chocolate bars, the null hypothesis had to be rejected in favour of the alternative at \( \alpha = 0.05 \) using both the Wilcoxon and paired \( t \)-tests. For the experiment valuing mugs, we were

\(^8\) A non-parametric test is included because individuals indicated their responses on an answer sheet with a payment scale. This may have affected the distribution of responses.
narrowly unable to reject the null (though could reject $H_0$ at $\alpha = 0.0508$) using the Wilcoxon test, but using the paired $t$-test the null was again rejected in favour of the alternative at $\alpha = 0.05$. 9 The finding that individual respondents change their stated WTA or WTP by a smaller magnitude between the last two trials than between the first two trials suggests that they have some degree of increased confidence in the value they assign the good. We interpret these results as a sign of learning.

Another possible indicator that people are learning would be if the number of respondents that stated the same WTA (WTP) in subsequent trials increased with each completed trial. That is, if individuals stick with the same value in consecutive trials, it would seem that they are confident in their response; if they keep changing their responses then it would seem that they are not yet sure of the value they place on the good. The number of people that stated the same value from one trial to the next is shown in Table 2 for all four pairs of trials. Considering first the chocolate bar experiment, we see that only two people (or only one person if we rearrange the answers of the respondent that seemingly reversed their ‘best estimate’ and ‘lower bound’ responses in trial (1) stated the same WTA in trials 1 and 2, but 7 of the 10 respondents stated the same value in trial 4 as in trial 5. It would appear, then, that members of this group were unsure of their WTA in the first trial, but that they had found the value that they were comfortable with by the fourth trial. The WTP respondents seem more confident initially with 5 out of 10 stating the same WTP in the trials 1 and 2, but then that number does not change over the repeated market trials. The same pattern can be seen in the experiment involving coffee mugs. Six of the 10 respondents stated the same WTP in the first two trials and that only grew to 7 in the last two trials. However, only three WTA respondents stated the same value in the first and second trials, and that number increased to 9 of the 10 individuals by the final two trials. Since, those that answered WTP questions in the chocolate experiment then answered WTA questions in the mug experiment, and vice versa, the different response patterns in WTP and WTA cannot simply be attributed to a difference in the samples. It stands to reason that people are more confident initially with their WTP responses than they are with their WTA responses since people regularly decide whether or not to purchase goods. The WTP format is probably unusual to people in that most of us are

---

9 Chocolate: Wilcoxon, $z = -2.5558$; paired $t$-test, $t = 2.54$. Mug: Wilcoxon, $z = -2.2357$; paired $t$-test, $t = 2.50$. 

faced with a price and simply decide whether or not to pay it, but a WTA question is still more foreign in that it must be decided how much compensation is needed to give up the good. 10

Although the number of respondents that maintained the same stated WTP from one trial to the next did not increase over the repeated market trials, the number of individual WTA respondents that did so at least tripled between the first two and last two trials in both experiments. So the repeated market trials seem to have enabled these individuals to learn. If we consider only the seven WTA and the five WTP respondents that gave the same value in the fourth and fifth trials the ratio of WTA to WTP in the chocolate experiment is 1.051:1, while it is 1.131:1 for the full samples. In the mug experiment this ratio is 2.013:1 when only considering the nine WTA and seven WTP respondents that gave consistent answers in the last two trials, which is

10 Hence the NOAA panel’s suggestion that only closed-ended (yes/no) WTP questions be used in contingent valuation surveys designed to elicit values from the public (Arrow, Solow, Portney, Leamer, Radner & Schuman, 1993).

<table>
<thead>
<tr>
<th></th>
<th>All equal</th>
<th>Trial 1–2</th>
<th>Trial 2–3</th>
<th>Trial 3–4</th>
<th>Trial 4–5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chocolate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTA: (n = 10)</td>
<td>0(1)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow $p_{t-1}^*$</td>
<td>6(7)a</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WTA$<em>t$ = WTA$</em>{t-1}$</td>
<td>2(1)a</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Away from $p_{t-1}^*$</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WTP: (n = 10)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow $p_{t-1}^*$</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5b</td>
<td></td>
</tr>
<tr>
<td>WTP$<em>t$ = WTP$</em>{t-1}$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Away from $p_{t-1}^*$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Mugs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTP: (n = 10)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow $p_{t-1}^*$</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WTP$<em>t$ = WTP$</em>{t-1}$</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Away from $p_{t-1}^*$</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WTA: (n = 10)</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow $p_{t-1}^*$</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WTA$<em>t$ = WTA$</em>{t-1}$</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Away from $p_{t-1}^*$</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

a See note in Table 1.
b Reject null of ‘no leading’ at $\alpha = 0.05$. 
only slightly smaller than the full group ratio of 2.188:1 but these respondents constitute a high proportion of the sample. So the disparity between the two measures of value is smaller for those that appear to have honed in on a value. This suggests that the respondents that did not change their answer in the last two trials had indeed learned.

By the same token, respondents exhibiting a general downward trend in WTA or increase in WTP over the repeated market trials would appear to be using experience to close in on a value. This is what is expected to happen over repeated market trials (e.g., Shogren et al., 1994). Using Friedman two-way ANOVA the hypothesis that the responses each individual gave were equal in all five trials could not be rejected, even at the 35% level of significance, for any of the four parts of the experiment. 11 So neither the WTP nor the WTA responses appear to be repeatedly revised upward or downward over the five trials. In order to assess whether such a pattern is evident at a pairwise level – that is, whether a significant number of respondents revised their stated values in the same direction from one trial to the next – the \( p \) values of Wilcoxon tests are presented in Table 3. One pair of WTP responses were found to be significantly different at \( z = 0.05 \), and one WTP and one WTA pair at \( z = 0.10 \). 12 The surprising thing about these results is that in all three of these cases, the significant change was in the wrong direction: WTP values were revised down while WTA values were increased. This is not incompatible with learning over repeated trials, but it is obviously incompatible with convergence. Such an adjustment of responses over trials would lead to the divergence of WTP and WTA.

In summary, there is some evidence that respondents hone in on a value. Firstly, examining the absolute difference between values stated in consecutive trials we found that the results involving the chocolate suggest that respondents are learning, while the results from the mug valuation do not refute that hypothesis. Secondly, the number of individuals stating the same value in consecutive trials tripled from the first two and final two trials for the WTA respondents, while the corresponding number among WTP respondents remained steady – the WTA responses support, and the WTP responses do not refute, the hypothesis that respondents learned. Thirdly, those that state the same value across the last two trials (and hence could be argued to

11 Chocolate WTA: \( \chi^2 = 4.14 \); Chocolate WTP: \( \chi^2 = 3.12 \); Mug WTA: \( \chi^2 = 0.38 \); Mug WTP: \( \chi^2 = 3.44 \). In all cases \( n = 10 \) with 4 degrees of freedom.

12 Paired \( t \)-tests echoed these results with the null being rejected for only one pair of values.
have honed in on their value for the good) also give a smaller WTP/WTA disparity, thus providing further evidence of learning. Finally, the finding that neither WTP nor WTA responses appear to be repeatedly revised upward or downward is not incompatible with learning. Hence, there is some evidence suggesting, and none refuting, the hypothesis that respondents do learn over the course of repeated trials.

4.3. Leading?

Now let us address the question of leading. This entails testing the hypothesis that responses are essentially following the randomly chosen exchange price. That is, if their stated WTP or WTA is less than (greater than) the exchange price in trial 1, $p_1$, do respondents revise their previous WTP or WTA response downward (upward) in trial 2? If they do, then they would appear to be following the exchange price. If they changed their stated value in the opposite direction or if they stated the same value in both trials, then they clearly are not following. To test whether respondents are following the exchange price in this way, individual responses in each trial were compared with the price of exchange in the previous trial ($p_{t-1}$). Table 2 shows, for each pair of trials, how many respondents fall into each of three categories: (i) $WTP_{t-1} > p_{t-1}^*$ and $WTP_t < WTP_{t-1}$ or vice versa; (ii) they state the same value in both trials, $WTP_t = WTP_{t-1}$; or, (iii) the respondent changes their response but moving away from $p_{t-1}^*$ if $WTP_{t-1} > p_{t-1}^*$ and $WTP_t > WTP_{t-1}$ or vice versa. The same relationships relate to categorising the WTA responses. The respondent is following $p_{t-1}^*$ if their responses fit into category (i), but not if they are of the types described for categories (ii) or (iii).

<table>
<thead>
<tr>
<th>Trial 1–2</th>
<th>Trial 2–3</th>
<th>Trial 3–4</th>
<th>Trial 4–5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chocolate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0$: WTA$ = $WTA$_{t-1}$</td>
<td>0.2361</td>
<td>0.0929$^a$</td>
<td>0.6121</td>
</tr>
<tr>
<td>$H_0$: WTP$ = $WTP$_{t-1}$</td>
<td>0.8927</td>
<td>0.2249</td>
<td>0.0431$^{a,b}$</td>
</tr>
<tr>
<td>Mugs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0$: WTP$<em>t = $WTP$</em>{t-1}$</td>
<td>0.1441</td>
<td>0.0679$^a$</td>
<td>0.4652</td>
</tr>
<tr>
<td>$H_0$: WTA$<em>t = $WTA$</em>{t-1}$</td>
<td>0.5541</td>
<td>0.7998</td>
<td>0.6547</td>
</tr>
</tbody>
</table>

$^a$Significant but in the wrong direction for convergence (i.e., WTP decreasing, WTA increasing).

$^b$Must reject at $z = 0.05$ using paired $t$-tests. In all cases $n = 10$.
Examination of the results presented in Table 2 might cause one to suspect that respondents are being led by $p_{t-1}$. In both the WTA and WTP sections of the chocolate experiment far more respondents changed their responses in a way that conforms with, rather than is incompatible with, leading. The data from the mug experiment does not provide as strong support for the hypothesis that respondents are being lead, however there are still more individuals who appear to be following rather than moving away from $p_{t-1}$. However, as striking as the results in Table 2 are, Wilcoxon matched-pairs signed-ranks tests (and paired $t$-tests) revealed that they are not significant. The hypothesis that there is no change in individual responses from one trial to the next could not be rejected at the 5% level of significance for all but one pair. WTP respondents in the chocolate experiment did change their value from the fourth to the fifth trial in such a way that they appeared to be following $p_{t-1}$. That is, a significant number of those that in the fourth trial stated a WTP greater than (less than) $p_{t-1}$ proceeded to reduce (increase) their WTP in the fifth trial.

Although the pattern of responses does appear to conform to the hypothesis that respondents are being lead by the exchange prices used in the experiment, there is no statistical evidence to support this. Thus, this experiment does find evidence that individuals are learning over repeated trials, but none to suggest that they are being led.

5. Conclusion

The purpose of this paper is to examine two questions. Firstly, does the difference between WTA and WTP values decrease over repeated trials and, secondly, regardless of whether or not they do, is there evidence of learning or do random elements in each trial appear to lead subjects’ responses in subsequent trials? This experiment finds no evidence that the disparity between WTA and WTP decreases over repeated trials, however there is evidence to suggest that respondents did learn over the course of the five trials. Most respondents did change their WTP (WTA) during the repeated trials and yet the majority had arrived at what was their final stated value by the penultimate trial. Restating the same answer in the last two trials suggests a degree of confidence in that response. Although cursory examination of the pattern of responses over the series of trials does seem to indicate that individuals modify their responses to follow the randomly chosen exchange price from the preceding trial, these suspicions are not supported by
statistical evidence. So there is no evidence of leading. Although the disparity between WTA and WTP did not decrease over the trials, the use of repeated trials does not appear to have introduced a bias, and yet does appear to have facilitated learning. Therefore, it is concluded that the practice of incorporating repeated trials in experiments is valuable in that it allows individual respondents to learn, and hence hone in on a value, over the series of trials.

Acknowledgements

My thanks to Shane Bonetti for his assistance in running the experiment, to the University of St Andrews students that took part, to the Department of Economics at St Andrews University for providing the funds to conduct this experiment, and to the referees for their comments and suggestions.

Appendix A. Instructions: Part I

This experiment consists of two parts. It is necessary to complete both parts in order to receive the participation fee. Please do not talk during the experiment.

Group 1: In the first part, you will be asked to indicate the minimum amount that you would be willing to accept (WTA) to sell the bar of chocolate that you have been given. You have been provided with answer sheets with a series of values upon which you can indicate your WTA for the chocolate bar. Specifically, you will be asked to mark key values on the list as follows:

(✓) beside the minimum amount that you are sure you would be willing to accept in exchange for the chocolate
(✗) beside the maximum amount that you are sure you would not be willing to accept in exchange for the chocolate
(*) beside the amount at which you feel indifferent between accepting the money and keeping the chocolate

N.B.: The list of values is only there to help you. If the value that you would like to use is not on the list provided, then you can add it to the list.

There will be five trials. In each, you will be asked to indicate your WTA (✓, ✗, *) on the relevant response sheet and your response sheet will be
collected from you. Then the price for that trial, which has already been
drawn at random from the price range on your answer sheet, will be an-
nounced. You may wish to keep a note of your WTA responses, and the price
drawn, for each trial on the back of your instruction sheet.

After all five trials have been completed, a number will be drawn indicating
which of the five trials will be binding. That is, if the number 3 is drawn, then
anyone that in the third trial stated a WTA less than or equal to the price
drawn will give up the chocolate bar in exchange for their stated WTA. All
exchanges will be made at the end of the experiment.

References

the NOAA panel on contingent valuation. *Federal Register, 58*(10), 4601–4614.


Coursey, D. L., Hovis, J. L., & Schulze, W. D. (1987). The disparity between willingness to accept and


Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the

47*(2), 263–291.


507–521.

Economics, 102*, 691–695.


