

# The Z-Effect: Why Good Is Good, but Better Is Better

Steven C. Huff<sup>1</sup>

<sup>1</sup> Woodbury School of Business, Utah Valley University, Orem, Utah, USA

Correspondence: Steven Huff, Woodbury School of Business, Utah Valley University, 800 W University Pkwy – MS 288, Orem, UT 84058, USA. Tel: 1-801-863-8863. E-mail: huff@uvu.edu

Received: January 13, 2014      Accepted: February 7, 2014      Online Published: May 21, 2014

doi:10.5539/ijms.v6n3p1      URL: <http://dx.doi.org/10.5539/ijms.v6n3p1>

## Abstract

Most preference construction research studies the response mode of choice. While such research is important, relatively little preference construction research has addressed the implications of constructing willingness to pay. Understanding willingness to pay is important for pricing because choice does not necessarily produce the same results or insights as willingness to pay. This research begins to extend the current literature on the construction of willingness to pay by investigating how it is influenced by the dispersion of quality in product menus. Two experiments demonstrate that willingness to pay is influenced by relative quality (i.e., an alternative's quality relative to other alternatives in the menu). Specifically, these two experiments demonstrate that willingness to pay for an alternative in a menu can be manipulated without changing the objective quality of those alternatives because willingness to pay is correlated with an alternative's quality z-score. This result is an artifact of the difficulty of translating psychological values (preferences) into numerical values (willingness to pay) combined with the comparative nature of the menu context.

**Keywords:** willingness to pay, preference construction, relative quality, quality dispersion

## 1. Introduction

### 1.1 Problem Introduction

When purchasing an iPad at Apple.com, shoppers are faced with three levels of quality that vary by the amount of internal memory installed—16 GB, 32 GB, and 64 GB. A wealth of research on preference construction can provide important insights into which level of memory is likely to be *chosen*. However, what can the research to date say about how much shoppers will be *willing to pay* for each of these options? Furthermore, if Apple offered instead a menu of 16 GB, 48 GB, and 64 GB, or a menu of 16 GB, 24 GB, 48 GB, and 64 GB, would these different menu configurations produce different levels of willingness to pay for the 16 GB and 64 GB configurations that are present in all three menus? This is the question investigated here.

### 1.2 Relevant Scholarship on Preference Construction

Much of the research on preference construction—the idea that people construct their preferences when faced with a choice rather than constantly maintaining a master list of preferences in memory (Bettman, Luce, & Payne, 1998)—focuses on the response mode of choice. For example, it is the main dependent variable in many demonstrations of preference reversals (Grether & Plott, 1979; Lichtenstein & Slovic, 1971; Tversky, Slovic, & Kahneman, 1990) and in a large literature on “context effects,” including the similarity effect (Tversky, 1972), the attraction effect or asymmetric dominance (Huber, Payne, & Puto, 1982), the compromise effect (Simonson, 1989), and tradeoff contrast and extremeness aversion (Simonson & Tversky, 1992) (See Heath & Chatterjee (1995) for a listing of studies up to that point).

Some work on preference construction has also investigated other response modes such as willingness to pay (Birnbaum & Sutton, 1992; Cox & Grether, 1996; Lichtenstein & Slovic, 1971), but the amount of preference construction research studying this response mode is significantly less than that of choice (one major exception here is the significant literature studying willingness to pay for public goods, otherwise known as contingent valuation (Venkatachalam, 2004)).

More research is needed to understand how preference construction influences estimates of willingness to pay. This is true for a number of reasons: first, choices do not necessarily produce the same results as willingness to pay judgments do, because judgments do not necessarily require the same level of commitment as choice

(Ganzach, 1995), and because attribute weights can be enhanced when the attribute is compatible with the mode of response (Slovic, Griffin, & Tversky, 1990). Second, the discrete nature of choice makes it difficult to observe the underlying value function needed by firms to price their products. Knowing a person's reservation price for a product is crucial to pricing decisions; thus, knowing how reservation prices are influenced by context (i.e., how they are constructed) is also crucial. Third, willingness to pay can measure not only the preference ranking of options but also the preference strength (e.g., how much higher a second ranked choice sits above a third ranked choice). And fourth, ultimately measuring choice captures only a part of the decision process associated with a purchase decision. In the absence of price, choice theories to date suggest a chooser might engage in a variety of qualitative choice strategies and heuristics including satisficing (Simon, 1955), lexico-graphic semi-order (Tversky, 1969), elimination-by-aspects (Tversky, 1972), the context effects mentioned previously, and others. But as soon as prices are involved (implying a purchase situation), people must consider whether they are willing to pay the posted price of a product and they may also need to compare the differences in attribute levels between products and consider whether such differences are worth paying the price difference. This requires mapping psychological values (i.e., preferences) to numerical values (i.e., currency) before a final choice can be made. Thus, without understanding willingness to pay, one can never completely understand the choices made in a purchase situation.

### *1.3 The Purpose of This Research*

The purpose of this paper is to begin to address this deficit by investigating how the dispersion of quality in a menu influences willingness to pay for any given item in the menu. Specifically, the dispersion of quality is operationalized by its standard deviation, which in turn influences both the relative quality of a given alternative and ultimately the willingness to pay for that alternative. What this means in the context of the iPad example is that changing the quality of the middle iPad option (i.e., from 32 GB to 48 GB) and/or adding another alternative to the menu (i.e., adding a 24 GB model), will change the dispersion of quality in the menu and consequently the relative quality in the menu. If quality dispersion influences willingness to pay, such changes should produce different levels of willingness to pay for the 16 GB and 64 GB iPads across these three menu configurations despite their constant levels of objective quality.

Two experiments demonstrate these effects. In these experiments, the dispersion of quality in a menu is explicitly varied by either manipulating the objective quality of some alternatives in the menu (Experiments 1 and 2), or by manipulating the number of alternatives in the menu (Experiment 2). The results of these experiments show that these manipulations of dispersion influence willingness to pay for target alternatives in the menu that remain constant in objective quality across experimental conditions. Specifically, these experiments show that willingness to pay is correlated to "standardized quality" (i.e., the product quality's z-score) in addition to objective quality. This relationship between willingness to pay and relative quality in turn produces two interesting artifacts. First, raising the average quality in the menu can actually lower the average willingness to pay for the items in the menu. Second, adding alternatives to the middle of the menu (when sorted by quality) can increase the willingness to pay range associated with the menu (i.e., willingness to pay for the highest quality alternative rises while willingness to pay for the lowest alternative drops).

## **2. Theoretical Background**

### *2.1 Willingness to Pay as an Uncertain Judgment Is Subject to Anchoring*

Current literature establishes that willingness to pay is a judgment made under uncertainty and is therefore susceptible to contextual influences. For example, willingness to pay has been shown to be influenced by a variety of anchor values including incidental prices (Adaval & Wyer, 2011; Nunes & Boatwright, 2004), extreme within-category prices (Krishna, Wagner, Yoon, & Adaval, 2006), and social security numbers (Ariely, Loewenstein, & Prelec, 2003; Simonson & Drolet, 2004); such anchoring effects should only occur when judgments are made under uncertainty (Epley & Gilovich, 2004, 2006; Jacowitz & Kahneman, 1995; Tversky & Kahneman, 1974). Under the insufficient adjustment explanation of anchoring (Epley & Gilovich, 2001, 2004, and 2006), anchoring occurs because adjustments away from the anchor value cease once the first plausible estimate is reached. Since there is a range of plausible values that surrounds the target value, terminating the adjustment process as soon as one reaches a plausible value biases the estimate toward the anchor. This explanation is consistent with the idea that willingness to pay estimates are made under uncertainty because willingness to pay has been shown to be more accurately modeled as a range rather than a single value (Venkatesh & Chatterjee, 2007).

One reason willingness to pay judgments are associated with significant uncertainty is because people find it difficult to translate psychological values into numerical values (Lichtenstein & Slovic, 2006). Ariely et al.

(2003) provide evidence of this when they show that willingness to pay for a good is susceptible to anchoring effects even after a person experiences that good. Specifically, the authors play annoying sounds for participants and ask how much they are willing to accept (the complement of willingness to pay) to listen to different lengths of that sound. Even after experiencing the sound and thus having perfect knowledge of what they were exchanging for cash, they were still influenced by anchoring effects. In another study, Ariely, Loewenstein, and Prelec (2006) show that in some circumstances, even the valence of psychological values can be manipulated. Specifically, they show that depending on the question asked, experimental groups either required payment or were willing to pay in order to listen to a professor with little oratory or performance skill read poetry aloud. This idea can also be seen when Hsee and Rottenstreich (2004) asked people how much they would pay for CDs. By priming study participants (through a questionnaire) to generate a valuation through either calculation or feeling, they were able to show significantly different value functions for sets of Madonna CDs of varying sizes.

### *2.2 Willingness to Pay as a Comparative Judgment*

Although significant difficulty is associated with determining the initial willingness to pay estimate for an item in a menu, once that value has been determined (and is thus available in working memory), it is available to influence subsequent willingness to pay estimates of other items in the menu. These additional estimates should vary as the product quality in the menu varies. But according to the evaluability hypothesis (Hsee, 1996), an attribute is hard to evaluate independently when the evaluator does not know how good a given value on the attribute is without comparisons. This can be the case with willingness to pay estimates, because of the difficulty of translating even differences in psychological values into differences in monetary values (e.g., how much more is one willing to pay for a 10% increase in quality). Consequently, in a shopping situation where in order to make a choice people must estimate their willingness to pay for multiple alternatives in a menu, this difficulty increases the strength and likelihood of comparative judgments (Hsee & Leclerc, 1998). This idea can be observed in a variety of research. For example, Ariely et al. (2003) show that people have arbitrary fundamental values for goods, but once the value for a good is established (or “imprinted” as they phrase it), preferences tend to be ordered and coherent (which comes from comparative judgments). Furthermore, Leclerc, Hsee, and Nunes (2005) show that people make judgments relative to the scope of the judgment task—a phenomenon they call “narrow focus”—and that the same stimuli can be judged very differently depending on the context. It has also been shown that comparative evaluations can even reverse the judgments made from individual evaluations (Hsee, 1996; Hsee, 1998; Hsee & Leclerc, 1998; Hsee, Loewenstein, Blount, & Bazerman, 1999; Hsee & Zhang, 2004). And finally, Tversky and Simonson’s (1993) model of comparative judgment suggests that the value function is relative to background choices and reference points.

While comparisons between items within the menu are likely, comparisons to items not in the menu are not likely. This is one of the central ideas of narrow focus (Leclerc et al., 2005) but can be found in other work as well. Kahneman and Miller’s (1986) norm theory suggests that people have a propensity to make comparisons only within category. There is evidence that people engage in choice bracketing (Read, Loewenstein, & Rabin, 1999; Read & Loewenstein, 1995), which refers to how people group individual choices into sets. Generally, it is better to group into large or broad brackets (i.e., consider alternatives both inside and outside a menu), but it has been shown that people tend to choose narrow brackets (i.e., consider only the alternatives within the menu). This also tends to be the case even when researchers don’t explicitly highlight the category participants might use (Biernat, Manis, & Nelson, 1991). Consequently, the menu context should play a significant role in determining willingness to pay, and different menu configurations should produce different willingness-to-pay estimates even when objective quality for the items in the menu remain constant across menu configurations.

### *2.3 Willingness to Pay Judgments of Alternatives in a Menu*

The fact that willingness-to-pay estimates will be based on comparative judgments of quality in addition to absolute judgments of quality has important implications for what the resulting estimates will be. This can be illustrated by returning to the iPad example. If the estimate for the 64 GB iPad is based on an absolute judgment of quality (meaning without comparison, but based solely on the quality level, meaning the amount of memory), the quality of the 32GB iPad should not influence the willingness to pay for the 64 GB model. However, if willingness to pay for the 64 GB iPad is based on a comparative judgment of quality (comparing the quality of the 64 GB iPad to the 32 GB iPad), then the difference in quality between the two products will matter. Changing the amount of memory in the 32 GB iPad to 48 GB and thereby raising its quality in a menu of iPads will increase the average quality of the menu, but this will reduce the relative quality of the 64 GB iPad (it is now only 16 GB better than the next alternative instead of 32 GB better in the case of the 32 GB iPad) and it will reduce the relative quality of the 16 GB iPad by increasing the distance between it and the next best alternative. Two different relative positions for the middle alternative can produce two different relative

valuations for the top and bottom alternatives and in turn produce two different estimates of willingness to pay, despite the fact that the objective quality of the highest and lowest quality products is unchanged. As a result, in a variety of scenarios where willingness to pay for products in a menu is determined, comparative judgments will often lead to significant differences in willingness to pay from the case where absolute judgments are made.

In large menus, the type of comparison performed is also important. Previous work suggests that a target product might be compared to common measures of central tendency and dispersion such as the mean, median, range, and standard deviation of menu quality. Anderson's (1981, 1982) information integration theory employs a comparison level (like average quality), as does the idea of reference prices in price judgments (Winer, 1986). Using the average quality as a comparison level provides information about the central tendency of the quality in the menu, but it does not provide any information about the spread of quality. This is also important to consider because it relates to the idea that people compare only to the other items in the menu, not necessarily outside the menu. Essentially, this is asking the question of what the scale for comparison is. One scaling method proposed by Parducci (1974) and employed by Bhargava, Kim, and Srivastava (2000) is the range of quality. This measure of spread serves Bhargava et al. (2000) well in their model of comparative judgment because they compare their model to many context effects studies where menus are small—choices between pairs and triplets is the norm. In a choice between a pair of options, the mean and range almost fully describe the distribution of the choice set, and with triplets the characterization is still close. However, in larger menus, there can be significant variety within many menu configurations without altering the range (or even the mean). For this reason, a better scaling factor in willingness to pay problems is the standard deviation of quality in the menu. In combination with the mean, the standard deviation provides information about how far from average quality an alternative is and how different this difference is with respect to other alternatives in the menu. Conveniently, these two statistics can be combined to calculate the z-score of quality—the number of standard deviations a given level of quality is from the average quality—as a way to operationalize the relative quality of an alternative in a menu.

#### *2.4 Summarizing the Implications of Theory*

The summary of this discussion is that when determining willingness to pay for products in a menu, people will be influenced by the relative quality of the product (which can be operationalized by its quality z-score). A more formal statement of the hypothesis is that a product's quality z-score will be positively correlated with willingness-to-pay estimates.

### **3. Method**

#### *3.1 General Overview of the Experiments and General Hypotheses*

Two experiments are conducted with a similar design where a questionnaire is administered through Qualtrics survey software on a computer. The instructions state that multiple hypothetical purchase situations will be presented. Participants are instructed to read each scenario carefully and imagine that they were actually faced with the situation described; they then are presented with different purchase situations. In the description of each situation, participants are told that multiple alternatives are available and that these alternatives are “essentially the same” on all attributes except for one differentiating attribute. Care was taken to ensure that in each purchase situation, the differentiating attribute was a significant driver of quality in the product category. Note that this produces a purchase situation that is not much different from the idea that people summarize all quality dimensions into one meta-attribute of quality (Green & Srinivasan, 1978; Wright, 1975). For example, one situation describes a scenario where the participant must purchase a flight from San Francisco to New York from one of the airlines listed. In this scenario, the differentiating attribute is the average on-time arrival rate for each airline.

Below the situation description, the questionnaire presents a menu of alternatives. In the menu, no brand names are used but rather entries are labeled “A”, “B”, “C”, etc. The menu then presents a measurement of how well the alternative performs on the differentiating attribute. For example, in the flight menu, on-time arrival rates are listed next to each airline and range from 87% to 59%. Each menu is sorted top-down from highest quality to lowest quality, where quality is determined solely by the differentiating attribute since all other attributes are considered to be essentially the same. In each study, the objective quality (e.g., on-time arrival rates in the flight scenario) of the target alternatives is constant across experimental conditions. Relative quality is manipulated by changing the quality levels of non-target alternatives in the menu or by changing the number of alternatives in the menu. This design allows the analysis to separate the influence of objective quality on valuation from the influence of the relative quality differences caused by changes in the dispersion of quality in the menu.

The purchase scenario ends by eliciting willingness-to-pay responses for each of the alternatives in the menu. The response text-boxes for the alternatives are also listed from high to low quality, but respondents may fill out

these boxes in any order they choose.

### 3.2 Experiment 1

#### 3.2.1 Participants and Design

Two-hundred and sixty-four undergraduates from a large university in the United States were recruited for the study and given course credit to participate. Experiment 1 employed a single factor, four-level design (dispersion type: Control vs. High-Mean vs. Low-Mean vs. Alternate-Variance).

#### 3.2.2 Procedure

Participants came to a computer lab and were given a session ID number—which unbeknownst to them randomly assigned them to an experimental condition—and directed to a computer terminal equipped with dividers so that no participant could see the stimuli or responses of another participant. The general instructions informed participants that they would be faced with hypothetical purchase situations, and that they should imagine themselves actually facing those situations in reality when determining the maximum they would be willing to pay for the products and services described. Participants then were presented with four different purchase situations related to 1) booking a flight from San Francisco to New York, 2) purchasing cell phone service after moving to a new city, 3) buying a digital camera, and 4) purchasing a laptop. For these categories, the differentiating attributes were respectively on-time arrival rates for each airline, percent area-coverage for each cell phone provider, the number of megapixels in each digital camera, and battery life for laptops.

After each description, a menu of five alternatives was presented. The quality for the five airline menu alternatives across the four conditions is listed in Table 1. Note that the Z-scores listed in the table were not shown to respondents. Menus for each of the other three product categories were similar. Note that all menus were sorted by quality from highest to lowest, and the quality of the highest- and lowest-quality alternatives remained constant across the four experimental conditions (i.e., these were the target alternatives of interest). As can be seen in the table, quality in the menu was distributed uniformly in the Control condition. The High-Mean (Low-Mean) condition differed from the Control condition in that the qualities of the three middle options were all very close to the quality of the highest (lowest) alternative. In the Alternate-Variance condition, the qualities of the three middle alternatives were clustered closely around the mean quality of the menu. In addition to objective quality, Table 1 also shows the quality z-score for each menu alternative, which was not shown to study participants. These z-scores show that relative quality range for the High-Mean (Low-Mean) condition was significantly lower (higher) than the relative quality range of the Control Menus. And the relative quality ranges of the Alternate-Variance condition menus were wider than those of the Control condition.

Table 1. Menu designs for experiment 1

Category and Attribute	Menu Label	Control		High-Mean		Low-Mean		Alt.-Variance	
		Q	Z	Q	Z	Q	Z	Q	Z
Airlines On-Time Record (%)	A	87%	1.27	87%	0.57	87%	1.78	87%	1.41
	B	80%	0.63	86%	0.49	62%	-0.32	74%	0.10
	C	73%	0.00	85%	0.40	61%	-0.40	73%	0.00
	D	66%	-0.63	84%	0.32	60%	-0.49	72%	-0.10
	E	59%	-1.27	59%	-1.78	59%	-0.57	59%	-1.41

After presenting the menu, an average price for the product category represented in the menu was given. This was done to reduce the variance in the willingness to pay responses. The average price presented for each category was \$400 for airlines, \$100 for monthly cell service, \$360 for digital cameras, and \$1,500 for laptops. These prices were accurate averages of real-world prices at the time of the experiment. After stating the average price, participants were asked the maximum they would pay for each of the five alternatives in the menu.

#### 3.2.3 Data

The data included 5,280 willingness to pay responses across the 1,056 instances of the four purchase situations; each participant responded to four purchase situations, so this represents 264 participants. Participants were divided across the four experimental conditions so that there was 78, 57, 53, and 76 respondents in the Control, High-Mean, Low-Mean and Alternate-Variance conditions, respectively.

### 3.2.4 Analysis

When performing an analysis of the experimental data, it is important to note that the experimental design collects multiple observations from each study participant, making observations within subjects not independent of each other. Because of this interdependence—where individual observations may be correlated because they came from the same source (i.e., each participant provided multiple willingness to pay responses in each product category)—, the experimental data are analyzed using a mixed-effects model with a random intercept to control for the repeated measures, also called a hierarchical linear model (HLM).

### 3.2.5 Results for Experiment 1

Table 2 contains the HLM estimates from Equation 3. The intercept term represents the average willingness to pay for the full set of alternatives in the menu across all conditions and *Quality* represents how much participants were willing to pay for a 1% increase in objective quality. For example, participants were willing to pay on average \$374.58 for a flight and \$1.67 for a 1% increase in the on-time arrival rate of the flight. The variable of interest, *Z*, is highly significant in all four product categories at least the 1% level. In terms of magnitude, for a standard deviation increase in relative quality, participants were willing to pay an additional \$28.94 in the airline category, \$9.14 in the cell phone category, \$24.23 in the digital camera category, and \$49.89 in the laptop category. These values are also substantial. Comparing these amounts to the average willingness to pay across the menu for each of the product categories shows that participants were willing to pay a premium of 7.7% for a standard deviation increase in relative quality in the airline category. In the cell phone, digital camera, and laptop categories, these premiums were 10.5%, 7.7%, and 5.8%, respectively. If these percentage increases were included in prices, their effects would be magnified in profits, especially in mature markets when margins are tighter.

Table 2. HLM coefficient estimates for experiment 1

Category	Airlines	Cell Service	Dig. Cameras	Laptops
Mean Category Price Given	\$400	\$100	\$360	\$1,500
Mean WTP in each Menu	374.58 ***	86.91 ***	316.01 ***	857.53 ***
Objective Quality <sup>†</sup>	1.67 *	0.53 **	12.56 **	46.59 ***
Rank	-10.66 ***	-2.21 ***	-7.76 ***	-21.09 ***
Relative Quality ( <i>Z</i> )	<b>28.94</b> **	<b>9.14</b> ***	<b>24.23</b> **	<b>49.89</b> **

Note. \*, \*\*, \*\*\* represent statistical significance at the 5%, 1%, and 0.1% levels respectively. † WTP increase for a unit increase in objective quality.

### 3.2.6 Discussion

These results provide evidence that in a menu context, the relative quality of a product influences willingness to pay. Despite objective quality for the highest and lowest quality alternatives in each menu remaining constant across conditions, and despite the HLM controlling for objective quality and rank effects, willingness to pay for the highest and lowest quality menu alternatives changed significantly due to variation in the quality of the alternatives in the middle of the menu.

While these results are significant, Experiment 1 has an important limitation. In order to control the inherent noise common in willingness to pay estimates, Experiment 1 provides participants with an average category price in each purchase scenario. This isn't a problem when comparing the Alternative-Variance condition to the Control condition because these two conditions have the same average menu quality. However, when comparing the High- or Low-Mean conditions to the Control condition, the results may be at least partially driven by the fact that the same average category price is used in each experimental condition. If participants use the average quality of the menu to infer what the average quality of the overall category is, and if the average category price for each of two conditions is the same, but the average menu quality across these conditions is different, then participants' estimates of their willingness to pay may be influenced by this difference.

For example, suppose two menus exist, each with two items. The first menu contains items with qualities of 100 and 80. The second contains items with qualities of 100 and 20. This produces average menu qualities of 90 and 60, respectively. Suppose further that the average category price for both menus is \$50, that people are willing to pay an additional dollar for a unit increase in quality, and that people only care about objective quality. Remember that people have difficulty translating psychological values into numerical values, so they are

susceptible to the anchoring effect of the stated average category price. Inferring average category quality from average menu quality suggests that people should be willing to pay \$50 for an item in the first menu with an average quality of 90 and that they would be willing to pay \$60 for the highest quality item in that menu. In comparison, these same rules suggest that they would be willing to pay \$90 for the highest quality item in the second menu (\$50 for a quality of 60 plus \$40 for a 40 percentage point increase in quality). Thus, raising the average quality of the menu may lower the willingness to pay for the highest item in the menu, even when relative quality does not influence willingness to pay. This issue is addressed in Experiment 2 by eliminating the category reference price in the scenario descriptions.

### 3.3 Experiment 2

#### 3.3.1 Participants and Design

1,978 people in the United States were recruited for the study through Amazon's Mechanical Turk website and paid a modest sum (\$0.50 per questionnaire) to participate. Experiment 2 employed a single factor, ten-level design (dispersion type: Four-Uniform vs. Four-High vs. Four-Low vs. Four-Wide vs. Four-Narrow vs. Three-Wide vs. Three-Narrow vs. Two-Wide vs. Two-Narrow vs. One; in the preceding condition names, the number refers to the number of alternatives in the menu, and the second term describes the quality distribution in the menu).

#### 3.3.2 Procedure

Experiment 2 was similar to Experiment 1 except for a few differences. First, the most important difference is that participants were not given an average category price before willingness to pay responses were elicited. This change fixes the limitation from Experiment 1, but as a consequence, a much higher number of observations is required to obtain enough power to produce significant differences in the appropriate statistical tests. Second, participants were recruited through Amazon's Mechanical Turk website and paid \$0.50 per questionnaire. This made it possible to recruit enough participants to overcome the high level of "noise" inherent in willingness to pay responses when no average category price was given in each purchase scenario. Third, only two categories were used (instead of four) and both of these were different from any of the categories in Experiment 1. Specifically, the first category was digital medical dictionary apps for smart phones and the second was headlamp-style flashlights. In the first category, the defining attribute of quality was the number of entries in the dictionary (a similar product category, music dictionaries, was used in Hsee (1996)). In the second category, the defining attribute of quality was the brightness of the headlamp measured by how many feet it cast its light. Using a only two categories made the Mechanical Turk questionnaire short enough that a payment of \$0.50 per completed questionnaire was a competitive rate for Mechanical Turk workers. Fourth and finally, a wider range of conditions were included in the study to capture more variation in dispersion and menu size. The menu alternatives in each of the two categories across all ten experimental conditions are shown in Table 3 along with the quality z-scores that were not shown in the experiment.

Table 3. Menu designs in each condition of experiment 2

Category	Item	Condition									
		1	2	3	4	5	6	7	8	9	10
Medical Dictionary App (entries in 000s)	A	21	21	21	21	21	21	21	21	21	21
	B	15	20	5	20	13	12	15	3	15	
	C	9	19	4	4	11	3	9			
	D	3	3	3	3	3					
Head-lamps (feet)	A	100	100	100	100	100	100	100	100	100	100
	B	70	90	30	90	60	55	70	10	70	
	C	40	80	20	20	50	10	40			
	D	10	10	10	10	10					

To increase the reliability of the data collected through Mechanical Turk in Experiment 2, only U.S.-based Turk "workers" who had completed at least 500 previous tasks with approval ratings of 95% or higher were used for the study, and information check questions were included to weed out any respondents who didn't carefully read experimental instructions (an indicator of low effort).

### 3.3.3 Data

The raw data included 12,284 willingness to pay responses across the 3,955 completed instances of the two purchase situations. Before analysis, the raw data were edited to remove responses that indicated a lack of effort in any way (an important precaution when working with Mechanical Turk data). First, any responses that contained a willingness to pay response that changed in the opposite direction of the change in quality were dropped. For example, if on time arrival rate dropped, but the respondent's willingness to pay increased, even if for only one alternative in the menu, then that respondent's responses were dropped for the category in question. This resulted in the removal of 33 purchase scenario instances. Second, any responses were dropped where the respondent failed the instruction check. This resulted in the removal of responses for 338 individuals (17% of the total set of respondents). The resulting number of respondents in each condition after these data were dropped ranged from 151 to 173.

### 3.3.4 Analysis

Just as in Experiment 1, Experiment 2 collects multiple observations from each study participant, making observations within subjects not independent of each other. Consequently, the data were again analyzed using a hierarchical linear model.

### 3.3.5 Results

Table 4 contains the HLM estimates from Equation 4 when applied to the data collected from Experiment 2. The results show that participants were willing to pay on average \$77.29 for the medical dictionary app and (interestingly) less than one cent for a 1% increase in the number of entries. For headlamps, participants were willing to pay \$19.00 on average and \$0.25 for an additional foot of light casting distance. The variable of interest,  $Z$ , is significantly different from zero in the dictionary category at the 5% level, but not in the headlamp category.

Table 4. HLM coefficient estimates for experiment 2

	Dictionaries		Headlamps	
Mean WTP in each Menu	77.30	***	19.00	***
Objective Quality <sup>†</sup>	0.003	***	0.25	***
Rank	-6.64	***	-3.75	***
Menu Size	-0.90		2.88	***
Relative Quality ( $Z$ )	<b>10.27</b>	*	<b>1.34</b>	

Note. \*, \*\*, \*\*\* represent statistical significance at the 5%, 1%, and 0.1% levels respectively. † WTP increase for a unit increase in objective quality.

This lack of significance is puzzling until one examines estimates of the variance inflation factors (VIFs) from the analysis. An analysis of the VIFs suggests that the quality z-scores are too correlated with the other variables in the model, meaning the data exhibit unacceptable levels of multicollinearity (VIF estimates for  $Z$  are 12.77 and 15.98 for the dictionary and headlamp categories, respectively).

The multicollinearity problem can be solved by analyzing each rank in the menu individually across the experimental conditions. When analyzing each menu rank individually, hierarchical models are no longer necessary because only one observation per respondent is included. Furthermore, the *Rank* variable is no longer necessary since rank is constant within each regression. The cleanest test of whether the z-score of quality influences willingness to pay will be to look at the highest quality alternative across all ten conditions (Rank 1), and the lowest quality alternative across the first five conditions (Rank 4) where the fourth item exists and has the same quality for each condition. When performing the analysis this way, the variable  $Q$  is no longer needed in the regression because it is also constant across all conditions in both regressions. Eliminating rank and quality from the regressions provides tighter control. Consequently, the only variable needed in addition to  $Z$  is *MenuSize*, and it is only needed in the first regression. Estimating this new regression using ordinary least-squares produces the results found in Table 5.

Table 5. OLS coefficient estimates for ranks 1 and 4 in experiment 2

	Dictionaries		Headlamps	
<b>Rank 1 Regression</b>				
WTP for Rank 1	113.53	***	34.74	***
Relative Quality (Z)	<b>40.60</b>	***	<b>12.35</b>	***
<b>Menu Size</b>	-10.09	*	-0.31	
<b>Rank 4 Regression</b>				
WTP for Rank 4	-100.61	***	34.41	***
Relative Quality (Z)	<b>19.73</b>	*	<b>11.64</b>	***

Note. \*, \*\*, \*\*\* represent statistical significance at the 5%, 1%, and 0.1% levels respectively.

Table 5 reveals that estimates of willingness to pay for a standard deviation increase in relative quality range from \$19.73 to \$40.60 in the dictionary category and from \$11.64 to \$12.35 in the headlamp category. In terms of magnitude relative to mean willingness to pay for the rank in question, these amounts are significant.

### 3.3.6 Discussion

Experiment 2 provides more evidence that willingness to pay is influenced by relative quality. Because Experiment 2 does not provide any category prices, respondents can't anchor on those prices and are free to provide any response they desire. This fixes the limitation of Experiment 1 and interestingly, when no anchor values are provided, the effect sizes are even bigger than before. In the Rank 1 regressions, participants were willing to pay premiums of 35.8% and 35.5% for a standard deviation increase in relative quality for medical dictionaries and headlamps, respectively. In the Rank 4 regressions, they were willing to pay premiums of 19.6% and 33.8%, respectively. This suggests that the presence of mean category prices was in fact attenuating the effect, rather than artificially magnifying the effect.

## 4. General Discussion

### 4.1 Practical Relevance

The importance of the influence of relative quality can be seen both online and off. Internet retailer menus can change with every consumer and even every mouse click. Manufacturers are always in the process of introducing new products, improving current products, and withdrawing old products. Even if the products that enter the effective menu have a low probability of purchase, the fact that willingness to pay is influenced by standardized quality (i.e., the dispersion of quality in the menu) suggests they could have a significant impact on willingness to pay for every product in the category.

Every item in a menu can potentially influence a consumer's willingness to pay for the item she actually purchases. These effects will be stronger when menus are relatively small, but note that they could be just as present in consumer-created assortments (consideration sets) as in firm-provided menus. These findings imply that manufacturers must consider relative quality as they consider additions to, deletions from, or modifications of products within their product lines. They also suggest that manufacturers and retailers should carefully consider what assortment of products they wish to offer within a given category, since every product in a menu may influence the relative quality of the other products in the menu.

### 4.2 Future Research

This research is limited to only demonstrating the existence of these effects. It does not attempt to fully characterize the relative component of the value function proposed by Tversky and Simonson (1993). There are directions of future research that naturally spring from this work. The most interesting direction is the combination of the z-effect with time and memory; how does relative quality manifest when people sample the menu sequentially rather than simultaneously? This may apply when the menu of most importance is the consideration set that is mentally constructed during the shopping process. Such a process introduces time between valuations and makes memory a salient factor in the evaluation process. A related question is whether some conditions exist where consumers' perception of relative quality includes qualities of products not in the menu, but stored in consumers' memories

## 5. Conclusion

This article investigates the effect of the dispersion of quality in a product menu on consumer preferences for the products found in that menu. The results of two experiments show that preferences are influenced by the dispersion of quality in the menu; in other words, the results demonstrate that willingness to pay is influenced by

relative quality (which can be operationalized by quality z-scores) in addition to absolute quality. This paper contributes to the current literature on context effects and menu dependence by suggesting a functional form that characterizes the nature of changes that result from the menu (context). It extends this literature by investigating larger menus through a continuous measure of preference.

## 6. Acknowledgements

This article benefited from comments by Bill Ross (University of Connecticut, Darron Billeter (Brigham Young University), Paul Dishman (Utah Valley University), and others who attend presentations of this work or reviewed this work in the review process of publication.

## References

- Anderson, N. H. (1981). *Foundations of integration information theory*. New York: Academic Press.
- Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.
- Adaval, R., & Wyer, R. S. (2011). Conscious and nonconscious comparisons with price anchors: Effects on willingness to pay for related and unrelated products. *Journal of Marketing Research*. <http://dx.doi.org/10.1509/jmkr.48.2.355>
- Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics*. <http://dx.doi.org/10.1162/00335530360535153>
- Ariely, D., Loewenstein, G., & Prelec, D. (2006). Tom Sawyer and the construction of value. *Journal of Economic Behavior and Organization*, 60(1), 1–10. <http://dx.doi.org/10.1016/j.jebo.2004.10.003>
- Benartzi, S., & Thaler, R. H. (2002). How much is investor autonomy worth? *The Journal of Finance*, 57(4), 1593–1616. <http://dx.doi.org/10.1111/1540-6261.00472>
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25(3), 187–217. <http://dx.doi.org/10.1086/209535>
- Bhargava, M., Kim, J., & Srivastava, R. (2000). Explaining context effects on choice using a model of comparative judgment. *Journal of Consumer Psychology*, 9(3), 167–177. [http://dx.doi.org/10.1207/S15327663JCP0903\\_4](http://dx.doi.org/10.1207/S15327663JCP0903_4)
- Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, 60(4), 485. <http://dx.doi.org/10.1037/0022-3514.60.4.485>
- Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior And Human Decision Processes*, 52(2), 183–215. [http://dx.doi.org/10.1016/0749-5978\(92\)90035-6](http://dx.doi.org/10.1016/0749-5978(92)90035-6)
- Chernev, A. (2003). When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of Consumer Research*, 30(2), 170–183. <http://dx.doi.org/10.1086/376808>
- Cox, J. C., & Grether, D. M. (1996). The preference reversal phenomenon: Response mode, markets and incentives. *Economic Theory*, 7(3), 381–405. <http://www.jstor.org/stable/25054933>
- Diehl, K. (2005). When two rights make a wrong: Searching too much in ordered environments. *Journal of Marketing Research*, 313–322. <http://dx.doi.org/10.1509/jmkr.2005.42.3.313>
- Diehl, K., Kornish, L., & Lynch, J. (2003). Smart agents: When lower search costs for quality information increase price sensitivity. *Journal of Consumer Research*, 30(1), 56–71. <http://dx.doi.org/10.1086/374698>
- Diehl, K., & Poynor, C. (2010). Great expectations?! Assortment size, expectations, and satisfaction. *Journal Of Marketing Research*, 47(2), 312–322. <http://dx.doi.org/10.1509/jmkr.47.2.312>
- Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, 12(5), 391–396. <http://dx.doi.org/10.1111/1467-9280.00372>
- Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality and Social Psychology Bulletin*, 30(4), 447–460. <http://dx.doi.org/10.1177/0146167203261889>
- Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, 17(4), 311–318. <http://dx.doi.org/10.1111/j.1467-9280.2006.01704.x>
- Ganzach, Y. (1995). Attribute scatter and decision outcome: Judgment versus choice. *Organizational Behavior And Human Decision Processes*, 62(1), 113–122. <http://dx.doi.org/10.1006/obhd.1995.1036>
- Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research. *Journal of Consumer Research*,

- 5(September), 103–123. <http://dx.doi.org/10.1086/208721>
- Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69(4), 623–638. <http://www.jstor.org/stable/1808708>
- Heath, T., & Chatterjee, S. (1995). Asymmetric decoy effects on lower-quality versus higher-quality brands: Meta-analytic and experimental evidence. *Journal of Consumer Research*, 22(3), 268–284. <http://dx.doi.org/10.1086/209449>
- Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, 67, 247–257. <http://dx.doi.org/10.1006/obhd.1996.0077>
- Hsee, C. K. (1998). Less is better: when low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, 11(2), 107–121. [http://dx.doi.org/10.1002/\(SICI\)1099-0771\(199806\)11:2<107::AID-BDM292>3.0.CO;2-Y](http://dx.doi.org/10.1002/(SICI)1099-0771(199806)11:2<107::AID-BDM292>3.0.CO;2-Y)
- Hsee, C. K., & Leclerc, F. (1998). Will products look more attractive when presented separately or together? *Journal of Consumer Research*, 25(2), 175–186. <http://dx.doi.org/10.1086/209534>
- Hsee, C., Loewenstein, G., Blount, S., & Bazerman, M. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125(5), 576–590. <http://dx.doi.org/10.1037/0033-2909.125.5.576>
- Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology-General*, 133(1), 23–30. <http://dx.doi.org/10.1037/0096-3445.133.1.23>
- Hsee, C. K., & Zhang, J. (2004). Distinction bias: Misprediction and mischoicedue to joint evaluation. *Journal of Personality and Social Psychology*, 86(5), 680–695. <http://dx.doi.org/10.1037/0022-3514.86.5.680>
- Huber, J., Payne, J., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9(1), 90. <http://dx.doi.org/10.1086/208899>
- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can desire be too much of a good thing? *Journal of Personality and Social Psychology*, 79(6), 995–1006. <http://dx.doi.org/10.1037//0022-3514.79.6.995>
- Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11), 1161–1166. <http://dx.doi.org/10.1177/01461672952111004>
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), 136. <http://dx.doi.org/10.1037/0033-295X.93.2.136>
- Krishna, A., Wagner, M., Yoon, C., & Adaval, R. (2006). Effects of extreme-priced products on consumer reservation prices. *Journal of Consumer Psychology*, 16(2), 176–190. [http://dx.doi.org/10.1207/s15327663jcp1602\\_8](http://dx.doi.org/10.1207/s15327663jcp1602_8)
- Leclerc, F., Hsee, C., & Nunes, J. (2005). Narrow focusing: Why the relative position of a good in its category matters more than it should. *Marketing Science*, 24(2), 194–205. <http://dx.doi.org/10.1287/mksc.1040.0090>
- Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology-General*, 89(1), 46. <http://dx.doi.org/10.1037/h0031207>
- Lichtenstein, S., & Slovic, P. (2006). *The construction of preference*. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511618031>
- Nunes, J. C., & Boatwright, P. (2004). Incidental prices and their effect on willingness to pay. *Journal of Marketing Research*. <http://dx.doi.org/10.1509/jmkr.41.4.457.47014>
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, 72(6), 407–418. <http://dx.doi.org/10.1037/h0022602>
- Parducci, A. (1974). Contextual effects: A range-frequency analysis. *Handbook of Perception*, 2, 127–141.
- Pope, D. G. (2009). Reacting to rankings: evidence from America's best hospitals. *Journal of Health Economics*, 28(6), 1154–1165. <http://dx.doi.org/10.1016/j.jhealeco.2009.08.006>
- Read, D., & Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, 1(1), 34. <http://dx.doi.org/10.1037/1076-898X.1.1.34>

- Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, 19(1), 171–197. <http://dx.doi.org/10.1023/A:1007879411489>
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99. <http://dx.doi.org/10.2307/1884852>
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16(2), 158. <http://dx.doi.org/10.1086/209205>
- Simonson, I., & Drolet, A. (2004). Anchoring effects on consumers' willingness-to-pay and willingness-to-accept. *Journal of Consumer Research*, 31(3), 681–690. <http://dx.doi.org/10.1086/425103>
- Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 29(3), 281–295. <http://dx.doi.org/10.2307/3172740>
- Slovic, P. G., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, 5–27.
- Suri, S., & Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS One*, 6(3), e16836. <http://dx.doi.org/10.1371/journal.pone.0016836>
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1), 31. <http://dx.doi.org/10.1037/h0026750>
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281–299. <http://dx.doi.org/10.1037/h0032955>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <http://dx.doi.org/10.1126/science.185.4157.1124>
- Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, 39(10), 1179–1189. <http://dx.doi.org/10.1287/mnsc.39.10.1179>
- Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, 204–217. <http://www.jstor.org/stable/2006743>
- Venkatachalam, L. (2004). The contingent valuation method: a review. *Environmental Impact Assessment Review*, 24(1), 89–124. [http://dx.doi.org/10.1016/S0195-9255\(03\)00138-0](http://dx.doi.org/10.1016/S0195-9255(03)00138-0)
- Wang, T., Venkatesh, R., & Chatterjee, R. (2007). Reservation price as a range: An incentive-compatible measurement approach. *Journal of Marketing*, 200–213. <http://dx.doi.org/10.1509/jmkr.44.2.200>
- Winer, R. (1986). A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research*. <http://dx.doi.org/10.1086/209064>
- Wright, P. (1975). Consumer choice strategies: Simplifying versus optimizing. *Journal of Marketing Research*, 12(February), 60–67. <http://dx.doi.org/10.2307/3150659>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).