The Robustness of Anchoring Effects on Market Good Valuation

SangSuk Yoon,[1] Nathan Fong,[2] and Angelika Dimoka[3]

Abstract

Ariely, Loewenstein, and Prelec (2003) showed that people's judgments of a product's value are strongly and systematically influenced by considering numbers (or "anchors") which should be irrelevant to their valuations. However, subsequent studies showed inconsistent results while using different experimental protocols. To bridge the gap between conflicting prior studies, we replicated the anchoring effect in product valuations while also varying the experimental protocol. We examined whether the type of anchoring number (Social Security number vs. random number), recruiting method (classroom vs. email recruiting), or cover story (in-class market research demonstration vs. participation in research experiment) contributed to the divergent results. Our results replicated the anchoring effect at economically significant magnitudes (near the middle of the range for previous studies), but we did not find evidence that the procedural differences can account for the conflicting prior results.

Keywords: anchoring effect, product valuation, preference elicitation

1. Introduction

The anchoring effect refers to the tendency for people's numeric judgments to be influenced by an initially considered value (Tversky & Kahneman, 1974). The anchoring effect has been studied in many domains, such as factual questions, physical length

[1] Temple University, Alter Hall, 1801 Liacouras Walk, Philadelphia, PA 19122, ssyoon@temple.edu
[2] Temple University, Alter Hall, 1801 Liacouras Walk, Philadelphia, PA 19122, nmfong@temple.edu
[3] Temple University, Alter Hall, 1801 Liacouras Walk, Philadelphia, PA 19122, dimoka@temple.edu

estimation, answers to math equations, legal judgment, performance judgment, and purchasing quantity (for a review, see Chapman & Johnson, 2002). It has been documented in a variety of product evaluation tasks; notably, Ariely, Loewenstein, and Prelec (2003) showed that, in consequential decision making situations, people's product valuations are strongly and systematically influenced by numbers that should have no bearing on their valuations. They showed several items to participants in a class and asked them whether they were willing to buy the given items for a dollar amount derived from the last two digits of their Social Security numbers (SSN, hereafter). The same participants were also asked their willingness-to-pay (WTP, hereafter) using an incentive compatible BDM mechanism (Becker, Degroot, & Marschak, 1964). The result showed that people's WTP for the given items was strongly and systematically influenced by the anchoring number, but their relative valuations across the given items were consistent. This finding contradicts the claim that consumers' choices reflect stable underlying preferences, and supports theories of constructed preferences (Slovic, 1995); thus, the robustness of such anchoring effects has important implications for economic theory.

Despite the importance of the result, evidence for the anchoring effect on market goods has only been documented in a handful of published studies. Furthermore, the results have been inconsistent. For example, Bergman, Ellingsen, Johannesson, and Svensson (2010) found a significant anchoring effect for five out of six items, and Simonson and Drolet (2004) showed similar results with different items. Fudenberg, Levine, and Maniadis (2012), however, did not find significant anchoring effects for any of six items (similarly, they did not find an effect in several willingness-to-accept studies). They suggest that anchoring effects are not robust in consequential valuation tasks. Fudenberg et al. (2012) do not provide a specific explanation for the difference, and leave open the possibility that it was due to differences in experimental protocol, arguing that if anchoring effects occur only with very specific procedures, they should not be considered particularly robust.

Our investigation has two objectives. The first is to reassess whether anchoring generates economically meaningful effects for market goods. To that end, our second objective is to reconcile the large discrepancies between prior studies attempting to measure this effect, as understanding the source of these differences would be informative about the stability of anchoring effects. Three probable sources of variation across studies are (1) procedural differences, (2) other unobserved differences, and (3) sampling variation. We address the first two possibilities by varying several procedural features that differed in previous studies. If, while varying experimental procedures and holding other unobserved factors constant, we can simultaneously replicate both the original Ariely et al. (2003) result, the Fudenberg et al. (2012) null result, and find a clear difference between the two, this would suggest that the procedural differences account for the different results. Depending on why the procedures cause a difference, such a finding could weaken claims about the prevalence of anchoring effects. However, we do not find evidence that procedural differences account for the discrepancies. Other unobservable differences, such as differences between the participant populations, other environmental differences, and subtle procedural differences that are impractical to fully document, are difficult to rule out. To the extent that the discrepancies were in part due to sampling variation, we contribute data from an additional 197 participants to a collective sample size of 249 participants from prior incentive compatible willingness-to-pay studies. We find large and robust anchoring effects near the middle of the range.

Compared with the Ariely et al. (2003) study, there were several procedural differences in Fudenberg et al. (2012) that could have moderated the anchoring effect. In particular, Fudenberg et al. (2012) used a different anchoring mechanism and recruiting method (in one study, they investigate whether the difference was due to whether participants were provided a thorough description of the BDM procedure, but find no evidence that this was the case). While Ariely et al. (2003) used the last two digits of participants' Social Security number as the anchoring number, Fudenberg et al. (2012) had participants actively generate random numbers (RN, hereafter). One possibility is that the perceived randomness of an anchoring number influences people's inferences about the anchoring number (Frederick, Mochon, & Danilowitz, 2013), and the actively-generated numbers may have been more transparently random to the Fudenberg et al. (2012) participants. A recent study showed that SSN can be predicted based on personal information, such as place and date of birth (Acquisti & Gross, 2009); even the perception that information is embedded in a person's SSN could cause inferences to be more prevalent than with randomly generated numbers.

Additionally, Ariely et al. (2003) conducted the experiment in a classroom as a part of a class activity, but Fudenberg et al. (2012) recruited people from a participant email list and conducted the experiment in sessions dedicated to running the experiment. The classroom demonstration context may have provided enough of a "cover story" that participants were less likely to perceive the anchoring number as transparently irrelevant to the valuation task. In the lab study, participants may have been more attuned to the arbitrariness of the anchoring manipulation, and thus less likely to draw inferences about the relevance of the anchor. This possibility was tested in two ways; in the first study, the experiment was conducted in both a classroom context and in a lab study using email recruiting. Surprisingly, we found a slightly stronger anchoring effect in the lab condition. To narrow down whether this was due to differences in expectations when it is salient to participants that the study is for academic research, a follow-up study conducted in a class, but the cover story was varied to emphasize that the activity was a demonstration of either marketing research or academic research.

2. Study 1: Recruiting Method and Anchoring Method

We used 2 X 2 between-subject design: (1) classroom or email recruiting, and (2) SSN or RN anchoring. We were not necessarily predicting an interaction effect; rather, we were exploring whether one or both factors together could account for differences in previous findings. A total of 116 undergraduate students at a major university in the northeast US participated in our study, receiving class credit or $5 in exchange for participation. Participants in classroom conditions were undergraduate students taking Marketing Research classes, and participants in email recruiting conditions were from a Sona system developed to recruit undergraduates for research. In the SSN conditions, the anchoring number was decided by transforming the last two digits of the participant's SSN into a dollar amount, and in the RN conditions, random numbers were generated by drawing from a deck of 100 numbered cards with a range of 0 to 99.

We conducted the experiment using an incentive compatible BDM mechanism (Becker et al., 1964). Beforehand, participants were informed that they would have a chance to buy one of the six items based on their answers. We showed the six items and, for each item, asked participants for three responses in the following order: whether they want to buy the given items for the random dollar amount (Y/N response), the highest price they are willing to pay, and their estimate of the item's market price. After responding to all the questions, the winners (who were endowed with enough additional funds to cover the purchase of the items), item, and resolution method (e.g., based on the Y/N or WTP response) were randomly decided by rolling dice (see Appendix for examples of the instructions and questionnaires).

To investigate the effects of the different anchoring mechanisms and recruiting methods, we regressed WTP responses on the anchoring values, interacted with the experimental conditions (Table 1). The main effect of the anchoring number was statistically significant in all specifications (β = 0.19, $t(685) = 6.67$, $p < .01$). The main effects of random number and recruiting method were not significant, nor was there an interaction between anchor type and anchoring number (Table 1). Not only do the results fail to explain the difference between previous studies, but the difference between recruiting conditions appears to run in the opposite direction: the anchoring effect appears stronger in the email recruiting conditions than that in the classroom conditions. This relationship is stronger when comparing high and low anchors, avoiding linearity assumptions in the Table 1 regressions (see Table 2). One possible explanation is that participants recruited specifically for the research study were more likely to consider the anchoring number as informative for the valuation task. Frederick et al. (2013) did not investigate the effect of recruiting method or experimental environment, but they do argue that participants' inferences about the relevance of random number ranges provided in lab experiments are well-founded.

[Table 1, here]

[Table 2, here]

3. Study 2: Cover Story

In this study, we provided participants with varying cover stories to investigate how their expectations affect the malleability of their preferences. We used a between-subject design, deploying two different cover stories for two different sections of the same

Marketing Research class. In the “Class Exercise” condition, students were told in the preceding class that we would perform a market research demonstration. During the instructions for the study, it was explained how the techniques being demonstrated could be used to measure consumer demand at various prices. In the “Experiment” condition, students were told in advance that we would conduct academic research at the following class. During the instructions for the study, it was explained that such studies were used to refine theories about consumer behavior that influence the content of their classes.

A total of 81 undergraduate students participated in our study, in exchange for a chance to purchase the items used in the experiment, using the BDM procedure (selected participants were endowed with enough funds to cover the purchase of the items). We regressed WTP responses on the anchoring values, interacted with the cover story condition. The main effect of the anchoring number was statistically significant in all specifications, but an interaction between the anchoring number and the cover story was not detected (Table 3).

[Table 3, here]

4. Combined Results

Both studies resulted in statistically robust anchoring effects. To estimate more precise effect sizes for the population used in these studies, we pooled the data. We examined the ratio of WTP in the highest quintile of anchoring number to that in the lowest quintile and compared the magnitudes with prior studies (Table 4). The range of WTP ratio in our study was from 1.27 to 2.09, which is near the middle of prior studies (the ratio in Ariely et al. (2003)’s study was 2.16 to 3.03 and that in Fudenberg et al. (2012)’s study was 0.77 to 0.92). We also tested for the presence of anchoring effects for each individual item using Pearson correlation coefficients, as used in previous studies (Ariely et al., 2003; Bergman et al., 2010;

Fudenberg et al., 2012). The main results are presented in Table 3. As seen in the table, the Pearson correlation coefficients between anchoring number and WTP ranged from .12 to .38. The range of the Pearson correlation coefficients was .32 to .52 in Ariely et al. (2003), and -.11 to .21 in Fudenberg et al. (2012). Our results replicated the anchoring effect with correlations near the middle of the range for previous studies.

[Table 4, here]

To allow visual inspection of the distribution of the effect, we used the WTP responses to generate demand curves for each of the items. To draw the demand curves, we inverted the empirical cumulative density functions of the participants' reported WTP (Figure 1). The demand curve shifts out for participants with high anchors, all along the range of reported prices. These figures suggest that the effect was pervasive, and not concentrated among a few participants with very high or low WTP.

[Figure 1, here]

We also note that the demand curve shifts out with increasing anchoring numbers (Figure 2). Consistent with a broad-based anchoring effect, the difference between the two lower quintiles is greater at lower price levels, and the curves for the three higher quintiles shift out more at higher prices. There does appear to be the possibility of a ceiling effect, since the differences between the three higher quintiles are relatively small, particularly for the fourth and fifth quintiles. This could result when the anchoring values move too far outside the plausible range for participants' WTP (Chapman & Johnson, 1994). This also suggests that an analogous floor effect could suppress anchoring effects in some cases. An example in our data is the Belgian chocolates, for which the undergraduate participants were, on average, only willing to pay $12.62. In Fudenberg et al. (2012), only one of the items used had an average willingness to pay over $20 (the cordless keyboard, at $31.50; the next highest was the cordless mouse, at $15.97). While they cannot fully explain the difference with Ariely et al. (2003), which also included items with low WTP, it is possible that floor effects contribute to the null findings in some cases.

[Figure 2, here]

5. Conclusion

In this study, we replicated the anchoring effect in product valuation, and investigated possible causes of divergent prior results. While our results more closely resembled Ariely et al. (2003)'s original results than Fudenberg et al. (2012)'s null results, the effects we observed appear to be lower in magnitude than the original study. However, the anchoring manipulation had an economically significant impact on consumer valuations for market goods. The interactions between the anchoring number and the three experimental factors could not explain the difference between prior results. We conclude that further investigation is needed to understand how environmental factors influence the perceived randomness of anchoring numbers, and its subsequent effect on WTP.

References

Acquisti, A., & Gross, R. (2009). Predicting Social Security numbers from public data. *Proceedings of the National Academy of Sciences, 106*(27), 10975-10980.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics, 118*(1), 73-105.

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*(3), 226-232.

Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters, 107*(1), 66-68.

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 120-138). New York, NY: Cambridge University Press.

Frederick, S., Mochon, D., & Danilowitz, J. (2013). *The role of inference in anchoring effects*. Unpublished manuscript.

Fudenberg, D., Levine, D. K., & Maniadis, Z. (2012). On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics, 4*(2), 131-145.

Simonson, I., & Drolet, A. (2004). Anchoring effects on consumers’ Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research, 31*(3), 681-690.

Slovic, P. (1995). The construction of preference. *American Psychologist, 50*(5), 364-371.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Table 1. Regression Analysis

	(1) No Item FE	(2) Item FE	(3) Anchor Type	(4) Recruit Type	(5) Both
Anchoring Number	0.19**	0.19**	0.19**	0.19**	0.19**
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Anchoring Number*Anchor Type			0.02		0.02
			(0.03)		(0.03)
Anchor Type			0.21		0.17
			(1.48)		(1.49)
Anchoring Number*Recruit Type				0.04†	0.04†
				(0.02)	(0.03)
Recruit Type				-0.20	-0.28
				(1.43)	(1.49)
Anchoring Number*Anchor Type*Recruit Type					0.01
					(0.03)
Anchor Type*Recruit Type					0.16
					(1.49)
Constant	19.34**	Item FE	Item FE	Item FE	Item FE
	(1.66)				
Observations	694	694	694	694	694

Standard errors in parentheses
† $p < .10$, * $p < .05$, ** $p < .01$

Note: Models (2) through (5) include fixed effects for each item. Anchor Type is coded as –1 for SSN and 1 for RN. Recruit Type is coded as –1 for classroom and 1 for email recruiting. Standard errors clustered on participant do not affect the pattern of results.

Table 2. Regression by High/Low Group

	(1) No Item FE	(2) Item FE	(3) Anchor Type	(4) Recruit Type	(5) Both
High/Low Anchor	9.86**	9.94**	9.75**	10.19**	9.97**
	(1.64)	(1.41)	(1.44)	(1.42)	(1.45)
High/Low Anchor*Anchor Type			0.90		1.18
			(1.44)		(1.45)
Anchor Type			-0.22		-0.13
			(1.03)		(1.05)
High/Low Anchor*Recruit Type				3.25*	3.29*
				(1.42)	(1.45)
Recruit Type				0.12	0.09
				(1.04)	(1.05)
High/Low Anchor*Anchor Type*Recruit Type					-0.03
					(1.45)
Anchor Type*Recruit Type					0.44
					(1.05)
Constant	23.80**	Item FE	Item FE	Item FE	Item FE
	(1.19)				
Observations	694	694	694	694	694

Standard errors in parentheses

* p<0.05, ** p<0.01

Table 3. Regression Analysis for Study 2

	(1) No Item FE	(2) Item FE	(3) Story
Anchoring Number	0.24***	0.24***	0.24***
	(0.05)	(0.05)	(0.05)
Anchoring Number*Cover Story			0
			(0.05)
Cover Story			-0.05
			(2.42)
Constant	27.19***	Item FE	Item FE
	(2.67)		
Observations	486	486	486

Standard errors in parentheses
† $p < .10$, * $p < .05$, ** $p < .01$

Note: Models (2) and (3) include fixed effects for each item. Cover Story is coded as –1 for Class Exercise and 1 for Experiment. Standard errors clustered on participant do not affect the pattern of results.

Table 4. Combined Results, Average Stated WTP Sorted by Quintile

Quintile	Art Book	Bluetooth Keyboard	Chocolate[a]	Cordless Mouse	Fancy Dinner	Normal Dinner	Headphone[a]
1 (N=28) 0-19	12.11	31.55	9.62	18.07	30.36	23.86	50.42
2 (N=30) 20-39	14.90	39.90	13.50	22.77	45.67	29.87	55.88
3 (N=34) 40-62	21.82	43.68	14.50	23.32	50.59	36.85	40.56
4 (N=35) 63-74	26.91	50.14	14.64	24.94	55.20	39.00	88.46
5 (N=36) 76-99	25.33	51.94	12.19	29.22	51.14	38.67	70.90
Average	20.42	42.78	12.62	23.85	45.70	32.53	56.22
Correlations	0.27	0.30	0.12	0.25	0.38	0.38	0.12
p-value	.0001	< .0001	0.1914	0.0003	< .0001	< .0001	.2828
Ratio	2.09	1.65	1.27	1.62	1.68	1.62	1.41
Rank Sum	-3.93	-3.64	-1.50	-3.43	-4.79	-4.09	-1.23
p-value	.0003	.0003	.1326	.0006	< .0001	< .0001	.2202
ALP Ratio	2.16	3.46	2.16	3.46			
BEJS Ratio	2.68		2.11				
FLM Ratio	0.81	0.92	0.87	0.77			

[a] The number of responses for Chocolate and Headphone were different with the number of responses for other items because we substituted Chocolate to Headphone in study 2. The number of responses per quintile for Chocolate were 16, 14, 16, 22, and 26 (from first to fifth quintile, respectively), and the number of responses per each quintile for Headphone were 12, 16, 18, 13, and 10 (from first to fifth quintile, respectively).

Note: Fudenberg et al. (2012) slightly changed the items used because of legal constraints (wine cannot be sold to people below 21 years old in the USA) and participant preferences (trackballs are not popular among current undergraduates). We substituted fancy and normal dinners for the fancy and normal wines, and used cordless mice instead of trackballs.

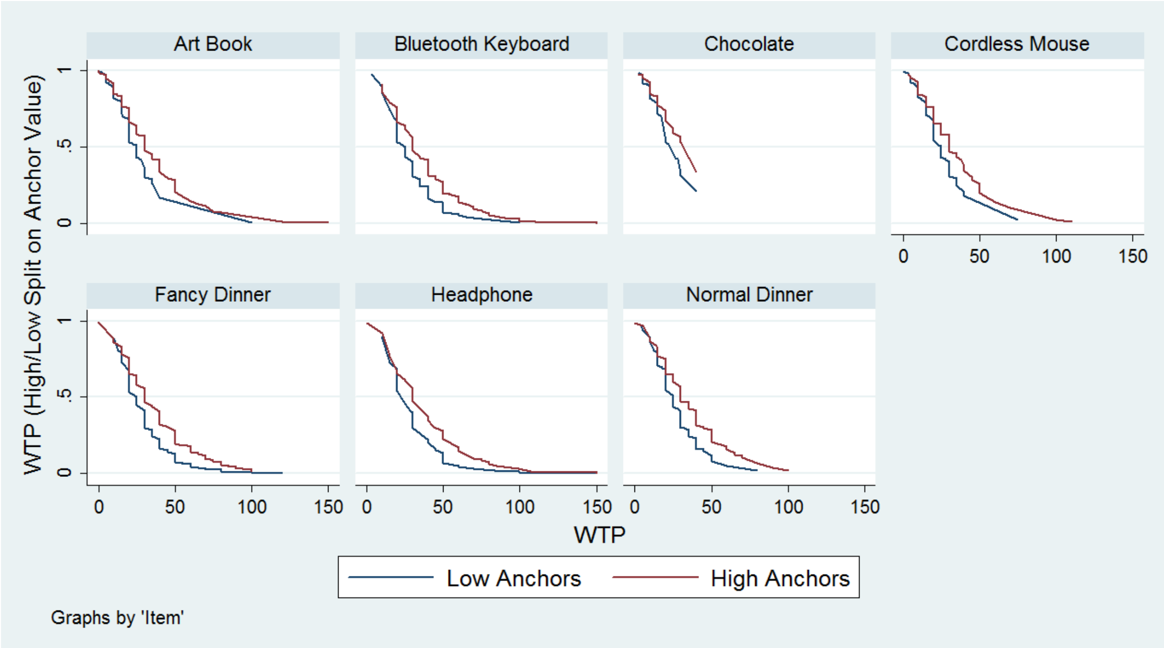

Figure 1. Demand Curves for Low vs. High Anchor Values

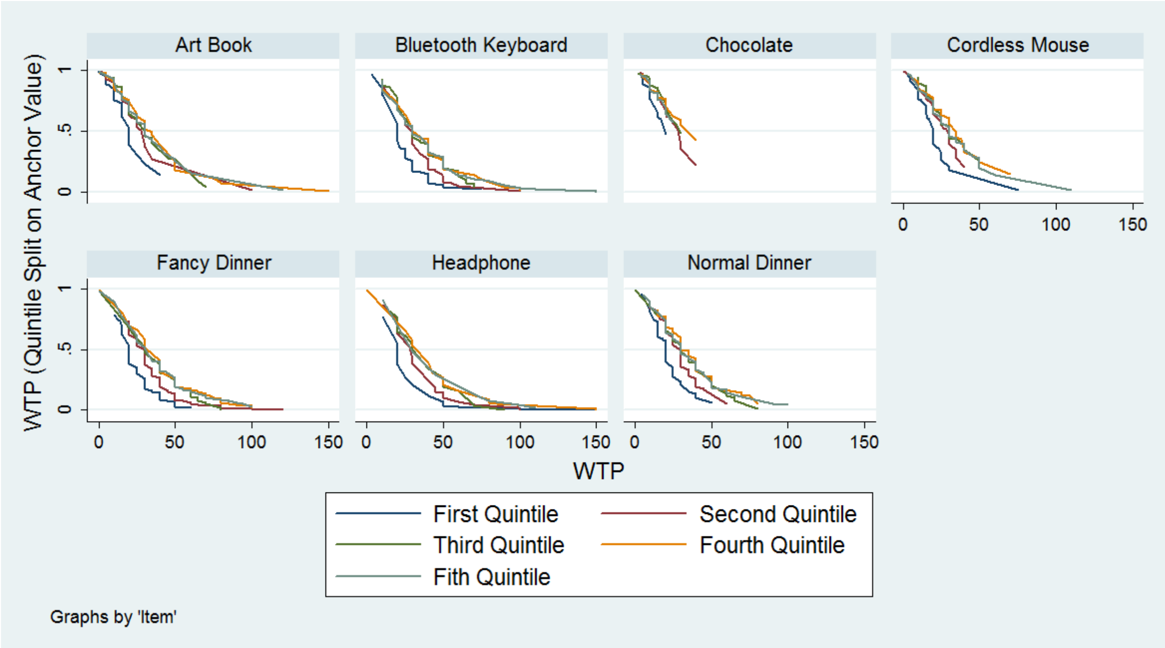

Figure 2. Demand Curves for Quintiles of Anchor Values

Appendix A. Instructions and answer sheet used in our study (classroom-SSN condition)

Thank you for participating in this pricing exercise. We are interested in how much you are willing to pay for various things. We will show you a couple of items and then ask you a pair of price-purchase intent questions about each item:

- First, we will ask whether you would like to buy the item at a particular price. That price will be determined randomly by having you convert the last two digits of your Social Security Number into a whole-dollar price.

- Next, we will ask you to state the most that you would be willing to pay for the item. We will pick three items randomly. Each item will be offered for purchase to one person, chosen randomly from the class. One item will be sold on the basis of answers to the first question, and two items on the basis of answers to the second question. Each person is in the running for only one item. We will endow selected three students enough money to buy the item.

<u>If you are chosen and the first question counts</u>, then we will look to see whether you decided to buy, or not to buy, the item at your random (SS#) price. If you stated that you wanted to buy the item at that price, then we will sell it to you at that price. If you stated that you did not want to buy the item at that price, then no sale will take place.

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

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

<u>Because your answer does not affect the amount you pay, only whether you buy, it's to your advantage to state the maximum that you would be willing to pay.</u>

First, please copy the last two digits of your SS# into each of the 6 boxes below

ITEM 1 Description ______Cordless Mouse_______________________

Would you buy this item for $ [ ] ? Circle: YES or NO

The most I would be willing to pay for this item is $ _______

My best guess of the actual market price for this item is $ ________

ITEM 2 Description ______Bluetooth Keyboard______________________

Would you buy this item for $ [ ] ? Circle: YES or NO

The most I would be willing to pay for this item is $ _______

My best guess of the actual market price for this item is $ ________

ITEM 3 Description ________Dinner Coupons at Sabrina's____________________

Would you buy this item for $ [ ] ? Circle: YES or NO

The most I would be willing to pay for this item is $ _______

My best guess of the actual market price for this item is $ ________

ITEM 4 Description ___Dinner coupons at Graces Trading Company____________________

Would you buy this item for $ [ ] ? Circle: YES or NO

The most I would be willing to pay for this item is $ _______

My best guess of the actual market price for this item is $ ________

ITEM 5 Description _______Art Book____________________________

Would you buy this item for $ [ ] ? Circle: YES or NO

The most I would be willing to pay for this item is $ _______

My best guess of the actual market price for this item is $ ________

ITEM 6 Description _______Chocolate__________________________

Would you buy this item for $ [ ] ? Circle: YES or NO

The most I would be willing to pay for this item is $ _______

My best guess of the actual market price for this item is $ ________

Appendix B. The random price matrix used in our study

	00	10	20	30	40	50	60	70	80	90
0	$36.00	$0.00	$30.00	$26.00	$38.00	$14.00	$32.00	$1.00	$5.00	$32.00
1	$40.00	$8.00	$42.00	$35.00	$36.00	$66.00	$45.00	$0.00	$19.00	$32.00
2	$24.00	$17.00	$38.00	$30.00	$10.00	$9.00	$48.00	$53.00	$37.00	$73.00
3	$24.00	$19.00	$3.00	$25.00	$33.00	$27.00	$7.00	$29.00	$9.00	$39.00
4	$16.00	$13.00	$51.00	$74.00	$5.00	$64.00	$40.00	$50.00	$70.00	$24.00
5	$28.00	$27.00	$43.00	$12.00	$65.00	$59.00	$17.00	$30.00	$24.00	$7.00
6	$5.00	$14.00	$7.00	$4.00	$51.00	$13.00	$1.00	$27.00	$35.00	$61.00
7	$2.00	$29.00	$12.00	$13.00	$30.00	$15.00	$6.00	$19.00	$35.00	$27.00
8	$50.00	$15.00	$49.00	$36.00	$26.00	$28.00	$53.00	$33.00	$59.00	$80.00
9	$46.00	$9.00	$22.00	$12.00	$82.00	$51.00	$5.00	$63.00	$10.00	$29.00