

## **How Multiple Anchors Affect Judgment: Evidence from the Lab and eBay**

YAN ZHANG, YE LI, AND TING ZHU \*

Yan Zhang is Assistant Professor of Marketing at the National University of Singapore, NUS Business School, 15 Kent Ridge Drive, Singapore 119245. Tel: (65) 6516-7724. Email: [yan.zhang@nus.edu.sg](mailto:yan.zhang@nus.edu.sg). Ye Li is a postdoctoral research scholar at Columbia University, Graduate School of Business, 3022 Broadway Avenue, Uris Hall, New York, NY 10027. Tel: 212-854-3698. Email: [yl2629@columbia.edu](mailto:yl2629@columbia.edu). Ting Zhu is Assistant Professor of Marketing at University of Chicago, Booth School of Business, 5807 S. Woodlawn Ave, Chicago, IL 60637. Tel: 773-834-8057. Email: [ting.zhu@chicagobooth.edu](mailto:ting.zhu@chicagobooth.edu). \* The three authors contributed equally. The order of the authorship was determined by die roll.

The authors thank Daniel Bartels, Pradeep Chintagunta, Benjamin Converse, Nicholas Epley, Cindy Kim, Margaret Lee, Yesim Orhun, Dilip Soman, Oleg Urminsky, George Wu, Robert Zeithammer, and Meng Zhu for valuable comments; Wei Ding for programming assistance; and the University of Chicago Booth School of Business and the National University of Singapore for research support.

## ABSTRACT

This article examines the impact of multiple anchors on subsequent estimates. Although past research has generally found that judgments assimilate toward single anchors, we hypothesize that the presence of additional anchors can reverse this effect. When presented with both an extreme anchor A and a moderate anchor B, people rely more on anchor B than when the anchor A is instead moderate. Extreme anchors in the two-anchor case can therefore generate more moderate estimates than less extreme anchors do, which implies a contrast effect. Three controlled experiments verified that although estimates assimilated to single anchors, the reverse occurred when people were simultaneously given a second anchor: extremely low (high) anchors generated higher (lower) estimates than moderately low (high) anchors. We found corroborating evidence in archival data from eBay auctions in the U.S. and China. This research has implications for pricing strategies when there is more than one price cue available.

**KEYWORDS:** anchoring; multiple anchors; anchor plausibility; product valuation; buy-it-now; pricing; eBay; auctions

Whether they are externally-provided or self-generated, anchors systematically influence people's estimates of uncertain quantities (Epley and Gilovich 2001; Mussweiler and English 2005; Tversky and Kahneman 1974). For example, Tversky and Kahneman (1974) asked people to guess whether a number generated by spinning a "wheel of fortune" was more or less than the percentage of African countries in the United Nations. Participants who spun lower numbers made significantly lower estimates than participants who spun higher numbers. The anchoring effect has been shown to be robust across many domains, including consumer decisions such as product valuation (Ariely, Loewenstein, and Prelec 2003; Green, Jacowitz, Kahneman, and McFadden 1998; Nunes and Boatwright 2004), purchase prices (Simonson and Drolet 2004), and purchase quantities (Wansink, Kent, and Hoch 1998). For example, Ariely, Loewenstein, and Prelec (2003) asked people to indicate whether they were willing to buy a list of products at a price equal to the last two digits of their social security number. They found that people with higher anchors (last two digits) were willing to pay more for the same products than those with lower anchors.

Despite the large number of articles examining anchoring effects, nearly all of it has focused on the effect of a single anchor on judgments. Our limited knowledge of how people make estimates when they encounter more than one anchor is incongruent with how commonly people encounter multiple anchors in daily life. For example, in order to forecast sales of a new product, managers often consult multiple sources of information such as the sales history of previous models and other similar products (Roggeveen and Johar 2004). Consumers also encounter multiple anchors when making purchases: potential buyers may estimate the value of a

house by referring to its list price, its private appraisal value, its last transaction price, and the average price of other houses in the neighborhood. Similarly, bidders in eBay auctions may form their willingness to pay by consulting both the starting price and the market price of similar products (Kamins, Drèze, and Folkes 2004).

The prevalence of judgments involving multiple anchors raises an important question: When people encounter more than one anchor, how do they arrive at a final estimate, and how does this process differ from that in the single-anchor case? Although previous research has tested the effect of presentation order on multiple anchors (Nunes and Boatwright 2004), anchoring as a result of multiple, sequential judgments (Mochon and Frederick 2009), and the effect one anchor in the presence of other anchors in a negotiation (Whyte and Sebenius 1997), no existing research has explicitly compared the effect of a single anchor with that of multiple anchors. In the present research, we investigate how an anchor's impact on subsequent estimates can change with the presence of additional anchors.

### **SINGLE VERSUS MULTIPLE ANCHORS**

Anchoring generally refers to the assimilation of people's estimates to a previously provided anchor, such that higher anchors generate higher or equal estimates than lower anchors. For example, when judging the length of a whale, participants provided with an anchor of 21 meters estimated the length of a whale to be 29.1 meters on average, whereas participants

provided with an anchor of 49 meters estimated the length of a whale to be 60.1 meters on average (Strack and Mussweiler 1997).

-----  
Insert Figure 1 about here  
-----

A natural question one might ask is whether the anchoring effect is proportional to anchor magnitude. Figure 1 shows three possible anchor response functions between anchor magnitudes and subsequent estimates (Chapman and Johnson 1994). Rather than finding a linear relationship (solid line in figure 1), research has generally found that the impact of anchors diminishes (dashed line in figure 1) as they become more extreme (Chapman and Johnson 1994; Strack and Mussweiler 1997). For example, for the length of a whale question, participants provided with an anchor of 0.2 meters estimated the length of a whale to be 20.5 meters on average. This extremely low anchor generated estimates that were only somewhat shorter than those generated by the moderately low 21 meters anchor. The effect of the extreme 0.2 meters anchor was diminished relative to the effect of the moderate 21 meters anchor, producing estimates lower than or equal to those produced by the moderate anchor.

Another possible relationship between anchor magnitudes and estimates, first suggested by Quattrone and colleagues (1981), is a non-monotonic one where extreme anchors could instead produce a contrast effect (dotted line in figure 1). That is, extremely low (high) anchors could actually generate *higher* (lower) estimates than moderately low (high) anchors. They suggested that this contrast effect could occur if extreme anchors caused people to abandon the anchoring process altogether and instead generate an unanchored value. However, empirical

demonstrations of such a contrast effect have been difficult to produce (although cf. Wegener et al 2001). For example, despite the fact that the 0.2 meters anchor was designed and pretested to be implausible for the length of a whale question, participants' estimates nonetheless assimilated toward the anchor, generating shorter estimates than for the 21 meters anchor.

Although researchers have generally found assimilation effects in the single-anchor case regardless of the plausibility of the anchor, we hypothesize that a contrast effect can emerge in a multiple-anchor situation when an extreme anchor is paired with a more moderate anchor. This hypothesis relies on anchor plausibility, which has been shown to affect people's absolute estimates (Chapman and Johnson 1994; Urbany, Bearden, and Weilbaker 1988). Research on cue diagnosticity also suggests that people evaluate how diagnostic a cue is when forming answers based on the cue (Skowronski and Carlston 1987). When people encounter more than one anchor, they will naturally evaluate which one makes more sense. This comparison process usually generates an evaluation of how similar the comparison targets are to each other in terms of plausibility (Mussweiler 2003). Two anchors will seem similar if they are both plausible; but if one anchor is plausible while the other is implausible, the anchors will seem dissimilar.

We suggest that dissimilarity between anchors causes people to rely on the anchor that seems more plausible. With two moderate anchors, both will seem similarly plausible, and therefore people will not have a reason to reject either anchor. However, when the first anchor is replaced with an extreme value and the second anchor remains moderate, people will rely more heavily on the moderate anchor. This decreased reliance on the first anchor when it is extreme implies that subsequent estimates will be closer to the second anchor than when the first anchor

is moderate, resulting in a contrast effect as shown in figure 1. Specifically, we expect that in the single-anchor case, estimates will assimilate toward the anchor regardless of whether the anchor is extreme or moderate, as found in previous research (Chapman and Johnson 1994; Strack and Mussweiler 1997); however, we also predict that estimates for an extremely low (high) anchor will be higher (lower) than estimates for a moderately low (high) anchor when each anchor is paired with a second, moderate anchor. This contrast effect is the opposite of what happens in the single-anchor case and can be considered a judgment reversal.

To illustrate our predictions more concretely, let us return to the length of a whale example. Recall that participants provided with single anchors of 0.2 meters and 21 meters (pretested to be implausible and plausible, respectively) estimated lengths of 20.5 meters and 29.1 meters on average, respectively. Now imagine that we paired each of these anchors with another moderately high anchor of 49 meters, which has been pretested to be plausible. We predict that participants simultaneously provided with anchors of 0.2 meters and 49 meters would judge the 0.2 meters anchor to be significantly less plausible than the 49 meters anchor and therefore rely mostly on the more plausible 49 meters anchor. However, participants simultaneously provided with anchors of 21 meters and 49 meters may find that both anchors seem plausible and therefore rely on both anchors. As a result, the 21 meters anchor will have a significant impact on estimates. Therefore, although the extremely low 0.2 meters anchor by itself generates lower estimates than the moderately low 21 meters anchor does, we expect that when presented along with the plausible 49 meters anchor, the extremely low anchor will generate *higher* estimates than the more moderate but plausible low anchor. That is, we expect a

contrast effect for the two-anchor case in the sense that estimates generated by extreme anchors will be more moderate (in this case, higher) than estimates generated by moderate anchors.

The process that accounts for this contrast effect is akin to the prediction by Quattrone and colleagues (1981) in that the contrast effect occurs because people abandon the extreme anchor and instead rely on the other more plausible anchor. Since people need other information to rely on for them to abandon a provided anchor, the contrast effect is difficult to find in single-anchor cases. The contrast effect becomes possible in two-anchor cases because the second anchor can serve as that additional information source that is required.

In our experiments, we vary the value of one anchor (A) between extreme and moderate values and add a moderate second anchor (B) for the two-anchor conditions. Although we expect moderate anchors to seem plausible to participants, we will refer to them as moderate anchors unless explicitly discussing plausibility. We predict that when anchor A is extreme, people will rely mostly on the more moderate anchor B. But when anchor A is also moderate, people will rely on both anchor A and B. This differential reliance on anchor A in the two-anchor conditions generates a contrast effect. That is, we predict that in single-anchor conditions, an extremely low anchor A will generate estimates lower than or equal to those generated by a moderately low anchor A; however, when each anchor A is paired with anchor B, the same extremely low anchor A will generate *higher* estimates than the moderately low anchor A, a contrast effect relative to anchor A magnitude. Using the same logic, we predict that the same occurs for high anchors but in the opposite direction: an extremely high single anchor will generate higher estimates than a

moderately high single anchor, but when each is paired with anchor B, the same extremely high anchor A will generate *lower* estimates than the moderately high anchor A.

## **OVERVIEW OF STUDIES**

We conduct four studies to examine the effect of two anchors on judgments. Study 1 compares the effect of extreme and moderate anchors on product evaluation with and without a second, moderate anchor. Study 2 generalizes our findings from evaluations of consumer products to general knowledge questions. Study 3 tests the prediction that estimates can be pushed even further from the extreme anchor and closer to the moderate anchor when the perceived difference between the two anchors is increased. Finally, study 4 uses archival data from eBay auctions to provide convergent evidence and external validity. Taken together, our research contributes to the existing anchoring research not only by demonstrating the differences between single anchor and double anchors, but also by providing a cognitive mechanism for understanding when anchoring generates contrast effect.

## **STUDY 1: PRODUCT VALUATION**

We predict that an extremely high anchor will generate estimates higher than or equal to the estimates generated by a moderately high anchor when each anchor appears alone, whereas the same extremely high anchor will generate *lower* estimates than a moderately high anchor

when each anchor is paired with a second moderate anchor. Study 1 tested this prediction by using consumer product valuation questions. High anchors are particularly relevant in this domain since sellers may use higher regular or suggested prices to manipulate the perceived value of a purchase.

## Methods

In this and all subsequent controlled experiments, participants volunteered to participate in the experiment by clicking a link on a survey website. They were roughly evenly split on gender (51.5% male), with a median age of 48 years ( $M = 46.4$ ,  $SD = 16.3$ ), and a relatively high level of education (81.2% reported at least some college education).

*Stimuli selection.* We first selected anchor values from a pretest. In the pretest, a total of 135 participants estimated the Amazon.com prices for six consumer products. For each product, participants saw a picture of the product, a paragraph of general description, and bullet points describing key features. Based on the pretest results, we selected the 95<sup>th</sup> percentile estimate as an extremely high anchor A, the 85<sup>th</sup> percentile as a moderately high anchor A, and the 50<sup>th</sup> percentile as anchor B, all rounded to the nearest \$10 to avoid potential variation due to perceptions of informational content based on anchor precision (Janiszewski and Uy 2008; Thomas, Simon and Kadiyali 2010). Table 1 lists the anchor values for the six products.

-----  
Insert TABLE 1 about here  
-----

*Main study.* A total of 152 participants were randomly assigned to one of the 4 conditions in a 2 (number of anchors: one vs. two)  $\times$  2 (extremity of anchor A: moderate vs. extreme)  $\times$  6 (product) mixed design with number of anchors and extremity of anchor A as between-participants variables and product as within-participants variable. All participants read the same product information as in the pretest. Participants in the one-anchor conditions judged whether each product's price on Amazon.com was less or more than anchor A (e.g., "Do you think the price of this clock on Amazon is less than \$420 or more than \$420?"). In the two-anchor conditions, participants instead judged whether the price was less than anchor B, between anchor B and anchor A, or more than anchor A (e.g., "Do you think the price of this clock on Amazon is less than \$90, between \$90 and \$420, or more than \$420?"). After the anchoring judgment, all participants estimated the actual price of the product on Amazon.com (e.g., "How much do you think this clock sells for on Amazon.com?").

After participants completed this procedure for all six products in random order, they then rated the plausibility of the anchor(s) for each product ("How close or far do you think \$420 is to the actual price of this clock on Amazon.com?") on a continuous scale labeled from 0 (very far) to 10 (very close).

## **Results and Discussion**

*Estimates.* The distribution of estimates was positively skewed so we performed the analyses by transforming participants' log estimates into z-scores for each product and then

pooling across products. After removing outliers that were more than three standard deviations from the mean, we were left with 876 responses across the six products. Figure 2 shows the z-scores of the log estimates averaged across questions.

We predicted a contrast effect for the two-anchor conditions but not the one-anchor conditions. Specifically, we expected that extremely high anchors would generate estimates that are higher than or equal to the estimates generated by moderate anchors in the one-anchor condition, but that the same extremely high anchors would generate lower estimates than moderate anchors in the two-anchor condition. To test this prediction, we submitted the standardized, logged estimates to a 2 (number of anchors: one vs. two)  $\times$  2 (extremity of anchor A: extreme vs. moderate)  $\times$  6 (product, repeated) repeated-measures ANOVA. The analysis returned a main effect of number of anchors,  $F(1,704) = 18.46, p < .0001$ , and the expected interaction between number of anchors and anchor A extremity,  $F(1,704) = 10.56, p < .01$ . As shown in figure 2, when there was only one anchor, the extreme 95<sup>th</sup> percentile anchor A generated higher estimates than the more moderate 85<sup>th</sup> percentile anchor A ( $M_{95\text{th}} = .37$  vs.  $M_{85\text{th}} = -.02$ ),  $F(1,704) = 22.73, p < .0001$ . However, after adding the 50<sup>th</sup> percentile anchor B, the 95<sup>th</sup> percentile anchor generated lower estimates than the 85<sup>th</sup> percentile anchor ( $M_{95\text{th}} = -.34$  vs.  $M_{85\text{th}} = -.12$ ),  $F(1,704) = 3.90, p < .05$ . These results were consistent with the predicted contrast effect: when an extremely high anchor was accompanied by a moderate anchor, final estimates were even lower than when a moderately high anchor was accompanied by the same plausible anchor.

-----  
 Insert Figure 2 about here  
 -----

*Plausibility Ratings.* Two types of participants were not included in this analysis: participants who estimated the Amazon price to be lower than anchor B but rated anchor A (which was higher than anchor B) to be more plausible; and participants who estimated the Amazon price to be higher than anchor A but rated anchor B to be more plausible. These participants either misused the plausibility scale or misunderstood the meaning of plausibility.

As expected, the extreme anchor A was judged as less plausible ( $M_{95\text{th}} = 4.14$ ) than the moderate anchor A ( $M_{85\text{th}} = 5.35$ ) in the single-anchor condition,  $F(1, 436) = 13.54, p < .001$ . More importantly, we found that the difference in plausibility between anchor A and anchor B was greater when anchor A was extreme than when it was moderate,  $F(1, 203) = 14.18, p < .001$ , as predicted. That is, the extremely high anchor A was judged as much less plausible than anchor B ( $M_A = 3.05$  vs.  $M_B = 6.79$ ), whereas the moderately high anchor A was judged to have a more similar plausibility as anchor B ( $M_A = 4.38$  vs.  $M_B = 6.05$ ). These plausibility results are consistent with our prediction that adding a more plausible anchor B to an extreme anchor A allows people to dismiss the extreme anchor and mostly rely on anchor B.

## **STUDY 2: GENERAL KNOWLEDGE QUESTIONS**

Study 2 generalizes our results from consumer product valuations to general knowledge questions. In addition, unlike study 1 which tested the effects of pairing high anchors with a second moderate anchor, study 2 examined the effects of pairing low anchor values with a second moderate anchor. In addition, our theory suggests that a contrast effect should only occur

when the perceived difference in plausibility between anchor A and B is sufficiently large. For larger plausibility differences, we expect estimates to negatively correlate with anchor magnitude, but for smaller plausibility differences, we expect the usual positive relationship between estimates and anchor magnitude. Study 2 thus explored this nonmonotonic relationship between estimates and anchor magnitude in the two-anchor condition.

### *Methods*

*Stimuli selection.* We selected our stimuli using a multiple-step procedure. First, we pretested 20 general knowledge questions and selected four as target questions. In the pretest, 250 participants made simple estimates in response to 10 questions randomly selected from a pool of 20 (e.g., “What do you think the height of the Sears Tower is (in feet)?”). We then selected four questions that generated answers with relatively large variances as target questions.

-----  
 Insert TABLE 2 about here  
 -----

Next, based on the percentiles of estimate distributions in the pretest, we chose the 5th, 20th, 30th, 40th, 50th and 90th percentiles as the anchor values for the main study (see table 2). We set the extremely low anchor A at the 5th percentile and four more moderate levels of anchor A at the 20th, 30th, 40th, and 50th percentiles. Since the pretest estimates were highly skewed such that the 90th percentiles were actually fairly close to the true values, anchor B was set at the 90th percentile. We rounded all anchors to the nearest multiple of 10.

Finally, we tested the plausibilities of these anchors. Another 72 participants from the same website were randomly assigned to one of the six anchor levels and rated anchor

plausibility for each of the four selected questions on a 0 (very implausible) to 10 (very plausible) scale. For example, participants judged “How plausible or implausible is it for the Sears Tower to be 560 feet?”

We analyzed plausibility ratings using a 6 (anchor percentile)  $\times$  4 (question) ANOVA with question as a within-participant variable. This analysis revealed a main effect of anchor percentile,  $F(5, 317) = 368.83, p < .0001$ . Relative to the 90th percentile anchor B ( $M_{90th} = 5.58$ ), the 5th percentile anchor was rated as less plausible ( $M_{5th} = 2.40$ ),  $F(1, 317) = 27.32, p < .0001$ , the 20th percentile anchor was rated equally plausible ( $M_{20th} = 5.15$ ),  $F(1, 317) = .01, ns$ , and the 30th, 40th anchors, and 50th percentile anchors were rated as more plausible ( $M_{30th} = 6.44, M_{40th} = 6.16, M_{50th} = 7.12$ ),  $F(1, 317) = 9.56, 16.15, \text{ and } 32.31$ , respectively, all  $p$ 's  $< .01$ . As expected, the 5th percentile anchor was rated as least plausible among all anchors. Note that because the distribution of estimates in the pretest was positively skewed, even the 90th percentile anchor was rated as reasonably plausible.

*Main study.* Another 820 participants participated in the main study. This experiment used a 2 (number of anchors: one vs. two)  $\times$  5 (anchor A percentile)  $\times$  4 (question, within-subjects) mixed design. Participants in the one-anchor conditions judged whether the true value was lower or higher than anchor A (e.g., “Do you think the height of the Sears Tower is lower or higher than 560 feet?”), with the order of the words “lower” and “higher” counterbalanced. Participants in the two-anchor conditions instead judged whether the true value was lower than anchor A, between anchor A and anchor B, or higher than anchor B (e.g., “Do you think height of the Sears Tower is lower than 560 feet, between 560 feet and 5700 feet, or higher than 5700 feet?”). The

order of the “lower” and “higher” ranges was also counterbalanced (e.g., “Do you think height of the Sears Tower is higher than 5700 feet, between 5700 feet and 560 feet, or lower than 560 feet?”). After the anchoring judgment, all participants provided an absolute estimate of the true value (e.g., “What do you think the height of the Sears Tower is (in feet)?”). Participants then repeated this procedure for all four questions in counterbalanced order.

We used 20th, 30th, 40th, and 50th percentiles as moderate levels of anchor A. In essence, this experiment is a 2 (number of anchors)  $\times$  2 (extremity of anchor A: extreme vs. moderate)  $\times$  4 (questions, within-subjects) mixed design with four levels of moderate anchor A. Using multiple levels of moderate anchor A enabled us to more carefully examine how the contrast effect evolves with anchor A extremity.

## **Results and Discussion**

Participants’ estimates for Beethoven’s birth year were negatively skewed and their estimates for other questions were positively skewed, so we performed the inverse of the logarithm of the inverted estimates for the Beethoven question and logarithmic transforms for other questions. After removing outliers more than three standard deviations from the mean, there were 3039 estimates across the four questions. We performed all analyses by z-scoring the log-transformed estimates for each question and pooling data across all questions. Table 3 lists the geometric mean for each question and Figure 3 plots the z-scores averaged across questions.

-----  
 Insert Table 3 about here  
 Insert FIGURE 3 about here  
 -----

*Estimates.* We first analyzed the data using an omnibus  $2(\text{number of anchors}) \times 5(\text{anchor A percentile}) \times 4(\text{question})$  ANOVA with question as a within-participant variable. The analysis revealed main effects of adding anchor B,  $F(1, 2189) = 368.83, p < .0001$ , of the percentile of anchor A,  $F(4, 3029) = 15.90, p < .0001$ , and an interaction effect between number of anchors and anchor A percentile,  $F(4, 3029) = 13.81, p < .0001$ .

We predicted that lower anchor A values would generate estimates that are lower than or equal to estimates generated by higher anchor A values when anchor A was presented alone. Conversely, when anchor A was paired with anchor B, we predicted that extremely low anchor A values would instead generate *higher* estimates than moderate anchor A values would. To test these predictions, we compared the extremely low 5th percentile anchor with the moderate 20th percentile anchor using a  $2(\text{number of anchors}) \times 2(\text{anchor A extremity: extreme [5th] vs. moderate [20th]}) \times 4(\text{question})$  ANOVA with question as a within-participant variable. This analysis revealed main effects of adding anchor B,  $F(1, 889) = 204.40, p < .0001$ , and of anchor A extremity,  $F(1, 889) = 3.81, p < .05$ . More importantly, the analysis returned the expected interaction between number of anchors and anchor A extremity,  $F(1, 889) = 3.93, p < .05$ . Specifically, estimates generated by the 5th and 20th percentile anchors were approximately equal when the anchors appeared alone ( $M_{5th} = -.51$  vs.  $M_{20th} = -.50$ ),  $F(1, 889) = 1.60, NS$ , whereas estimates were significantly higher for the 5th percentile anchor than for the 20th percentile anchor when each appeared together with anchor B ( $M_{5th} = .54$  vs.  $M_{20th} = .28$ ),  $F(1, 889) = 9.66, p < .01$ .

For instance, participants who indicated whether the height of the Sears Tower was lower or higher than 50 feet (5th percentile, mean plausibility = 0.71) generated estimates ( $M_{5th} = 1016$ ) that were not significantly different from estimates generated by those who were given an anchor of 560 feet (20th percentile; plausibility rating = 5.50,  $M_{20th} = 1095$ ),  $F(1, 318) = .07$ , *NS*. However, when each of these anchors appeared together with a second, 5700 feet anchor (90th percentile, plausibility rating = 5.50), the 5th percentile anchor generated higher estimates ( $M_{5th} = 3699$ ) than did the 20th percentile anchor ( $M_{20th} = 2992$ ),  $F(1, 318) = 4.99$ ,  $p < .05$ . These results suggest that although estimates generally assimilate to anchors regardless of their extremity in single-anchor conditions, people rely less on extreme anchors when they are paired with a second moderate anchor, leading to a contrast effect where lower values of anchor A generated higher estimates than higher, but more moderate values of anchor A.

We ran additional 2 (number of anchors)  $\times$  2 (anchor A extremity)  $\times$  4 (questions) repeated-measures ANOVAs to make pairwise comparisons between the 5th percentile anchor with each of the more moderate 30th, 40th, and 50th percentile anchors. All ANOVAs returned significant interactions between number of anchors and anchor A extremity,  $F_s = 17.99, 21.94$ , and 35.78, respectively, and all  $p$ 's  $< .0001$ , as well as significant paired contrasts between the 5th and the 30th, 40th, and 50th percentile anchors. The extreme 5<sup>th</sup> percentile anchor A generated significantly smaller estimates than the moderate anchor A values in the one-anchor condition ( $M_{5th} = -.51$  vs.  $M_{30th} = -.30$ ,  $M_{40th} = -.10$ , and  $M_{50th} = .09$ ),  $F_s = 6.10, 29.29$  and 72.92, all  $p < .05$ , but significantly higher estimates in the two-anchor condition ( $M_{5th} = .54$  vs.  $M_{30th} = .22$ ,  $M_{40th} = .41$ , and  $M_{50th} = .42$ ),  $F_s = 20.09, 5.54$  and 5.36, all  $p$ 's  $< .05$ .

We also found a nonmonotonic, U-shaped relationship between anchor A values and estimates (see figure 3) that is very similar to the shape of the contrast curve in Figure 1. Although this nonmonotonicity is not the focus of the current research, our theory predicts a nadir in the anchor response function for a certain anchor value. To the left of the nadir, increasingly extreme values of anchor A create more dissimilarity in plausibility with anchor B, leading to increasing reliance on anchor B and higher estimates. To the right of the nadir, the assimilation effect dominates, leading to estimates increasing with anchor A. We will show a similar pattern in study 4.

*Plausibility Ratings.* We have shown that the extremely low anchor A generated higher estimates than the more moderately low anchor A when each was paired with anchor B. We suggest that this is because when paired with anchor B, an extremely low anchor A is judged as less plausible than anchor B, leading people to discount anchor A and rely more on anchor B when making estimates. On the other hand, a moderately low anchor A seems similarly plausible as anchor B, allowing people to rely on both anchors. To test this hypothesis, we had 117 participants evaluate the plausibilities for both anchor A and anchor B in the two-anchor conditions in a 5 (percentile of anchor A)  $\times$  4 (question, within-subjects) mixed design. As expected, when paired with anchor B, the extremely low anchor seemed much less plausible than anchor B ( $M_A = 1.50$  vs.  $M_B = 7.00$ ),  $F(1, 197) = 42.22, p < .0001$ . Although the moderate values of anchor A were also judged as less plausible than anchor B ( $M_A = 4.56$  vs.  $M_B = 5.87$ ),  $F(1, 197) = 9.35, p < .001$ , the difference in plausibility between anchor A and anchor B in the two anchor

condition was greater for the extreme anchor A,  $F(1, 197) = 15.82, p < .001$ , suggesting that people are more likely to rely on anchor B when anchor A is more extreme.

The results of study 2 were consistent with our hypotheses: When there was only one anchor, estimates assimilated to anchor values. However, when those anchors appeared along with a plausibly high anchor B, extremely low anchors generated higher estimates than moderately low anchors. Participants' plausibility ratings showed that the extremely low anchor A seemed significantly less plausible than anchor B whereas more moderately low values of anchor A seemed more similarly plausible as anchor B. Consequently, participants relied more on anchor B and generated higher estimates when anchor A was extremely low than when it was moderately low.

### **STUDY 3: MANIPULATING THE PERCEIVED DIFFERENCE BETWEEN ANCHORS**

We have explained the contrast effects found in studies 1 and 2 by suggesting that the difference in plausibilities between an extreme anchor and a second, plausible anchor causes people to rely more on the more plausible anchor for generating estimates. If the perceived differences in plausibility really matters, then the contrast effect can be heightened by enlarging the perceived plausibility difference between anchors.

Whether people evaluate two anchors as having similar or different plausibility can be influenced by seemingly unimportant factors. For instance, people's judgments of their own physical attractiveness relative to a comparison target can be influenced by whether participants

believe that they were born on the same day as the target (Brown, Novick, Lord, and Richards 1992), presumably because the birthday information influences people's judgments about the similarity between themselves and the target. In this study, we manipulated perceived anchor similarity using seemingly unrelated factors: the text colors that were used to present the anchors. People may infer that the difference in colors implies differences in plausibilities, or they may misattribute the difference in presentation colors to differences in anchor plausibilities. We therefore predict that people are more likely to judge two anchors to be more different in plausibility when they are presented in different colors than when they are presented in the same color. By demonstrating that heightening perceived difference between anchors generates the same results as enlarging differences between anchor values, we provide additional evidence that perceived difference in plausibility drives the contrast effect we demonstrated in studies 1 and 2.

Study 3 fixed the anchor values and manipulated the perceived difference between two anchors by changing the text colors used to present them. In addition, whereas study 1 varied low anchors and study 2 varied high anchors separately, study 3 included both extremely low and extremely high anchors.

## **Methods**

*Anchor selection.* From the pretest question pool of study 2, we selected a new general knowledge question regarding the number of bones in the adult human body. We also followed

similar procedures to select the anchor values. We selected 78 for the extremely low anchor A, 2432 for the extremely high anchor A, and 228 for the moderate anchor B.

*Main study.* A total of 502 participants took part in a 2 (magnitude of anchor A: low vs. high)  $\times$  3 (color manipulation: no vs. same vs. different) between-subjects design. The no color manipulation conditions presented only anchor A, whereas the same and different color conditions presented anchor A along with anchor B. To further manipulate the perceived difference between the two anchors, we varied the text color of the anchor values. In the same-color conditions, both anchors were displayed in the same font color, either red or blue. In the different-color conditions, one anchor was in red and the other was in blue, with the colors counterbalanced between anchors. We expected that presenting the anchors in different colors would increase the perceived difference between the two anchors and thus make the extreme anchors seem even more implausible, which in turn would cause people to rely even less on the extreme anchors in making their estimates. To control for the effects that comes from the color itself rather than the perceived difference between colors, we also displayed the single anchor in either red or blue in the no-difference conditions. Finally, in order to measure the effect of anchor B by itself, a subset of the participants was assigned to an additional anchor-B-only condition. Like in the first two studies, participants first answered the anchoring question and then made an estimate about the number of bones in human body. After providing estimates, participants rated the plausibility of the anchor(s).

## **Results and Discussion**

*Estimates.* We performed all analyses on log-transformed estimates, and report results using geometric means where appropriate (*GM*; exponential of the mean of the log-transformed data). A 2 (magnitude of anchor A)  $\times$  3 (color manipulation: no vs. low vs. high) ANOVA revealed a main effect of anchor A magnitude,  $F(1, 440) = 96.06, p < .0001$ , and an interaction between anchor A magnitude and color manipulation,  $F(2, 440) = 16.41, p < .0001$ . Adding a covariate of whether anchor A appeared in red or blue did not affect these results,  $F(1, 439) = .01, ns$ .

To better understand the results, we next performed a 2 (magnitude of anchor A)  $\times$  2 (color manipulation: no vs. same color) ANOVA on only the no-color-manipulation and same-color conditions. The analysis revealed a main effect of anchor A magnitude,  $F(1, 261) = 105.25, p < .0001$ , where higher anchors generated higher estimates ( $GM_{low} = 211, GM_{high} = 522$ ), qualified by the expected interaction between anchor A magnitude and color manipulation,  $F(1, 261) = 10.09, p < .01$ . As seen in Figure 4, participants who saw the extremely low anchor A (78) and anchor B in the same color made higher estimates ( $GM = 253$ ) than did participants who saw only the extremely low anchor A ( $GM = 163$ ),  $F(1, 261) = 8.93, p < .01$ . Likewise, participants who saw the extremely high anchor A (2432) and anchor B together in the same color made somewhat lower estimates ( $GM = 555$ ) than did those who only saw the extremely high anchor A ( $GM = 725$ ),  $F(1, 261) = 2.58, p = .11$ .

Next, to test whether increasing the perceived difference between the two anchors generated estimates that were even further away from anchor A, we performed a 2 (magnitude of anchor A)  $\times$  2 (color manipulation: same vs. different) ANOVA on only the same-color and

different-color conditions. This analysis again returned a significant main effect of anchor A magnitude,  $F(1,323) = 28.14, p < .0001$ , where the extremely high anchor A generated higher estimates than the extremely low anchor A ( $GM_{\text{high}} = 462, GM_{\text{low}} = 280$ ), qualified by the interaction between anchor A's magnitude and color manipulation,  $F(1,323) = 7.60, p < .01$ . As predicted, presenting the extremely low anchor A and anchor B in different colors increased estimates ( $GM = 305$ ) compared to presenting them in the same colors ( $GM = 253$ ),  $F(1,323) = 2.85, p < .10$ . Similarly, presenting the extremely high anchor A and anchor B in different colors decreased estimates ( $GM = 400$ ) compared with presenting them in the same colors ( $GM = 555$ ),  $F(1,323) = 5.44, p < .05$ . These results confirmed our hypothesis that increasing the perceived difference between two anchors lead people to generate estimates that are even further away from the extreme anchor, suggesting that they were relying more on the moderate anchor.

-----  
 Insert Figure 4 about here  
 -----

*Plausibility Ratings.* The perceived differences between anchors, we suggest, can be captured by differences in plausibility ratings. As in study 1, we removed participants who either misused the plausibility scale or misunderstood the meaning of plausibility for this analysis. Figure 5 shows the mean plausibility rating for anchor A as a function of anchor magnitude and contrast. A 2 (magnitude of anchor A)  $\times$  3 (color manipulation) ANOVA on plausibility ratings for anchor A revealed a significant main effect of perceived difference,  $F(2, 403) = 10.25, p < .0001$ . Plausibility did not depend specifically on whether anchor A was red or blue,  $F(1, 402) = .07, ns$ .

More specifically, a 2 (magnitude of anchor A)  $\times$  2 (color manipulation: no vs. same) ANOVA on anchor A plausibility returned a main effect of color manipulation,  $F(1, 244) = 5.54$ ,  $p < .05$ . Participants rated anchor A as less plausible when it was presented with anchor B in the same color ( $M_{\text{same}} = 3.99$ ) than when it was presented alone ( $M_{\text{no}} = 5.01$ ). More importantly, a 2 (magnitude of anchor A)  $\times$  2 (color manipulation: same vs. different) ANOVA on anchor A plausibility also returned a main effect of color manipulation,  $F(1, 287) = 4.25$ ,  $p < .05$ . Participants rated anchor A as even less plausible when anchors A and B appeared in different colors ( $M_{\text{different}} = 3.22$ ) than when they appeared in the same color ( $M_{\text{same}} = 3.99$ ). There was no effect of contrast on anchor B,  $F(2, 344) = .01$ , *ns*.

In addition, showing the two anchors in different colors increased the perceived plausibility difference between anchor A and anchor B ( $M_{\text{same}} = 2.03$  vs.  $M_{\text{different}} = 2.63$ ),  $F(1, 287) = 3.30$ ,  $p = .07$ . We suggest that this enlarged plausibility difference made people rely more on anchor B, which led to higher estimates for the extremely low anchor A and lower estimates for the extremely high anchor A.

-----  
 Insert FIGURE 5 about here  
 -----

In summary, these results showed that increasing perceived difference between two anchors makes extreme anchors seem even more implausible, providing additional evidence that people rely more on the more plausible anchor when two anchors seem very different.

## STUDY 4: EBAY AUCTIONS

The first three studies demonstrated the difference between one and two anchors in experimental settings. In study 4, we replicate this effect in a real purchase setting using auction data from eBay.com. Past research has demonstrated that price cues can serve as anchors to affect bidders' bidding behavior. For example, increasing Buy-It-Now (BIN) prices increases bidders' willingness to pay (Popkowski Leszczyc, Qiu, and He 2009). Thus, eBay auctions provide a real world test of anchoring effects.

To test the difference between one and two price anchors, we exploited an exogenous difference in the BIN auction format between eBay's websites in the United States and China. In BIN auctions, bidders can choose between placing a bid and paying the BIN price to get the item immediately. More relevant to this research, an important difference in the BIN auction formats between the United States and China creates a natural experiment, analogous to our controlled experiments. On the U.S. eBay website, the BIN option becomes invisible (and thus unavailable) if any bidder submits a bid higher than the reserve price (i.e., the minimum selling price that sellers can optionally set, but which is not shown to bidders). As a result, the only salient anchor seen by most bidders in U.S. BIN auctions is the starting price, similar to the one-anchor case in our controlled experiments. However, when eBay acquired eachnet.com—the Chinese equivalent of eBay—and made it eBay China, eBay kept the BIN auction format of eachnet.com where the BIN option remained available throughout the auction. As a result, both the BIN price and

starting price serve as anchors on Chinese eBay BIN auctions, similar to the two-anchor case in our previous studies.

The BIN price, which is close to the market price, usually seems plausible because buyers can actually buy the item at that price. However, starting price, which has to be lower or equal to the BIN price, can seem implausible when it is extremely low. For example, for a digital camera with a BIN price of \$300, a \$1 starting price seems implausible because consumers are not likely to assume that \$1 reflects the camera's true value.

We predict that after controlling for other factors that may affect auction outcomes, extremely low starting prices will lead to lower final prices in BIN auctions than will moderate starting prices in the U.S. market, where starting price serves as the only salient anchor. In contrast, because the BIN price serves as a plausible second anchor, we instead expect that extremely low starting prices will lead to higher final prices than will moderate starting prices in the Chinese market.

One challenge in using eBay data was separating the differences arising from the intrinsic differences between the two markets from the differences caused by the number of anchors. Unlike in controlled experiments in which we can manipulate each variable without affecting others, many factors in eBay auctions may covary with the number of anchors and contribute to differences in the final prices. For example, Chinese eBay users may have different characteristics and preferences from U.S. eBay users, which can possibly contribute to the differences in final prices between the two markets.

To control for these differences between the two markets, we used pure auctions—in which bidders can only bid on the auction and never have the option to buy-it-now—as a control condition. As the auction format that is identical across countries, pure auctions can be used to capture the intrinsic differences between the two countries. Thus, any additional differences that exist for BIN auctions but not for pure auctions can be attributed to the number of anchors. We expect a significant between-country difference in the pattern of final prices for BIN auctions above and beyond the between-country differences in pure auctions.

### **Data Description**

We used Canon digital cameras as our target product. Canon cameras are available in both China and the United States with identical specifications and are therefore comparable across the two markets. We collected data for successful transactions of Canon digital point-and-shoot cameras for pure auctions and for BIN auctions that were sold through the auction option (rather than the BIN option) from February 10 to June 17, 2007, on eBay U.S. and eBay China. Because the total number of listings on eBay China was much smaller than that on eBay U.S., and to meet a database access quota specified by eBay, we randomly selected 10% of the listings from eBay U.S. but used all listings from eBay China.

For each transaction, we collected (a) product specifications, including the camera model, whether additional accessories (e.g., memory cards, camera case) were included, and shipping costs; (b) buyer and seller ratings as a proxy for experience, consisting of the net number of

positive and negative evaluations buyers have received from sellers they have purchased from and vice versa; and (c) auction characteristics, such as auction format, BIN price for BIN auctions, starting price, final price, and bidding history.

### **Model Specification**

Our data included different camera models with a range of market values. We aggregated the information across all models and conducted the analysis in relative terms. Specifically, we defined the relative starting price and relative final price as respective ratios of these prices to the market price of the model, including any bundled accessories. The market prices were collected from the most popular online shopping websites in the United States (Amazon.com) and China (pchome.net). We did not include used or refurbished products as we would not be able to determine their market value without a full appraisal of their conditions. The final data set included 288 pure auctions and 83 BIN auctions for the United States and 56 pure auctions and 83 BIN auctions for China.

As we have discussed, both starting price and BIN price can serve as anchors and affect bidders' bidding behavior. However, starting price can also affect the final price through traffic and escalation of commitment. Ku, Galinsky, and Murnighan (2006) showed that low starting prices led to higher final prices than did medium starting prices, because a lower starting price decreased the entry barrier, thus attracting more bidders (i.e., more traffic) who drove the final price up. Related research demonstrated that bidders also were willing to bid higher with more

bidders in the auction, exhibiting an irrational escalation of commitment (Heyman, Orhun, and Ariely 2004; Ku, Galinsky, and Murnighan 2006). Demonstrating the anchoring effect would require controlling for these participation and competition effects, so we included two additional variables—traffic and escalation—in our analysis. Similar to Ku et al. (2006), we defined traffic as  $(\text{number of bidders} + \text{number of bids})/2$  and escalation as  $(\text{number of bids}/\text{number of bidders})$ . Because traffic and escalation were on different scales, we used within-country z-scores for each variable in our model.

We also included ratings of both buyers (bidders who actually won auctions) and ratings of sellers in the model to control for factors related to buyers' experience and sellers' reputation (Cheema 2008; Melnik and Alm 2002). Similar to traffic and escalation, we used the within-country z-scores for buyer and seller ratings.

Our theory predicts that after controlling for these factors, we should find an interaction between starting price, country, and auction format. A starting price paired with a BIN price (BIN auctions in China) should act like the two-anchor case, with the BIN price serving as a plausibly high anchor. Therefore, the final price should be higher when the plausible BIN price appears together with an extremely low starting price than when it appears with a less extreme starting price. In contrast, we expect BIN auctions in the United States to mirror the one-anchor case. The starting price should have a positive monotonic effect on the final price, with higher starting prices resulting in higher final prices. Using pure auctions as a baseline, we expect that the differences in BIN auctions are beyond the differences in pure auctions between the two countries.

To test our prediction, we estimated the following model specification:

$$\text{Relative Final} = a * (\text{Country}) + b * (\text{Auction Format}) + c * (\text{Relative start}) + d * (\text{Country} \times \text{Auction Format}) + e * (\text{Country} \times \text{Relative Start}) + f * (\text{Auction Format} \times \text{Relative Start}) + g * (\text{Country} \times \text{Auction Format} \times \text{Relative Start}) + h * \text{Traffic} + i * \text{Escalation} + j * \text{Seller Rating} + k * \text{Buyer Rating} + \varepsilon$$

In this model, the dependent variable is the ratio of final price to market price of the same camera model. Country is a dummy variable, with a value of 0 for the United States and 1 for China. Auction format is also a dummy variable with a value of 0 for pure auctions and 1 for BIN auctions. Relative start is a continuous variable that captures the ratio of starting price to market price. This model also included the two-way and three-way interaction terms between these variables. Finally, we adopt variables such as traffic, escalation, buyer rating, and seller rating to control for other known effects, and use  $\varepsilon$  to represent unobserved factors that cannot be captured by the observed independent variables. In essence, this model is a three-way ANOVA (country  $\times$  auction format  $\times$  relative start), with relative start as a continuous variable. Table 4 lists the variable definitions, and Table 5 contains summary statistics.

-----  
 Insert TABLE 4 about here  
 Insert TABLE 5 about here  
 -----

## Results and Discussion

Recall that study 2 found a nonmonotonic, U-shaped relationship between low anchor values and estimates in the two-anchor condition. That is, to the left of the nadir in the anchoring response function, further decreasing the value of anchor A instead generated higher estimates

than the optimal anchoring point. To the right of the nadir, subsequent estimates became positively correlated with anchor A value. Similarly, we expected that final prices would exhibit a U-shape relationship with starting prices, initially decreasing with starting price up to a point and then increasing with starting price above that point. Therefore, we only expect to find an interaction between starting price, auction format, and country for starting prices below the nadir. For starting prices above the nadir, final prices should increase along with starting prices regardless of whether people see one anchor (starting price) or two anchors (starting price and BIN price), and the interaction between starting price, auction format, and country may not be as significant. Because we had no *a priori* predictions for the location of the nadir, we tested a few different splitting points by including only data for which the relative starting price is lower than that of the nadir.

First, we included only auctions with a relative starting price equal to or less than 60% of the market price of the product in our analysis. This regression returned a significant three-way (country  $\times$  auction format  $\times$  relative start) interaction,  $F(1, 368) = 2.12, p = .03$ . Specifically, for BIN auctions, we found an interaction between starting price and country,  $F(1,96) = 2.17, p = .03$ . After controlling for traffic and escalation effects, starting price in BIN auctions correlated positively with final price in the United States but negatively with final price in China. For pure auctions, however, the effect of starting price on the final price did not differ between the countries,  $F(1, 268) = 1.08, ns$ . Since the interaction between starting price and country was only found in BIN auctions and not in pure auctions, we suggest that the interaction is caused by differences in number of anchors rather than differences between the two countries. We repeated

this analysis using 50% and 70% as alternative nadir points and found similar results (three-way interactions,  $p = .10$  and  $.03$ , respectively).

As noted previously, in the two-anchor cases (BIN auctions in China), when the starting price is below the nadir, final prices negatively correlated with starting prices. When the starting price was over that point, the correlation of final price with starting price flipped, so that final price positively correlated with starting price as in single-anchor cases. This flip suggests a quadratic relationship between starting price and final price for BIN auctions in China, but not for pure auctions and BIN auctions in the U.S. We therefore supplement the preceding analysis (using only data for which the starting price was lower than a pre-selected threshold) with quadratic regressions that fit all auction data separately for each country:

$$\text{Relative Final} = a * (\text{Relative Start})^2 + b \text{ Relative Start} + c * \text{Traffic} + d * \text{Escalation} + e * \text{Seller Rating} + f * \text{Buyer Rating} + \varepsilon$$

We included the quadratic relative starting price term to allow for a nonmonotonic relationship between final price and starting price. As a control condition for each country, we separately estimated the same equation for pure auctions in each country. Table 6 shows the regression results.

Based on the regression results, Figure 6 depicts the predicted final price as a function of the starting price, given other covariates at their means. We used pure auctions as a baseline to capture differences between the two markets that were not caused by the number of anchors. We expect the fitted lines for pure auctions and BIN auctions in the U.S. to be similar, as both formats consist only of one anchor. If two anchors (starting price and BIN price) generate

different final prices than a single anchor (starting price), we expect the fitted line for BIN auctions in China to be different from the fitted line for pure auctions in China. The results were consistent with our predictions. In the U.S. market, the two fitted lines were almost flat and similar to one another. In the Chinese market, although the final prices for pure auctions monotonically increased as a function of starting prices, the final prices for BIN auctions were higher for extremely low starting prices than for less extreme ones. When starting prices were closer to the BIN prices, final prices increased along with starting prices.

-----  
 Insert FIGURE 6 about here

Insert TABLE 6 about here  
 -----

The regression analysis also provided evidence consistent with past research finding that more traffic, escalation of commitment, and higher seller's ratings lead to higher final prices. A one standard deviation increase of traffic led to an increase of 3% in final price for the U.S. market ( $M = 13.94$ ,  $SD = 6.62$ ,  $p = .02$ ), and an increase of 5% in final price in the Chinese market ( $M = 6.29$ ,  $SD = 8.36$ ,  $p = .04$ ). Similarly, a one standard deviation increase in escalation led to an increase of 3% in final price for the U.S. market ( $M = 2.09$ ,  $SD = .85$ ,  $p < .01$ ) and an increase of 3% in final price for the Chinese market ( $M = 1.59$ ,  $SD = 1.09$ ,  $p = .08$ ), though only for pure auctions. Furthermore, sellers with higher ratings sold cameras for a marginally higher price in pure auctions in United States ( $p = .12$ ) and in BIN auctions in China ( $p = .10$ ). More experienced buyers tended to pay less for a given camera model, though this effect was only significant for BIN auctions in China ( $p < .001$ ).

In summary, study 4 generalized the differential impact of one versus two anchors to product evaluations in a dynamic, competitive auction environment. After controlling for other variables, extremely low starting prices led to lower final prices when the BIN price was not observable but higher final prices when the BIN price served as a second plausible anchor.

Although we earnestly tried to control for the potential differences caused by differences between countries rather than number of anchors (such as using pure auctions as the control condition), field data are usually subject to the influences of other factors that we cannot perfectly control. For example, people from collectivistic cultures, relative to those from individualistic cultures, tend to care more about other people's outcomes in reward allocation tasks (Heydenfeldt 2000). Therefore, one possible alternative explanation of our results is that an extremely low starting price, when juxtaposed with the BIN price, triggered concerns about the seller's well-being among Chinese buyers, which may have resulted in higher final prices compared to when there was no BIN price. This concern for others is less likely to manifest in the individualistic U.S. market, regardless of the presence of a BIN price. Although this cultural difference could technically explain this effect, we believe it is less of a concern in the eBay context. Unlike in other situations such as reward allocations, we believe that eBay shoppers' only aim is to find good deals and they are therefore unlikely to care about sellers' welfare, irrespective of their cultural backgrounds.

## GENERAL DISCUSSION

We have examined differences in judgments arising from the influence of one versus two anchors in three controlled experiments and a field study. The results consistently demonstrated that estimates in response to two anchors did not vary monotonically with the anchor values, unlike the monotonic response to a single anchor. Instead, we found a contrast effect where an extremely low (high) anchor paired with a moderate second anchor led to a higher (lower) estimate than when a moderately low (high) anchor was paired with the same moderate second anchor. That is, estimates were less extreme for the extreme anchor than for the moderate anchor when each was paired with an additional moderate anchor (studies 1 and 2). We also demonstrated that increasing the perceived difference between anchors leads people to rely even less on the less plausible anchor and thus produces a stronger contrast effect (study 3). Finally, we found the same relationships between starting prices and final prices with and without a second price anchor in actual eBay Buy-It-Now auctions (study 4). After controlling for other variables that could affect final price, we found that when bidders only observed the starting price, a lower starting price led to a lower final price; but when bidders could also see the BIN price, the implausibly low starting prices generated higher final prices than did more plausible starting prices.

This research makes both theoretical and practical contributions. Theoretically, this investigation represents a first step in exploring the effect of multiple anchors, shedding light on how people generate estimates when they encounter more than one anchor. Specifically, we have

demonstrated that when an extreme anchor A is paired with a moderate anchor B, the large gap in plausibility between the two anchors causes people to rely less on the extreme anchor than when it is presented alone. However, when a moderate anchor A is paired with anchor B, both anchors seem plausible, so estimates are influenced by both anchors. Because there is less reliance on anchor A when it is extreme than when it is moderate, estimates are closer to anchor B when anchor A is extreme, producing a contrast effect relative to when anchor A is moderate.

Although this research has only demonstrated the effect for two anchors, we expect the same mechanism to generalize to even more anchors. The presence of additional anchors should continue to influence people's usage of the original anchor. We expect that since adding a single moderate anchor is sufficient for minimizing the impact of an extreme anchor, adding even more moderate anchors may not have much additional effect on decreasing the impact of extreme anchors. We also suspect that a single moderate anchor may allow people to ignore multiple extreme anchors. That being said, it is possible that extreme anchors can begin to seem more plausible than the single moderate anchor when there are many extreme anchors, simply because people may infer plausibility from number of anchors. Further experiments are needed to explore how increasing the number of moderate or extreme anchors would influence judgments of anchor plausibility and subsequent estimates.

Practically, this research has important implications for pricing strategies. According to our findings, managers need to consider the plausibility of prices when there is more than one price available for comparison. When introducing a new product, managers often set a manufacturer suggested retail price (MSRP) along with a sale price. The MSRP provides a high

reference point and therefore is expected to increase consumers' willingness-to-pay. However, our findings suggest that a very high MSRP may backfire. Because MSRP is always accompanied by the actual sale price, which by definition is a plausible anchor, an extremely high MSRP might actually lead to a lower willingness-to-pay than a less extreme MSRP. Although MSRP should not be set very high, our research suggests that historical prices should be presented together with the current price only when they are very low, such as a one-time Black Friday loss leader. Extremely low historical prices may not seem as relevant to a product's current value and can therefore be more easily ignored. Consumers might exhibit a higher willingness-to-pay than if they were told a more moderately low historical sale price.

There are a number of questions we did not address that may provide fruitful avenues for future research. First, our research has considered the effect of pairing one moderate or extreme anchor with a second moderate anchor but has not examined what happens when both anchors are extreme. Will pairing two extreme anchors lead to a contrast or assimilation effect? Recall that a contrast effect in the two anchor conditions in our studies meant that, controlling for the same moderate anchor B, extreme values of anchor A generated less extreme estimates than moderate values of anchor A. That is, we used the condition where both anchors A and B are moderate as a baseline for comparison. However, if we replace the moderate anchor B with an extreme one, the condition with moderate anchor A and extreme anchor B is no longer a good baseline for the condition where both anchors A and B are extreme. The combination of a moderate anchor A and an extreme anchor B already produces a contrast effect by itself (relative to the condition where both anchors are moderate). Therefore, we do not have an

appropriate baseline to test for a contrast effect in the condition where both anchors are extreme.

It would require a different experimental design to examine whether there is a contrast or assimilation effect when both anchors are extreme.

Second, our investigation of the effect of multiple anchors has focused more on the relationship between anchor values and final estimates but less on how the anchoring effect works. Research on anchoring effect has identified two main mechanisms: selective accessibility (Mussweiler and Strack 2001) and insufficient adjustment (Tversky and Kahneman 1974).

Adding another anchor does not make one mechanism more likely than the other. Hence, our findings would hold regardless of which mechanism is driving the anchoring effect. We therefore remain agnostic about whether the effect in the multiple-anchor cases results from selective accessibility or insufficient adjustment.

## **Conclusion**

Our world is full of anchors and in many cases there may be more than one relevant anchor for a decision. Consumers see both MSRP and retail prices, managers receive multiple sales estimates from employees, and bidders on eBay in China observe both BIN and starting prices. Using data from controlled experiments and eBay auctions, we have shown that when presented along with a second, moderate anchor, extreme anchors can have the unexpected effect of generating more moderate estimates than less extreme anchors do.

## REFERENCES

- Ariely, Dan, George Loewenstein, and Drazen Prelec (2003), "Coherent arbitraries: Stable demand curves without stable preferences," *The Quarterly Journal of Economics*, 118 (1), 73-105.
- Brown, Jonathon D., Natalie J. Novick, Kelley A. Lord, and Jane M. Richards (1992), "When Gulliver travels: Social context, psychological closeness, and self-appraisals," *Journal of Personality and Social Psychology*, 62 (5), 717-27.
- Chapman, Gretchen B., and Eric J. Johnson (1994), "The limits of anchoring," *Journal of Behavioral Decision Making*, 7 (4), 223-42.
- Cheema, Amar (2008), "Surcharges and seller reputation," *Journal of Consumer Research*, 35 (1), 167-77.
- Epley, Nicholas (2004), "A tale of tuned decks? Anchoring as accessibility and anchoring as adjustment," in *Blackwell Handbook of Judgment and Decision Making*, Derek J. Koehler and Nigel Harvey, eds. Oxford, Blackwell Publishers, 240-56.
- \_\_\_\_\_ and Thomas Gilovich (2001), "Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors," *Psychological Science*, 12 (5), 391-96.
- Green, Donald, Karen E. Jacowitz, Daniel Kahneman, and Daniel McFadden (1998), "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods," *Resource and Energy Economics*, 20 (2), 85-116.

- Heydenfeldt, Jo Ann G. (2000), "The influence of individualism/collectivism on Mexican and US business negotiation," *International Journal of Intercultural Relations*, 24 (3), 383–407.
- Heyman, James E., Yesim Orhun, and Dan Ariely (2004), "Auction fever: The effect of opponents and quasi-endowment on product valuations," *Journal of Interactive Marketing*, 18 (4), 7-21.
- Janiszewski, Chris and Dan Uy (2008), "Anchor precision influences the amount of adjustment," *Psychological Science*, 19 (2), 121-27.
- Kamins, Michael, A., Xavier Drèze, and Valerie S. Folkes (2004), "Effects of seller-supplied prices on buyers' product evaluations: Reference prices in internet auction context," *Journal of Consumer Research*, 30 (3), 622-28.
- Ku, Gillian, Adam D. Galinsky, and J. Keith Murnighan (2006), "Starting low but ending high: A reversal of the anchoring effect in auctions," *Journal of Personality and Social Psychology*, 90 (6), 975-86.
- Melnik, Mikhail and James Alm (2002), "Does a seller's ecommerce reputation matter? Evidence from eBay auctions," *Journal of Industrial Economics*, 50 (3), 337-49.
- Mochon, Daniel and Shane Frederick (2009), "Anchoring in sequential judgments," unpublished manuscript, Marketing Department, Rady School of Management, University of California, San Diego, CA 92093.
- Mussweiler, Thomas (2003), "Comparison processes in social judgment: Mechanisms and Consequences," *Psychological Review*, 110 (3), 472-89.

- \_\_\_\_\_ and Birte Englich (2005), "Subliminal anchoring: Judgmental consequences and underlying mechanisms," *Organizational Behavior and Human Decision Processes*, 98 (2), 133-43.
- \_\_\_\_\_ and Fritz Strack (2001), "Considering the impossible: Explaining the effects of implausible anchors," *Social Cognition*, 19 (2), 145-16.
- Nunes, Joseph C. and Peter Boatwright (2004), "Incidental prices and their effect on willingness to pay," *Journal of Marketing Research*, 41 (4), 457-66.
- Popkowski Leszczyc, Peter T.L., Chun Qiu, and Yongfu He (2009), "Empirical testing of the reference price effect of buy-now prices in internet auctions," *Journal of Retailing*, 85 (2), 211-21.
- Quattrone, George A., Cheryl P. Lawrence, Diann L. Warren, Kristina Souza-Silva, Steven E. Finkel, and D. Calvin Andrus (1981), "Explorations in anchoring: The effects of prior range, anchor extremity, and suggestive hints," *Unpublished manuscript*, Stanford University.
- Roggeveen, Anne L. and Gita V. Johar (2004), "Integration of discrepant sales forecasts: The influence of plausibility inferences based on evoked range," *Journal of Marketing Research*, 41 (1), 19-30.
- Simonson, Itamar and Aimee Drolet (2004), "Anchoring effects on consumers' willingness-to-pay and willingness-to-accept," *Journal of Consumer Research*, 31 (3), 681-90.

- Skowronski, John J. and Donal E. Carlston (1987), "Social Judgment and Social Memory: The role of cue diagnosticity in negativity, positivity, and extremity Biases," *Journal of Personality and Social Psychology*, 52 (4), 689-99.
- Strack, Fritz and Thomas Mussweiler (1997), "Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility," *Journal of Personality and Social Psychology*, 73 (3), 437-46.
- Thomas, Manoj, Daniel Simon, and Vrinda Kadiyali (2010), "The Price Precision Effect: Evidence from Laboratory and Market Data," *Marketing Science*, 29 (1), 175-90.
- Tversky, Amos and Daniel Kahneman (1974), "Judgment under uncertainty: Heuristics and biases," *Science*, 185 (4157), 1124-131.
- Urbany, Joel, E., William O. Beardon, and Dan C. Weilbaker (1988), "The effect of plausible and exaggerated reference prices on consumer perceptions and price search," *Journal of Consumer Research*, 15 (1), 95-110.
- Wansink, Brian, Robert J. Kent, and Stephen J. Hoch (1998), "An anchoring and adjustment model of purchase quantity decisions," *Journal of Marketing Research*, 35 (1), 71-81.
- Wegener, Duane T., Richard E. Petty, Brian T. Detweiler-Bedell, and W. Blair G. Jarvis (2001), "Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness," *Journal of Experimental Social Psychology*, 37 (1), 62-9.

Whyte, Glen, and James K. Sebenius (1997), "The effect of multiple anchors on anchoring in individual and group judgment," *Organizational Behavior and Human Decision Processes*, 69 (1), 75-85.

Table 1. Products and anchor values used in study 1.

Product	Medium (50 <sup>th</sup> )	Moderately High (85 <sup>th</sup> )	Extremely High (95 <sup>th</sup> )	True Amazon Price
Dell ST2410 Flat Panel Monitor	\$200	\$400	\$800	\$270
Canon PowerShot SD780IS Camera	\$150	\$250	\$350	\$207
Garmin nüvi 855 GPS Navigator	\$180	\$300	\$380	\$126
Zero-Gravity Lounge Chair	\$50	\$100	\$160	\$70
Dyson DC14 Vacuum Cleaner	\$280	\$440	\$580	\$280
Seiko Wall Pendulum Clock	\$90	\$180	\$420	\$173

Table 2. Questions and anchor values for study 2.

Anchor Percentile	5%	20%	30%	40%	50%	90%	True Value
Nile River length (miles)	10	150	310	630	1100	3900	4135
Land-speed record (mph)	4	210	240	280	330	750	760
Sears Tower height (ft)	50	560	850	1000	1200	5700	1730
Beethoven's year of birth	1110	1650	1720	1760	1780	1870	1770

Table 3. Geometric mean for estimates following various anchors in study 2.

Anchor A Percentile	Condition	5%	20%	30%	40%	50%
Nile River length	one-anchor	281	378	587	786	1219
	two-anchor	1706	1294	1822	2247	2615
Land-speed record	one-anchor	208	315	324	328	378
	two-anchor	453	446	399	412	406
Sears Tower height	one-anchor	625	868	938	1331	1410
	two-anchor	2320	1983	1778	2524	2604
Beethoven's year of birth	one-anchor	1730	1704	1739	1761	1769
	two-anchor	1808	1782	1761	1778	1764

Table 4. Variable definitions for the regression equation in study 4.

Definitions of Variables	
Relative start	Starting price divided by the market price of the particular model.
Relative final	Final price divided by the market price of the particular model.
Number of bids	Number of bids a particular auction received.
Number of bidders	Number of bidders in a particular auction.
Buyer rating	Buyer's cumulative rating.
Seller rating	Seller's cumulative rating.
Escalation	Bids per bidder.
Traffic	Average of number of bids plus number of bidders.

Table 5. Summary statistics for study 4. Standard deviations in parentheses.

Auction Format	United States		China	
	Pure Auction	BIN Auction	Pure Auction	BIN Auction
Relative start	0.14 (0.29)	0.63 (0.38)	0.61 (0.34)	0.07 (0.23)
Relative final	0.93 (0.16)	0.95 (0.1)	0.76 (0.19)	0.86 (0.18)
Number of bids	18.94 (10.11)	9.17 (9.81)	8.63 (12.37)	13.49 (13.55)
Number of bidders	8.94 (3.84)	4.73 (4.16)	3.95 (4.65)	5.87 (4.09)
Buyer rating	80.36 (175.05)	117.13 (227.92)	4.59 (8.96)	0.46 (1.28)
Seller rating	98450.18 (123384.21)	52100.16 (103380.47)	3402.36 (2689.78)	293.8 (203.69)
Escalation	2.09 (0.85)	1.61 (0.77)	1.59 (1.09)	1.96 (1.18)
Traffic	13.94 (6.62)	6.95 (6.87)	6.29 (8.36)	9.68 (8.60)
N	288	83	56	83

Table 6. Regression results for study 4.

U.S. Market								
Variable	Pure Auctions				BIN Auctions			
	Parameter Estimate	Standard Error	t Value	Pr >  t	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	0.92	0.01	82.42	<.0001	0.90	0.04	20.44	<.0001
Relative start <sup>2</sup>	0.51	0.10	5.28	<.0001	0.28	0.15	1.89	0.062
Relative start	-0.32	0.10	-3.2	0.002	-0.17	0.15	-1.12	0.268
Traffic	0.03	0.01	2.34	0.020	0.05	0.02	2.13	0.037
Escalation	0.03	0.01	3.55	0.001	-0.01	0.02	-0.59	0.557
Seller rating	0.02	0.01	1.58	0.115	0.00	0.01	0.26	0.797
Buyer rating	0.01	0.01	1.16	0.245	-0.01	0.01	-0.9	0.369

Chinese Market								
Variable	Pure Auctions				BIN Auctions			
	Parameter Estimate	Standard Error	t Value	Pr >  t	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	0.50	0.09	5.05	<.0001	0.86	0.02	46.55	<.0001
Relative start <sup>2</sup>	0.71	0.22	3.26	0.002	2.00	0.82	2.44	0.017
Relative start	-0.07	0.26	-0.28	0.779	-1.66	0.75	-2.23	0.029
Traffic	0.13	0.03	4.19	0.000	-0.03	0.03	-1.06	0.295
Escalation	0.03	0.02	1.79	0.079	0.01	0.02	0.57	0.573
Seller rating	0.04	0.04	1.15	0.368	0.04	0.02	1.64	0.104
Buyer rating	0.02	0.01	0.26	0.257	-0.06	0.02	-3.34	0.001

Figure 1. Three possible forms for the anchoring response function, modified from Chapman and Johnson (1994).

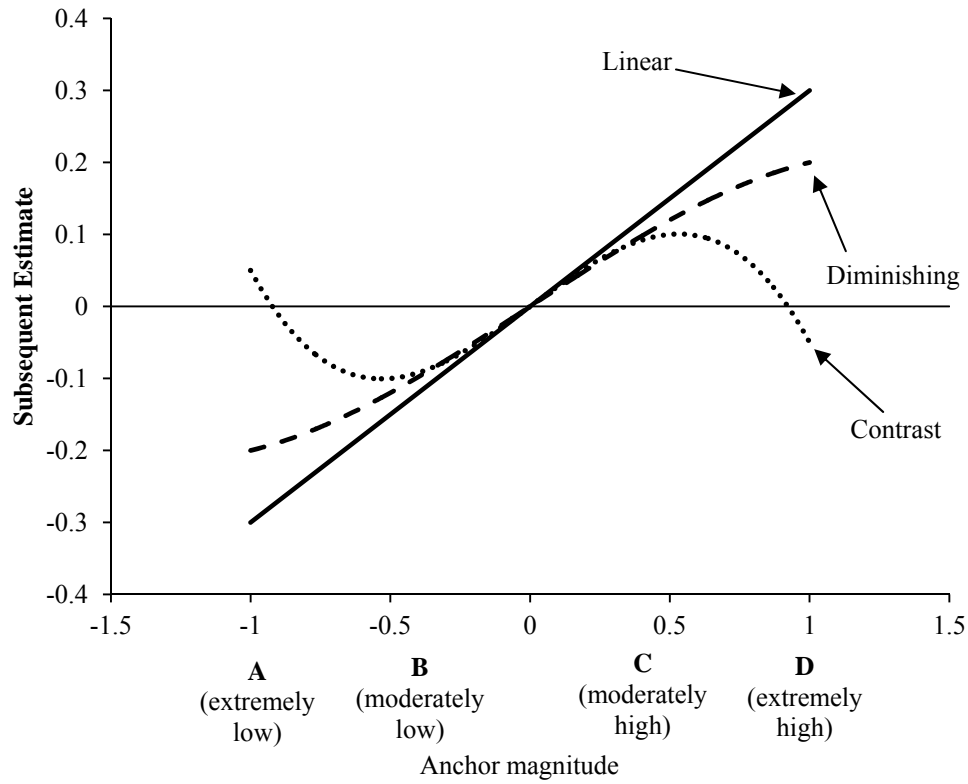


Figure 2. Standardized, logged estimated prices on Amazon.com in study 1. Error bars represent +/- one standard error.

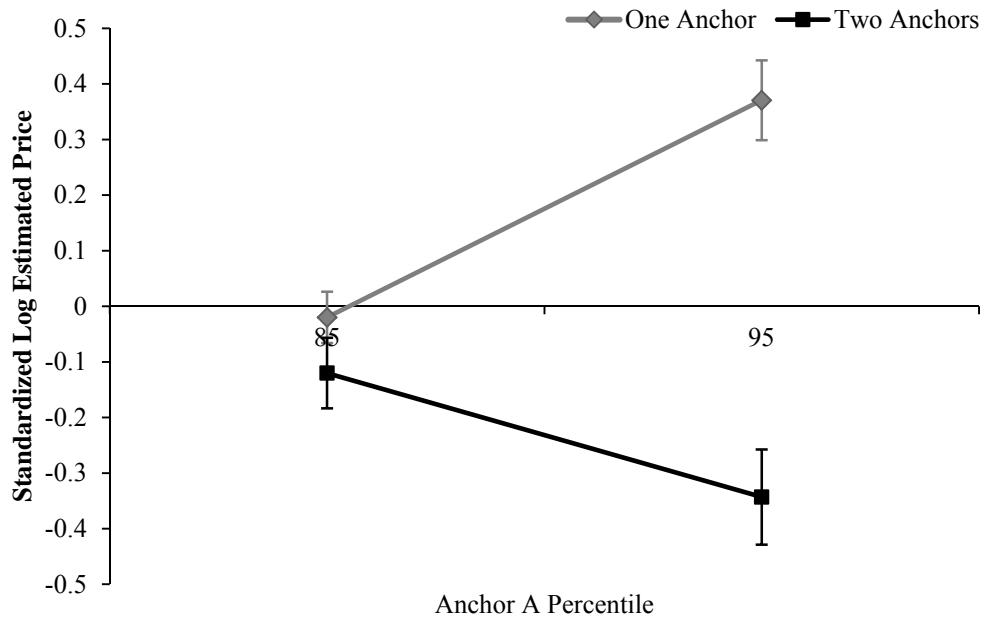


Figure 3. Standardized, logged estimated values in study 2. Error bars represent +/- one standard error.

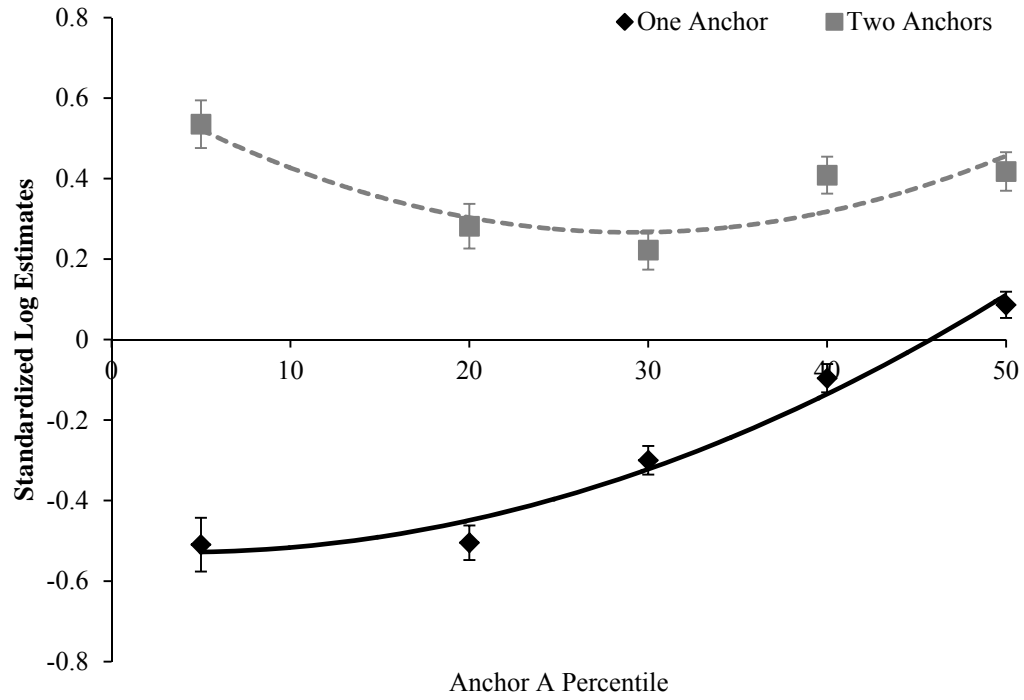


Figure 4. Log estimates of number of bones in the adult human body in study 3. Error bars represent +/- one standard error.

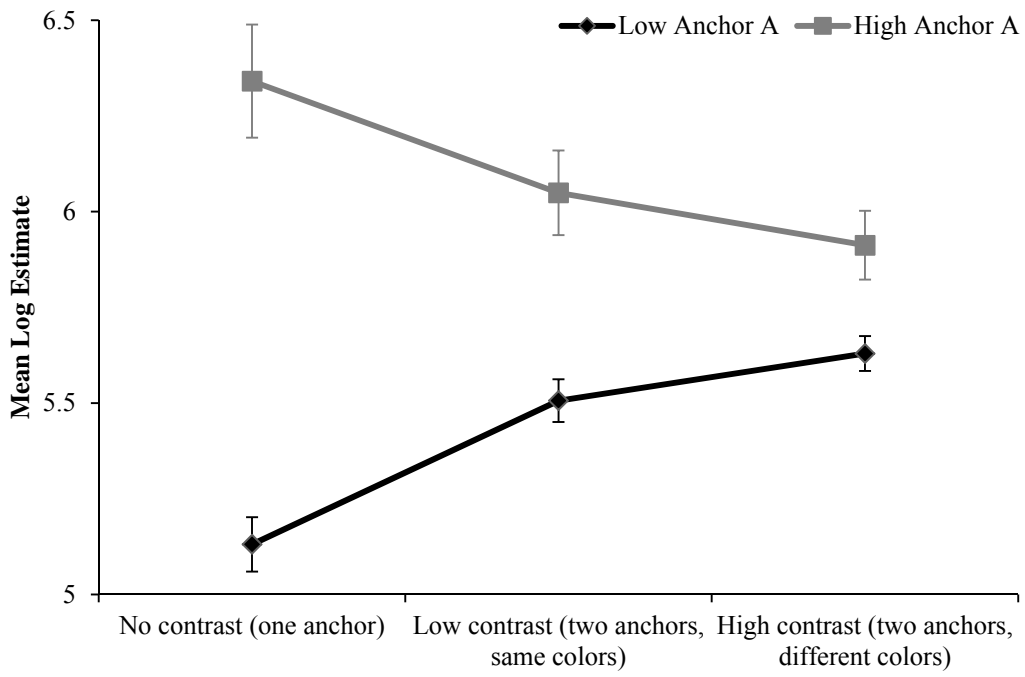


Figure 5. Anchor plausibility ratings from study 3. Error bars represent +/- one standard error.

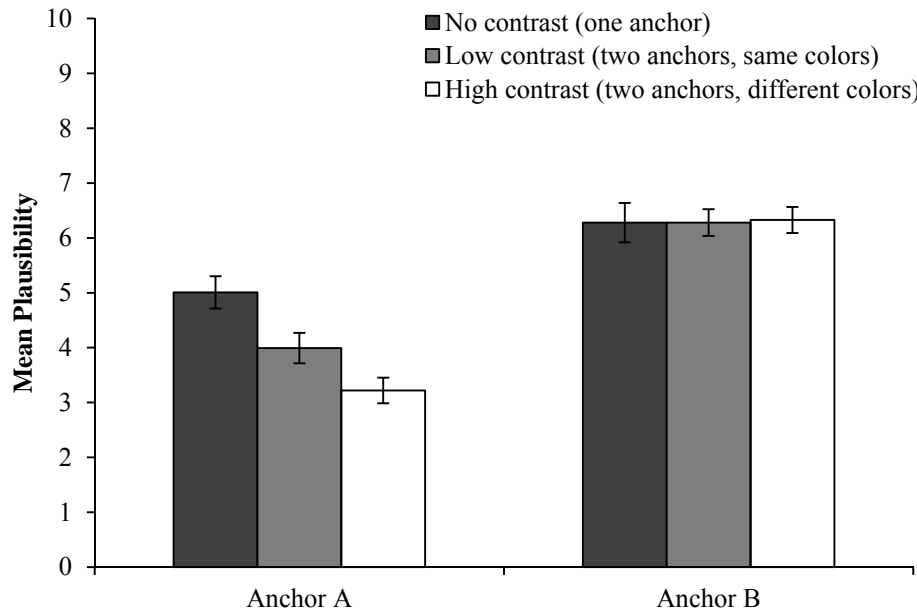


Figure 6. Fitted curves for the final price as a function of starting price, holding other variables constant. The solid black line is for BIN auctions, and the dotted grey line is for pure auctions.

