Reexamining Coherent Arbitrariness for the Evaluation of Common Goods and Simple Lotteries[§]

Drew Fudenberg[*], David K. Levine[**] and Zacharias Maniadis[***]

Abstract

We reexamine the effects of the anchoring manipulation of Ariely, Lowenstein and Prelec (2003) on the valuation of common market goods and found very weak anchoring effects. We perform the same manipulation on the evaluation of binary lotteries, and find no anchoring effects at all. This suggests limits on the robustness of anchoring effects.

[§] We thank Drazen Prelec for many helpful conversations, and for sharing the experimental instructions and data from Ariely, Lowenstein and Prelec (2003). We are also grateful to Tore Ellingsen and Oscar Bergman for sharing their instructions and data with us, and to Kathy Zeiler and for sharing the lottery data from Plott and Zeiler (2005). We would also like to thank Nikolaos Georgantzis, Fabio Tufano, Charlie Plott and especially Jane Klobas, for useful comments. NSF grants 0646816 and SES-08-51315 provided financial support.

[*] Harvard University
[**] Washington University of St Louis
[***] Bocconi University, Department of Decision Sciences, and Carlo F. Dondena Center

1. Introduction

Decades of evidence from individual-choice experiments have documented that subjects do not always satisfy the assumptions of consumer theory and of expected utility theory. Some departures from the standard model can be accommodated while maintaining the assumption that people do have stable preferences. Experiments on anchoring effects are more troubling as they suggest that many people do not have fixed or slowly changing preferences. A defender of the standard models might argue that this is not relevant to most everyday consumer decisions, and that an average person with enough experience will have relatively fixed preferences for common market goods. This suggests that experience with the goods and decisions in question may be important in the stability of preferences. Since people are relatively familiar with risky choices, this in turn suggests that decisions about uncertain prospects (lotteries) might be relatively stable.

If the monetary valuation of goods and gambles is based on constant underlying preferences, choices should be independent of irrelevant prior "anchoring" questions. In the case of lotteries, Johnson and Schkade (1989) showed that asking subjects whether they would prefer a fixed amount (the anchor) to a given gamble affected subsequently stated certainty equivalents. Chapman and Johnson (1999) had subjects generate a random anchor from their SSN, and found very strong anchoring effects on the stated willingness to accept (WTA) for selling a purely hypothetical lottery.

In the case of goods Ariely, Lowenstein and Prelec (2003, henceforth ALP) used a simple anchoring manipulation to provide evidence that people do not have constant preferences. They first generated a personal random number between 0 and 99 for each subject (the anchor), asking them to turn the last two digits of their Social Security Number (SSN) into a dollar price. They then asked subjects whether they would buy particular consumer goods for this price. Subsequently, they elicited subjects' willingness to pay (WTP) for the goods using the Becker-DeGroot-Marschak (1964) (BDM) mechanism,[1] and found that WTP was significantly correlated with the anchors.

[1] This mechanism is incentive compatible under risk neutrality, but not in general. However, there is strong evidence that experimental subjects are not risk neutral even for small gambles.

These results seem troubling for economic theory. First, they are large. ALP write: "The effect is even more striking when examining the valuations by quintiles of the social security number distribution […] the valuations of the top quintile subjects were typically greater by a factor of *three.* For example, subjects with SSN in the top quintile were willing to pay $56 on average for a cordless computer keyboard, compared with $16 on average for subjects with bottom quintile numbers. Evidently, these subjects did not have, or were unable to retrieve personal values for ordinary products." Similarly, the average stated WTP for a bottle of average wine, for subjects whose SSN-based anchor price was in the highest quintile, was $27.90, more than three times as high as the analogous average WTP for subjects, whose SSN-based anchor price was in the lowest quintile. Moreover, ALP present evidence that their results are robust across many dimensions: repetition, different levels of simplicity of the "hedonic experience" (consuming a good is one such experience), different subject pools, different levels of monetary incentives, and markets vs. individual choice. We believe that the ALP experiments are important and influential enough that they deserve to be replicated and tested for robustness.

We conducted four sets of experiments in order to examine whether preferences for common market goods and for simple lotteries are "arbitrary" in the sense of ALP. We examined choices for both lotteries and goods. In the case of goods we examined WTP and WTA and whether or not explaining the incentive properties of the BDM procedure matters. For lotteries we focused on eliciting WTA with the BDM incentives explained. These experiments enable us to examine new questions as well as examining robustness: Are preferences for lotteries or common goods more arbitrary? Are anchoring effects more pronounced for WTA or WTP?

Our first finding is that we are unable to replicate the results of ALP: we find very weak anchoring effects both with WTP and with WTA. The Pearson correlation coefficients between the anchor and stated valuation are generally much lower than in ALP, and the magnitudes of the anchoring effects (as measured by the ratio of top to bottom quintile) are smaller. Repeating the ALP procedure for lotteries we do not find any anchoring effects at all.

Unlike ALP we carried out laboratory rather than classroom experiments. This necessitated some minor changes – discussed below – from ALP's procedures. It is conceivable that these changes are responsible for the differences in our findings; if so the robustness of their results is limited.

2. Experiment 1: Eliciting WTA for Common Market Goods

All of our experiments were paper-and-pencil as was the case in ALP. Our first experiment reexamined the effects of random anchors on subjects' valuation for common market goods. The experiment took place in the California Social Science Laboratory (CASSEL) at UCLA, in August of 2009, and subjects were UCLA students. Subjects were recruited using the standard procedure of the lab, which is sending an announcement to a large email list. A total of 84 subjects participated,[2] and each of the three sessions had 26-28 subjects.

Our subject pool consisted almost exclusively of undergraduate students, and the goods were chosen to be of interest for them: an academic planner, a cordless keyboard, a financial calculator, a designer book, a pack of quality chocolates, and a cordless mouse. The average retail price of the goods was $51.70. When the experiment started, subjects were asked to press the "Enter" key, and generate a random number between 0 and 99 (we used the program "Excel"). Subjects were then asked to copy this number six times in their answer sheet. After subjects had done so, we started reading the instructions loudly, so that everyone could hear.

For each good, subjects had to answer two questions. The first (anchor) question asked them whether they would be willing to sell the good for an amount of money equal to the number that they had written down. In the second question, subjects were asked to state the minimum price for which they would be willing to sell the good. They were instructed to act as if the good was theirs to keep since there was a positive probability

[2] Although 84 subjects participated, five subjects were excluded from the data analysis. One subject failed to write a WTA for any of the goods, and another subject failed to write a WTA for one of the six goods. One subject failed to write the same anchor number for the six goods. One subject wrote a random number equal to $100, which was not in the range of possible random numbers (0-99). Finally, one subject failed to answer any of the anchoring questions (the first question for each of the goods), leaving the space blank. It is worth emphasizing that we have not excluded a subject whose WTA for the chocolates and the cordless mouse was $100 and $300, respectively, because it is not a priori impossible for someone to have such WTA for these common goods. So our actual sample size consists of 79 subjects.

that they would truly own it. We explained that three subjects would be randomly drawn. Each of the three chosen persons would be assigned one of the goods (a different good for each of the three subjects drawn). For each of them, one of her answers (corresponding to her assigned good) would be carried out for real.[3]

If a subject was the first person randomly drawn, her answer to the first question would be consequential. So if she stated that she wanted to sell the good for an amount equal to her random anchor she did so. Otherwise she kept the good and received it at the end of the experiment. If a subject was the person randomly drawn second or third, her answer to the second question (for the assigned good) would matter. We explained that in this case, a price would be randomly drawn from a matrix containing 100 values. The values of the matrix were drawn before the experiment from a triangular distribution in [0,100] with a mode equal to zero. Subjects were shown the matrix on the screen. If the stated WTA was higher than the random price, the subject kept the good, and if not, she received a monetary amount equal to the random price, and sold the good. We also explained why the BDM mechanism gives an incentive to truthfully reveal preferences.[4]

We explained that all uncertainty would be resolved using physical devices. The random price would be chosen from the 100 possible prices by throwing a 10-sided dice twice. The three consequential goods would be determined by throwing a six-sided dice three times. The three chosen subjects would be chosen using a bingo cage with balls numbered according to numbers pre-assigned to the subjects.

After we had finished explaining the instructions, we showed the six goods to subjects. We displayed each good in clear view, and a photograph of each good was also

---

[3] This process emulates the procedures of ALP.

[4] Prelec explained the BDM procedure as part of his class prior to conducting the experiment. As we used paid participants from a laboratory pool, we instead explained the procedure as part of the experiment. We based our explanation on the instructions of Plott and Zeiler (2005). In particular, we said: "Please, leave your instructions aside for a while. With the second question, we want to know how much the opportunity to have each item is worth to you. This method simply gives you an incentive to state the true minimum you would be willing to accept for selling each item. Why is your best strategy to write the minimum you would be willing to accept? Because you will not receive the amount you ask for. Instead, if you receive anything, you will receive the random price. Let's try to understand this with an example. Imagine that I own the item. Say I believe that this item is worth 500$ for me. What happens if I falsely say that the minimum I would be willing to accept is more than 500? Assume I say 600$. If the random price is, say, 530, I do not sell the item. But, had I said 500, I could receive 530 for an item that I think is only worth 500. I lose out. What happens if I say less than 500? For example, I say only 400. If the random price is, say, 420, then I have to sell my item. I lose out, because I have to sell for 420 an item that I think is worth 500." Note that these instructions implicitly tell subjects to be risk neutral.

projected on a screen. After showing each good we gave subjects time to complete the answers to both questions. After they finished, they were asked to copy their responses to a second answer sheet. After we picked up the second answer sheet, we resolved all uncertainty as explained, and concluded the experiment.

Given the results of ALP, our prior expectation was that we would find a positive relationship between the personal random number and the stated valuation. Thus in all our experiments, we use one-tailed statistical tests. For example, we reject the null hypothesis of a zero Pearson correlation between the personal random number and the WTA at a certain significance level if we get a high enough positive value of the sample correlation.

The average cash payment per subject was $13.80. Moreover, subjects received goods of total retail value equal to $298.00, which implies that the average value of earned goods per subject was about $3.50.[5]

Table 1 presents the main results. For each of the five quintiles of the distribution of the personal random number, we report the mean and the median WTA (the medians are in the parentheses). We also report the Pearson correlation coefficients, both to understand whether there is a linear relationship between the WTA and the anchor and in order to allow comparisons with ALP. As seen in the table, the Pearson correlation coefficients between the random anchor and stated WTA were relatively small: they range from –0.108 to 0.207, and only one of them is statistically significant. We also report the (non-parametric) Spearman correlation coefficient and find similar values. In contrast, the Pearson correlation coefficients in ALP's study range from 0.319 to 0.516 and are all highly statistically significant, despite their smaller sample size.[6]

To get further evidence for the importance of anchoring, we examine the ratio of stated WTA for subjects in the highest quintile of random numbers to those in the lowest quintile. Recall that ALP report ratios of about 2.3 and 2.2 for two of the goods, and between 3 and 3.45 for the remaining four goods. Table 2 reports our results. For one of our goods the ratio was 2.9, for the other five the ratio ranged from 1.1 to 1.4. Using the

---

[5] Each of the participants who were not chosen received a participation fee of $13. Chosen persons also received this fee, in addition to being assigned one good.
[6] ALP's sample was 55 subjects, we had 79.

(nonparametric) Wilcoxon rank-sum test we are unable to reject the null hypothesis of homogeneity at the 5% significance level for the five goods with ratios between 1.1 and 1.4. For the remaining good (the financial calculator) the hypothesis is rejected at the 5% significance level. Notice that if we test six different true hypotheses with different data it is not unlikely that one of them will be rejected at the 5% level. On the other hand all the ratios are greater than 1, which is not likely if there is no anchoring at all.[7]

Table 1. Experiment 1, average and median WTA per quintile of anchor's distribution, and correlations.

Quintile (# Obs.)	**Academic Planner**	**Cordless Keyboard**	**Financial Calculator**	**Designer Book**	**Milk Chocolates**	**Cordless Mouse**
0-19 (13)	7.00 (5)	38.85 (35)	10.23 (10)	11.30 (10)	5.62 (4)	22.07 (25)
20-39 (16)	12.44 (10)	50.19(47.5)	20.37 (20)	14.75 (15)	17.88 (7)	42.88 (23)
40-59 (16)	11.50 (10)	51.94 (50)	20.94 (15.5)	18.12(17.5)	8.47(5.25)	24.56 (20)
60-79 (20)	10.05 (8)	38 (30)	18.78 (11)	16.5 (13.5)	7.47 (5)	19.95 (15)
80-99 (14)	7.64 (7.5)	47.28 (30)	31.92 (23.5)	15.57 (15)	7.36 (6.5)	26.85 (20)
All data	9.90 (9)	45.07 (40)	20.46 (15)	15.46 (15)	9.46 (6)	27.10 (20)
Pearson	-0.047	-0.022	0.207	0.134	-0.096	-0.108
p-value	0.66	0.58	0.034	0.12	0.80	0.83
Spearman	-0.0458	-0.0590	0.2075	0.1261	0.0704	-0.1334
p-value	0.66	0.70	0.033	0.134	0.27	0.73

Table 2. Experiment 1, Wilcoxon test of homogeneity of the distributions of WTA for subjects in the first and fifth quintile of the range of random numbers

Item	**Academic Planner**	**Cordless Keyboard**	**Financial Calculator**	**Designer Book**	**Milk Chocolates**	**Cordless Mouse**
z	-0.587	-0.685	-3.008	-1.424	-1.342	-0.171
p-value	0.279	0.247	0.001	0.077	0.09	0.432

3. Experiment 2: Eliciting WTA for Lotteries

In our second experiment we elicited WTA for six lotteries with a range of expected values similar to the range of the market prices of the six goods. The experiment also took place in the CASSEL lab at UCLA during November of 2008 and August of

the sample of ALP consisted of 55 subjects, whereas we had 79 subjects.

[7] The statistical tests reported in the test included all of our data; in the online appendix we verify that our conclusions about which effects are significant does not change if we exclude outliers and subjects with inconsistent responses, as has been done in some past work on this topic.

2009. A total of 110 subjects participated,[8] in four sessions of 26-28 subjects each. Subjects were UCLA students, and the great majority was undergraduate students. The average total earnings per subject were equal to $21.7.[9]

The six gambles are shown in Table 3. The procedures and the instructions were similar to Experiment 1, with lotteries instead of goods.[10] In each experimental session, three subjects would be drawn at random, and each would get to own one lottery for real, and would either keep it or sell it depending on the answers given to the experimental questions. The lottery was carried out for real, using a dice. Subjects, who chose to keep them, by indicating a WTA higher than the random price, would receive the prize drawn from the lottery. Otherwise, they simply received the random price from selling the lottery. All of this was made common knowledge to subjects.

We again examine the ratio of stated WTA for subjects in the highest quintile of random numbers to those in the lowest quintile. Three of the ratios lie between 0.99 and 1.05, and the remaining three lie between 1.2. and 1.4. This is smaller even than the modest effect we found for goods. Despite the larger sample size none of Wilcoxon rank-sum tests reported in Table 4 are significant at the 5% level. Nor are any the Pearson and Spearman correlation coefficients in Table 3.

---

[8] We exclude from the data a person who did not state a WTA for one of the lotteries. We also exclude a participant who stated a WTA equal to $300 for the first lottery, and a WTA $400 for the sixth lottery. Since the maximum prize of any lottery is $100, we felt that it is impossible that someone could actually have such a WTA for the lotteries. Therefore, our final sample size is 108.

[9] All subjects received a participation fee of $16. The three subjects who were randomly chosen to receive a lottery could earn money on top of that.

[10] The only difference was that in this experiment we further explained to subjects how they should answer the questions. In particular, after presenting the two questions, we drew attention away from the written instructions, and told subjects: "Please go back to the last two pages of your instructions but please do not write anything yet. Please have a look at the lotteries and the questions you are asked to answer. The first question for each lottery is easy: do you prefer to have the opportunity to play this lottery or the specified amount of money? An example will illustrate what you should answer to the second question. Imagine that I own the right to play a particular lottery. Say this sample lottery ticket. How do I know what amount is the minimum I 'd be willing to accept for the lottery? First of all, notice that, if I keep it, this lottery ticket gives me the opportunity to earn some money with a certain probability. I am not willing to give it up without getting some money. The question is, what is the minimum I would accept? Start with 100$. Would I be willing to give up my lottery in exchange for 100? If so, decrease the amount to 95. If I am willing to accept 95 in exchange for the lottery, then decrease further. I keep decreasing until I come to the amount that makes me indifferent between keeping the lottery and getting the money. Your objective is to think like this for all six lotteries. You should think carefully before answering these questions because these questions are not purely hypothetical. There is a chance you will actually own each lottery, and each of your decisions may be carried out for real money."
This part was also based on the instructions by Plott and Zeiler (2005).

The lack of anchoring in our experiment might be due to the simplicity of the lotteries we considered. Chapman and Johnson (1999) used a SSN-based anchor, but for the complicated binary lottery ($287, 17%; $18.5, 83%), and found a Pearson correlation coefficient equal to 0.45, whereas the largest coefficient that we find for any lottery is equal to 0.082.

Table 3. Experiment 2, average and median WTA per quintile of anchor's distribution, and correlations.

Quintile (# Obs.)	**0.5, $100; 0.5, $0**	**0.25, $100; 0.75, $0**	**0.1, $100; 0.9, $0**	**0.75, $60; 0.25, $20**	**0.5, $60; 0.5, $20**	**0.9, $60; 0.1, $20**
0-19 (18)	42.11 (40)	24.67 (20)	14.67 (10)	41.11 (40)	34.39 (35)	49.39 (52)
20-39 (13)	49.69 (50)	26.62 (25)	15.77 (11)	44.69 (45)	37.00 (40)	48.69 (50)
40-59 (22)	51.54 (50)	31.82 (32.5)	23.55 (20)	46.05 (45)	37.86 (40)	55.32 (55)
60-79 (21)	41.67 (50)	22.19 (25)	10.23 (10)	39.67 (45)	34.67 (40)	49.52 (55)
80-99 (34)	50.02 (50)	31.45 (27)	21.15(13.5)	43.15(42.5)	34.18 (30)	51.38(52.5)
All data	47.35 (50)	28.01 (25)	17.78 (10)	42.90 (45)	35.40(39.5)	51.17 (55)
Pearson	0.057	0.082	0.071	-0.010	-0.076	0.032
p-value	0.28	0.20	0.23	0.54	0.78	0.37
Spearman	0.119	0.090	0.034	-0.026	-0.104	0.025
p-value	0.11	0.17	0.36	0.60	0.86	0.40

Table 4. Experiment 2, Wilcoxon test of homogeneity of the distributions of WTA for subjects in the first and fifth quintile of the range of random numbers.

Lottery	**0.5, $100; 0.5, $0**	**0.25, $100; 0.75, $0**	**0.1, $100; 0.9, $0**	**0.75, $60; 0.25, $20**	**0.5, $60; 0.5, $20**	**0.9, $60; 0.1, $20**
z	-1.577	-1.312	-1.125	-0.371	0.304	-0.485
p-value	0.058	0.095	0.13	0.355	0.619	0.313

4. Experiment 3: Willingness to Pay

In order to examine whether the differences of our results from ALP's were driven by the fact that we elicited WTA and ALP elicited WTP, we performed an additional treatment. This treatment also took place at the CASSEL lab at UCLA, in April of 2010, and the participants were UCLA students. A total of 79 subjects participated, in three sessions of 26 or 27 students.[11] The average payment per subject was equal to $21.8.[12]

[11] One subject failed to answer any of the six anchoring questions and was excluded for the analysis. Hence, our final data set consists of 78 subjects.

[12] The subjects who were not chosen received a participation fee of $13.

The sessions were nearly identical to those of Experiment 1. The main difference was that subjects were asked to state the most they would be willing to pay for the goods.[13] Hence, the subjects that would be chosen at random would not own the good, but they would have the chance to buy it. To conform to the standards of the CASSEL lab, a large participation fee ($93) was offered to the three chosen subjects to make sure they could buy each good, and in order to prevent subjects from losing money in the experiment.

The results are shown in Table 5. We again examine the ratio of stated WTP for subjects in the highest quintile of random numbers to those in the lowest quintile. One ratio is quite large, 1.4. Four are less than one and lie between 0.9 and 1.0, and the sixth, the financial calculator, has a very low ratio of 0.7, while in other treatments it had a high ratio. However none of the ratios are statistically significant at the 5% level according to the Wilcoxon rank-sum test. Moreover the Pearson and the Spearman correlations are all very close to zero.[14]

From this data, we can conclude that anchoring effects are not stronger for WTP than WTA. Our data is consistent with the possibilities that anchoring effects are weaker for WTP than WTA and that there are negative anchoring effects, but we cannot make definitive statements about these issues.

5. Experiment 4: Is it the BDM explanations?

In Experiments 1-3 we included part of the explanations of the BDM mechanism that was used by Plott and Zeiler (2005); our objective was to try to account for the explanations that were provided as a part of a classroom lecture before the experiments of

ALP. As we could not replicate their classroom training it is possible that this is why our results differ from theirs. To examine the importance of the explanation of the BDM elicitation procedure our final experiment provides no explanations of the concept of WTP/WTA or of the BDM mechanism.

[13] Four of the six goods were the same as in Experiment 1, while two goods were not the same, but very similar. A Class Organizer was used instead of an Academic Planner, and a very similar Cordless Mouse was used instead of the one used in Experiment 1.

Table 5. Experiment 3, average and median WTP per quintile of anchor's distribution, and correlations

Quintile (# Obs.)	**Class Organizer**	**Cordless Keyboard**	**Milk Chocolates**	**Cordless Mouse**	**Designer Book**	**Financial Calculator**
0-19 (20)	3.25 (1)	30.70 (21)	4.82 (4)	16.75 (14.5)	7.70 (4.5)	15.05 (11.5)
20-39 (11)	2.95 (2)	33.45 (30)	4.77 (4)	15.63 (10)	5.55 (4)	12.09 (10)
40-59 (12)	4.00 (2)	32.33 (35)	4.83 (5)	17.58 (15)	8.91 (5)	7.66 (6)
60-79 (19)	4.37 (3)	33.37 (25)	5.00 (5)	16.95 (20)	8.32 (4)	13.74 (10)
80-99 (16)	4.81 (1)	28.31 (25.5)	4.19 (3)	12.87 (10)	6.25 (3)	9.81 (6)
All data	3.92 (1.5)	31.50 (25)	4.73 (4)	15.97 (15)	7.43 (4.5)	12.1 (8.5)
Pearson	0.096	-0.003	-0.043	-0.039	-0.013	-0.08
p-value	0.2	0.51	0.65	0.63	0.54	0.76
Spearman	0.045	0.009	-0.007	-0.029	-0.013	-0.074
p-value	0.35	0.47	0.52	0.6	0.55	0.74

Table 6. Experiment 3, Wilcoxon test of homogeneity of the distributions of WTP for subjects in the first and fifth quintile of the range of random numbers

Item	**Class Organizer**	**Cordless Keyboard**	**Milk Chocolates**	**Cordless Mouse**	**Design Book**	**Financial Calculator**
z	0.148	-0.191	0.562	0.959	0.500	1.024
p-value	0.558	0.424	0.713	0.831	0.691	0.847

The experiment was also conducted at the CASSEL laboratory at UCLA, in February of 2011, and all subjects were UCLA students. A total of 81 subjects participated in three sessions of 27 subjects each.[15] Subjects' average cash payment was equal to $15.64 and in addition subjects received goods of total value equal to $153, for an average value of earned goods per subject of about $1.88.[16]

The experiment was conducted in exactly the same way as Experiment 1, except without the additional explanations of the BDM mechanism.[17] The type and quality of each good was the same as in Experiment 1, except for using a telephone address book instead of an academic planner.

[14] As have earlier researchers, we find a possible WTP-WTA gap, however it is a relatively small one, in the order of 20%-30%.

[15] One subject failed to state a WTA for one of the goods, and another one failed to copy the correct random number (wrote different numbers rather than a single one). These subjects were excluded from the observations, leaving a total of 79 subjects.

[16] Each of the participants who were not chosen received a participation fee of $13. Chosen persons also received this fee, in addition to being assigned one good.

[17] The written instructions were short. Appendix 2 contains the instructions for the lotteries experiment, which are similar to the written instructions used in the other three experiments.

The results, shown in Table 7, are similar to our other three experiments. We again examine the ratio of stated WTA for subjects in the highest quintile of random numbers to those in the lowest quintile. One of these is 1.5 (this time the cordless mouse rather than the financial calculator). Four lie between 1.0 and 1.1, and the sixth is 1.2. Needless to say the Wilcoxon rank-sum tests, shown in Table 8, are not significant. The Pearson correlation coefficients are very low although for one good the null hypothesis of a zero correlation is rejected at the 5% significance level.

Table 7. Experiment 4, average and median WTA per quintile of anchor's distribution, and correlations

Quintile (# Obs.)	**Address Book**	**Cordless Keyboard**	**Financial Calculator**	**Designer Book**	**Milk Chocolates**	**Cordless Mouse**
0-19 (16)	8.31 (7.5)	39.69 (35)	20.81(18.5)	14.31 (15)	7.28 (6.5)	19.93 (20)
20-39(15)	10.06(8.95)	31.60 (26)	21.40 (20)	17.13 (10)	5.93 (5)	17.26 (17)
40-59(15)	12.60 (10)	42.53 (40)	17.80 (18)	20.2 (20)	11.13 (6)	25.46 (20)
60-79(20)	6.55 (5)	41.05 (40)	21.80 (15.5)	17.35 (15)	5.10 (5)	20.85 (19)
80-99(13)	10.38 (10)	41.62 (40)	21.69 (20)	16.3 (15)	7.69 (6)	29.30 (25)
All data	9.35 (8)	39.35 (35)	20.74 (18)	17.06 (15)	7.27 (6)	22.25 (20)
Pearson	0.0007	0.0974	0.0217	0.0414	0.0028	0.2024
p-value	0.497	0.197	0.423	0.359	0.49	0.037
Spearman	-0.0127	0.1228	0.0166	0.0120	-0.0415	0.1521
p-value	0.544	0.140	0.442	0.458	0.642	0.091

Table 8. Experiment 4, Wilcoxon test of homogeneity of the distributions of WTP for subjects in the first and fifth quintile of the range of random numbers

Item	**Address Book**	**Cordless Keyboard**	**Financial Calculator**	**Designer Book**	**Milk Chocolates**	**Cordless Mouse**
z	-0.904	-0.529	-0.507	-0.199	-0.309	-1.323
p-value	0.183	0.298	0.306	0.421	0.379	0.092

6. Conclusion

There were several differences between our experimental design and ALP. They elicited subjects' WTP for six common goods: Belgian chocolates, a wireless keyboard, a design book, a bottle of rare wine, a bottle of cheap wine, and a cordless trackball. They performed the experiment in a class with 55 MBA students, and decisions were

consequential for only six subjects, drawn at random, one for each good. We conducted the experiments in a lab environment, and we changed some of the goods, because of technology changes and age restrictions.[18] In the first three treatments we used part of the Plott and Zeiler (2005) explanation of the BDM mechanism, but the fourth treatment suggests that this deviation from ALP was not a key factor in the earlier results.

We also preferred not to use the SSN-based anchor, since during a pilot some subjects showed privacy concerns. This might be another possible reason for the difference in our results. There is evidence that subjects might not view such numbers as truly arbitrary. For example, one-third of participants in one of the experiments of Chapman and Johnson (1999) stated that they though the SSN-based anchor number was informative. In another experiment, 57% said that they thought that the experimenter wanted the number to influence their judgment. Perhaps a SSN-based anchor is not seen as random, in the sense that the value of the SSN has been determined long ago.[19]

Our results do not confirm the very strong anchoring effects found in ALP. They are more in agreement with the results of Simonson and Drolet (2004), who used the same SSN-based anchor as ALP, and found no anchoring effects on the stated WTA, and moderate anchoring effects on WTP for four common consumer goods. Bergman et al. (2010) also used the design of ALP for six common goods, and found anchoring effects, but of smaller magnitude than in ALP.

Tufano (2009) and Maniadis and Tufano (2011) also failed to confirm the robustness of the magnitude of the anchoring effects of ALP, using hedonic experiences, rather than common goods. Tufano used the anchoring manipulation to increase the variance in subjects' WTA for a bad-tasting liquid, but the manipulation had no effect. Notice that this liquid offers a simple (negative) hedonic experience, like the "annoying sounds" of ALP. Maniadis and Tufano replicated the "annoying sounds" treatment (Experiment 2) of ALP and found weaker (and non-significant) anchoring affects.

Overall our results suggest that anchoring is real – it is hard to reconcile otherwise the fact that in the WTA treatment with goods the fact that the ratios between highest and

---

[18] Wine could not be offered to students less than 21 years old, who are the majority of UCLA undergraduate students. Cordless trackballs were largely unknown to undergraduate students at the time of the experiment.

[19] We would like to thank Bill Zame for pointing out this possibility to us.

lowest quintile is always bigger than one – but that quantitatively the effect is small. Additionally our data supports the idea that anchoring goes away when bidding on objects with greater familiarity, such as lotteries. Our results do not suggest that anchoring is stronger with WTP than WTA – if anything the reverse is true.

References

Ariely, Dan, George Lowenstein, and Drazen Prelec. " 'Coherent Arbitrariness': Stable Demand Curves without Stable Preferences". The Quarterly Journal of Economics, Vol. 118, No. 1, 2003, pp. 73-104.

Becker, Gordon, Morris DeGroot, and Jacob Marschak. "Measuring Utility by a Single-Response Sequential Method". Behavioral Science 9, No. 3, 1964, pp. 226-232.

Bergman, Oscar, Tore Ellingsen, Magnus Johannesson and Cicek Svensson. "Anchoring and Cognitive Ability". Economic Letters, Vol. 107, No. 1, 2010, pp. 66-68.

Chapman, Gretchen and Eric Johnson. "Anchoring, Activation, and the Construction of Values". Organizational Behavior and Human Decision Processes, Vol. 79, No. 2, 1999, pp. 115-153.

Johnson, Eric and David Schkade. "Bias in Utility Assessments: Further Evidence and Explanations". Management Science, Vol. 35, No. 4, 1989, pp. 406-424.

Maniadis, Zacharias and Fabio Tufano. "One Swallow does not Make a Summer: Research Discoveries and (non)Replications: the Case of Anchoring in Economics. Unpublished manuscript.

Plott, Charless and Kathryn Zeiler. "The Willingness to Pay/ Willingness to Accept Gap, the 'Endowment Effect', Subject Misconceptions and Experimental Procedures for Eliciting Valuation". The American Economic Review, Vol. 95, No. 3, 2005, pp. 530-545.

Simonson, Itamar and Aimee Drolet (2004), "Anchoring Effects on Consumers Willingness-to-Pay and Willingness-to-Accept". Journal of Consumer Research, Vol. 31, No. 4, 681-690.

Tufano, Fabio. "Are 'True' Preferences Revealed in repeated Markets? An Experimental Demonstration of Context-Dependent Valuations". Experimental Economics, Vol. 13, No. 1, 2009, pp. 1-13.

Web APPENDIX

1: Robustness Check of our Results

We tried to follow the ALP experiment as much as possible, so we did not include a long practice process, or very detailed instructions. Nevertheless, as we explained above, there is reason to believe that the notions of WTA and WTP, as well as the BDM mechanism, are not very easy to understand. Accordingly, some participants in our experiment showed signs of bidding "strategically", or otherwise showed signs that they have not fully understood the valuation task. Moreover, several subjects had their second answers inconsistent with their response in the first question, for the same good or lottery.

In Experiment 1, one subject assigned the same WTA ($20) to all goods, and another stated implausibly similar WTA for each of the six goods (in a range from $22 to $30). We believe that this behavior indicates insufficient understanding, because it is highly implausible that the actual WTA for so different goods lies in such a narrow range. Another subject stated a WTA of equal to $5 for goods 1-4 and equal to $100 and $300 for goods 5 and 6, respectively, which highly unlikely to express true valuation. Moreover, several subjects had "inconsistent" responses. Accordingly, we considered a total of 15 subjects to have "questionable responses".

In Experiment 2, one subject assigned the same WTA ($5) to all lotteries. Four subjects stated only two different WTA values for the six lotteries (for example, one specified $97 or $30, another specified $10 or $56). Three subjects stated an implausibly similar WTA for the different lotteries (for one subject, the WTA of all lotteries ranged from $15 to $18, for another from $48 to $60 and for the third from $46 to $60). This behavior might indicate improper understanding of the task, for the reasons explained above. Moreover, six other subjects stated implausibly large WTA values (of at least $50 for each lottery) or wrote a WTA higher than the highest possible prize. We believe that this might be caused by the tendency to behave "strategically", asking for as much as possible in a bargaining situation. It is generally difficult to understand the BDM mechanism, so this is a real possibility. A few subjects also had inconsistent responses. We assigned a total of 16 subjects in Experiment 2 as having "questionable responses".

In Experiment 3, some subjects stated improbably low willingness to pay for all the goods (less than $10 for any of the goods). The above tendency to behave strategically in a bargaining situation is a plausible reason. Moreover, one subject stated a WTP of 50 for the good with the lowest average WTP (the first good, the class organizer), and a WTP less than 10 for all the other goods. This might indicate confusion. There were also a few inconsistent subjects, and, in total, 11 subjects in Experiment 3 had "questionable responses".

In Experiment 4, two subjects stated the same WTA for all goods, one subject stated a WTA of either $20 or $5, and four subjects stated and excessively narrow range of WTA for the six goods (all WTA were within $10 from each other). Several other subjects stated inconsistent responses, and a total of 18 observations were characterized as "questionable".

We should note that in Session 4 of Experiment 2, one subject complained that some other subject/s had not written down the actual excel random number, but a different one (but clicking "enter" multiple times). We should note that in this type of experiments, subject might write a different anchor number than the one that they are instructed to write. If the anchor question is incentivized, then it is in the subjects' best interest to write a number as large as possible in WTA experiments, and a number as low as possible in WTP experiments.[20] Excluding from the sample the eight subjects with anchor numbers higher than 80, in Session 4, we believe that we exclude the most suspicious observations. The possibility of cheating can never be ruled out, and excluding the suspicious observations leaves a very reasonable distribution of random numbers for all our experiments.

Characterizing observations as "questionable" or "suspicious" has necessarily a subjective, and perhaps arbitrary, element. However, we feel that it is useful to examine whether our results are driven by confusion or cheating. For doing this, we can see whether excluding these subjects significantly changes the results. Table 7 describes the correlations between the anchor number and WTA/WTP for the samples where the questionable and suspicious observations are excluded: 15 observations from the data of

Experiment 1, 23 observations from the data of Experiment 2, 11 observations from the data of Experiment 3, and 18 observations of Experiment 4.

Table 1. Correlations for the samples that exclude objectionable and suspicious observations

Experiment 1 N=64	**Academic Planner**	**Cordless Keyboard**	**Financial Calculator**	**Designer Book**	**Milk Chocolates**	**Cordless Mouse**
Pearson	0.033	-0.063	0.264	0.12	0.054	-0.137
p-value	0.399	>0.5	0.017	0.172	0.336	>0.5
Experiment 2 N=85	**0.5, $100; 0.5, $0**	**0.25, $100; 0.75, $0**	**0.1, $100; 0.9, $0**	**0.75, $60; 0.25, $20**	**0.5, $60; 0.5, $20**	**0.9, $60; 0.1, $20**
Pearson	0.041	0.112	0.052	-0.047	-0.083	-0.123
p-value	0.35	0.133	0.318	>0.5	>0.5	>0.5
Experiment 3 N=67	**Class Organizer**	**Cordless Keyboard**	**Milk Chocolates**	**Cordless Mouse**	**Designer Book**	**Financial Calculat.**
Pearson	-0.026	-0.03	-0.067	-0.055	-0.020	-0.018
p-value	>0.5	>0.5	>0.5	>0.5	>0.5	>05
Experiment 4 N=61	**Address Book**	**Cordless Keyboard**	**Milk Chocolates**	**Cordless Mouse**	**Designer Book**	**Financial Calculat.**
Pearson	-0.001	-0.005	-0.022	0.164	-0.017	-0.090
p-value	>0.5	>0.5	>0.5	0.103	>0.5	>0.5

The data of Table 1 reveal that the anchoring effects, measured as the Pearson correlations of the anchor with the stated valuation, are similar for the sample that excludes objectionable and suspicious observations. Note that very similar results are obtained, in Experiment 2, if we exclude only the objectionable observations, but not the suspicious ones, and vice versa. Therefore, the subjective element does not seem to affect the main results of our experiments.

[20] The mean of the anchor number that subjects wrote in Experiment 1 was 51.4 and the median 55, in Experiment 2 the mean was equal to 56.7 with a median equal to 60. In Experiment 3, the mean 49.7 was and the median was equal to 53. In Experiment 4, the mean was 50.1 and the median 54.

2: Instructions for the Lotteries Experiment (Experiment 2)

Welcome to CASSEL. Please turn off pagers and cellular phones now. It is important that you do not talk, or with any other way try to communicate during the experiment. Thank you for participating in this pricing exercise. We are interested in how much you value various lotteries. We will show you six different lotteries, and then ask you a pair of price-selling intent questions about each lottery:

- First, we will ask whether you would like to sell the lottery at a particular price. Your computer generated that price randomly.
- Next, we will ask you to state the minimum that you would be willing to accept for selling the lottery.

Three persons will be chosen at random from the class and the decisions of each about one of the lotteries will be carried out for real money. One lottery will be sold on the basis of answers to the first question, and two lotteries on the basis of answers to the second question. Each person is on the running for only one lottery.

<u>If you are chosen and the first question counts for you</u>, then we will look to see whether you decided to sell, or not to sell, the lottery at your random computer-generated price. If you stated that you wanted to sell the lottery at that price, then we will buy it from you at that price. If you stated that you did not want to sell the lottery at that price, you will keep it.

<u>If you are chosen and the second question counts</u>:

- The price of the lottery will be determined by drawing a number at random from the matrix in the overhead slide.
- If the value is smaller than this random price, then the lottery will be sold *at that randomly selected price*.
- If the value stated is higher than the price, then the lottery will not be sold.

Because your answer does not affect the amount you get paid, only whether you sell, it's to your advantage to state the minimum that you would be willing to be paid for selling the lottery.

Each of you will receive a 12-dollar participation fee. In addition, if you are one of the three persons chosen, all results from lotteries you have chosen to keep will be drawn right away. Any amount that you will earn from the sale of a lottery, or from the results of a lottery you keep, will be added to your participation fee. You will know your total earnings when you leave the experiment.

First, please copy your computer-generated random number in each of the 6 boxes

LOTTERY TICKET 1

50% chance of	50% chance of
$ **100**	$ **0**

Would you sell this lottery ticket for $ ____ Circle YES or NO

The least I would be willing to receive to sell this lottery is $ _____

LOTTERY TICKET 2

25% chance of	75% chance of
$ **100**	$ **0**

Would you sell this lottery ticket for $ ____? Circle YES or NO

The least I would be willing to receive to sell this lottery is $ _____

LOTTERY TICKET 3

10% chance of	90% chance of
$ **100**	$ **0**

Would you sell this lottery ticket for $ ____? Circle YES or NO

The least I would be willing to receive to sell this lottery is $ _____

LOTTERY TICKET 4

75 % chance of	25% chance of
$ **60**	$ **20**

Would you sell this lottery ticket for $ ____? Circle YES or NO

The least I would be willing to receive to sell this lottery is $ _____

LOTTERY TICKET 5

50% chance of	50% chance of
$ **60**	$ **20**

Would you sell this lottery ticket for $ ____? Circle YES or NO

The least I would be willing to receive to sell this lottery is $ _____

LOTTERY TICKET 6

90% chance of	10% chance of
$ **60**	$ **20**

Would you sell this lottery ticket for $ ____? Circle YES or NO

The least I would be willing to receive to sell this lottery is $ _____