Does Price Rounding Lead to More Favorable Product Perception?

Mike Fan, Prakhar Maini, Sean Norris, Claudia Ventresca

ABSTRACT

Why is it so often that supermarkets, online retailers, and even car dealerships price their products at $59.99, $29.99, or $9,999.99? More specifically, why do all those prices end in ".99"? Why not just price them at $60, $30, and $10,000? Researchers have pondered why this could be so and many have settled on what is known as the "left digit effect", which reflects how the left-most digit disproportionately affects our perception of price. While there have been a number of experiments that have studied the "left digit" effect, an important question that remains unclear is whether round pricing has an impact on consumers' perception of the product's quality? In this study, we conducted a randomized controlled experiment to study the effect of round versus non-round pricing on the perception of common consumer goods' quality under hedonic purchasing contexts.

In order to mitigate research bias, a pretest was conducted with friends and families to ascertain that the products and the purchasing contexts were viewed in accordance with their intended category. After revising the products and purchasing contexts based on the results from the pretest, we created surveys that included pictures, a description of the product's features, and product prices. These surveys were then exposed to subjects on Amazon Mechanical Turk (MTurk). Each subject was asked a set of questions with respect to perceived quality and anticipated satisfaction from use before and after they were exposed to the price (treatment event). We then compared the "before and after" results and found that we were unable to reject the null hypothesis: lack of evidence to suggest that round prices under hedonic purchasing led to a more favorable quality perception and anticipated satisfaction.

BACKGROUND

Pricing has been shown to be one of the most important dimensions that can have a substantial impact on consumers' product evaluation and purchase decisions. Consumers also tend to assume the price of the product to be positively correlated with the quality of the product. Moreover, price may not only affect perceived quality, but can potentially affect the actual satisfaction of owning and using the product (i.e. buyer's remorse). Specifically, a controlled experiment conducted by Robert Schindler and Thomas Kibarian [2] argues that products ending in digits ".99" rather than rounded (i.e., $29.99 versus $30.00) lead to an increase in consumer purchasing. The same said authors also argued in a previous study conducted in 2001 [4] that the use of the".99" rather than the 00 (i.e. rounded) price ending also had negative effects on the quality image of products in advertisements. A more recent study in the *Journal of Consumer Research* by Monica Wadhwa and Kuangjie Zhang [1] found that shoppers deal with pricing information differently when prices feature round numbers, as opposed to non-round ones. For example, when something costs $100, consumers tend to rely on their emotions, whereas when something has an irregular price, such as $98.97, people have to use reason and cognition to decide whether it's an attractive price.

If rounded prices can elicit positive quality perception under certain purchasing conditions and settings, then firms can adapt their prices to the context in which products are sold. For example, online retailers could use historical or in-the-moment behavioral data collected from customers (i.e. based on browsing patterns) that indicate a tendency to make more deliberate (vs. impulsive or time-constrained) decisions and dynamically adapt pricing.

Our research aims to expand the scope of a previous study completed by Monica Wadhwa and Kuangjie Zhang [3]. To do this, we used multiple product and context pairs.

RESEARCH QUESTION

The objective of this study is to test whether people's evaluation of various consumer products would be more favorable when the price is rounded, versus non-rounded. This perception was tested under two consumption goals: feeling/hedonic (i.e., buying a camera for an upcoming family vacation) and feeling/utilitarian (i.e., buying a camera for a class project). Past research suggests that the degree to which a consumer relies on feelings versus thinking depends on the consumption goal associated with the product [2]. When the consumption goal is primarily hedonic driven, people tend to rely more on feeling where in a utilitarian context, people tend to rely more on cognition. Thus, the hypothesis is that when the decision context and motivation for the purchase is primarily driven by feeling/hedonic driven goals, then rounded pricing will lead to a more favorable perception of the product. In contrast, when the purchase motivation is primarily driven by a feeling/utilitarian goal, non-rounded pricing should lead to a more favorable perception of the product. However, the scope of this study focuses on the former rather than the latter; primarily due to the fact that the latter requires a 2x2 factorial design, which was beyond the team's budget at the time this research was conducted.

Hypothesis

Hedonic Purchase Context
- **Null**: Product quality perception is identical when the price is rounded vs. unrounded
- **Alternative**: Product quality perception *is higher* when the price is rounded vs. unrounded

Utilitarian Purchase Context
- **Null**: Product quality perception is identical when the price is rounded vs. unrounded
- **Alternative**: Product quality perception *is higher* when the price is rounded vs. unrounded

EXPERIMENT DESIGN

The experiment is split into two milestones:
1. A **pretest** to ascertain that the product and purchasing contexts are aligned with the viewpoints of the broader population in order to mitigate personal biases

2. The **main experiment** exposes validated product and purchase contexts from the pretest to random survey participants on MTurk to answer a set of questions with respect to perception of quality and anticipated satisfaction of usage

In the following subsections, we clarify the rationale and definition of the experiment and measurement approach.

Study Timeline

Start Pretest	Complete Pretest	Pretest Analysis	Launch Experiment	Close Experiment	Experiment Analysis
3/1	3/9	3/10 - 3/14	3/15	3/29	3/30 - 4/6

Product and Purchase Context Definition

- **Hedonic vs. Utilitarian Purchasing Context**: There are very few products that neatly fit in only one of the buckets across all usage contexts possible. Most often, it is the usage motivation that defines whether the purchase of a given product is to seek more pleasure or simply achieve utility (i.e. buying a camera for a family vacation versus for a class project). We acknowledge this complexity and employed a pre-test to validate the purchase contexts we developed for a given product.

- **Perception of Product Quality Measures**: Product quality is a broad term and usually refers to a collection of attributes. We formalized this definition by using a composite score representing product quality. Specifically, we measured the total score across five dimensions provided by respondents in the test and control groups.

Product Quality Definition

Because our hypothesis is hinged on the pricing scheme's potential effect on the product's perceived quality, we used the following dimensions as proxies to derive a composite quality score. This not only enabled us to use a single numerical score to compare across treatment and control, but also provided the flexibility to decompose the score into its attributes to help identify any unexpected and interesting findings.

Evaluation Dimensions

The evaluation dimensions focused on the four (4) key below and asked these specific questions:

1. **Durability and Reliability**: Do participants perceive that the product will remain functional, without requiring excessive maintenance over a five year period?
2. **Performance**: Do participants perceive that the performance of the product's attributes will live up to their expectations?
3. **Features**: Do participants perceive the product to have a rich set of features that will increase their willingness to buy?

4. **Aesthetics**: Do participants perceive the product to exhibit beautiful design and craftsmanship?

To measure the above quality dimensions, a 10-point-likert scale (0 = low and 10 = high) was used, thus creating a 40-point-aggregate score. Here, we assumed an uniform weighting across all the five dimensions, but expected the dimensions to vary from subject to subject in terms of weighting and importance.

Anticipated Satisfaction

In addition to perceived quality, we asked participants to rate their anticipated satisfaction from using or consuming the product under each purchase context. Similarly, we used a 10-point scale (0 = low satisfaction and 10 = high satisfaction) for this question.

Pretest: Preliminary Evidence for Rounded Price Hypothesis

Objective

As a first step, we needed to ascertain that the purchasing contexts (hedonic and utilitarian) for the set of products (link to full set of product associated with this study) that we compiled were reflective of the perception of the broader population. We saw this as a critical step to mitigate personal biases in the overall experiment design. Additionally, we wanted to narrow the field of products by only using product/context pairs that resonated the most with the subjects and elicited the strongest responses.

Recruitment

For the pre-test survey, we solicited help from friends, families, and acquaintances via direct emails and messages as well as virtual group discussions.

Methodology

Each participant was asked to indicate their intention for the product and purchase context as either associated with a hedonic or utilitarian goal (i.e., champagne for a friend's birthday; a calculator for school). After seeing the product images, feature description and its context, participants rated their purchase context on two 10-point-likert scales to measure whether the purchase decisions are driven by feelings (hedonic) versus cognition (utilitarian). Below is an example to illustrate.

Product: Digital Camera
- **Hedonic Context**: Two years after the outbreak of COVID-19, life is almost back to normal, so you decide to book an extended family vacation in Italy. Not wanting to miss a moment, you decide to buy a new digital camera that will capture all of the memories and historic landmarks.

  On a scale of 1-10 (1 - Low, 10 - High) choose how much you agree with the following statements:

- ○ **Statement 1**: The purchase decision in the above context is driven by needs
- ○ **Statement 2**: The purchase decision in the above context is driven by desires

- ● **Utilitarian Context**: You are a relatively new self-employed photographer and just booked your first photo shoot for a wedding. You decide to buy a new camera as an investment that will improve the quality of the pictures and accelerate your new business' growth.

  On a scale of 1-10 (1 - Low, 10 - High) choose how much you agree with the following statements:
  - ○ **Statement 1**: The purchase decision in the above context is driven by needs
  - ○ **Statement 2**: The purchase decision in the above context is driven by desires

Analysis

In order to facilitate the goal of choosing the product-context pairs, which elicited the strongest response, we employed employed the following methodology:

For each product-context pair, the Median Signed Difference (MSD) was computed between the two responses where the MSD represents the median need rating minus the median desire rating. This number is expected to be positive for utilitarian contexts and negative for hedonic contexts.The results of the pre-test produced a set of product-context pairs with reduced bias that were used as input for the main experiment.
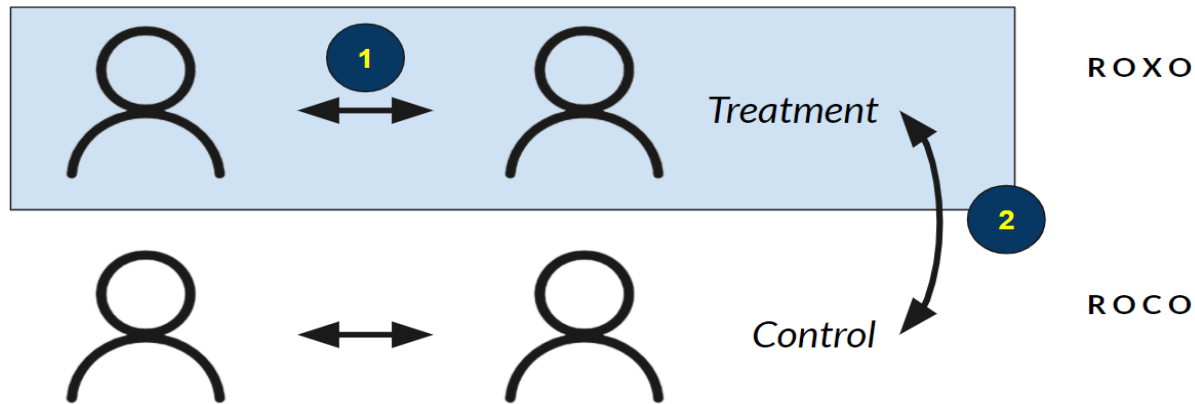
Main Experiment

Type of Study

The main experiment was survey based. The exposure was defined as showing the product pictures, features, and price.

Price Roundness and Consumption Goal

We initially planned for a 2x2 design (i.e. purchase context and pricing scheme) in which within the product context, there can be two types of prices (i.e. $89.95 and $90.95) compared to a given round price. Here, both options can be randomized to assess if the effects are different compared to the round price.

However, as will be shown in the later sections, the decision was made to descope the study, and focus only on the hedonic purchase hypothesis; the results from the pretests failed to return any strong signals in our utilitarian product-purchase contexts. Rather than spending additional time and effort on weak signals, the team moved its focus and efforts to validated signals (please reference the Results section for additional information).

For our final design, we decided on a before-and-after placebo design (see figure below).

Recruitment

MTurk was used to recruit survey respondents due to its large worker base and the ability to geo-target only users that resided in the United States.

Methodology

The methodology used in the main experiment is as follows:
- Randomization between the test and the control groups
- User perception across quality questions for all three products were measured in both groups
- Exposed the price of the product to subjects; **treatment was deemed delivered after price exposure**
- The same questions were asked again to the users, which would help determine if product perception had changed after the treatment had been delivered

Response Measurement

Once the exposure was established, standard questions were asked again to assess the user's ratings for the key aspects of product quality (final metric was the sum of ratings across all of these quality parameters).

Control Variable Measurement

The list of control variables (i.e. metadata) collected and used for control during the experiment is as follows:

- Age range
- Gender
- Household income
- Education level

Analysis

Power Analysis

Because we did not know the actual effect size prior to the experiment, we computed necessary sample size based on various effect sizes. The table below illustrates the number of samples/responses needed per experiment condition. Since three (3) independent surveys were utilized for this study, the sample sizes below are for *each* condition and not in *totality*. Moreover, for each effect size, an alpha of 0.5 and power of 0.8 was assumed.

Effect size	Samples per experiment condition
0.05	6,280
0.1	1,571
0.2	393
0.3	175
0.4	99
0.5	64
0.6	45
0.7	33
0.8	26
0.9	20

Power of Test
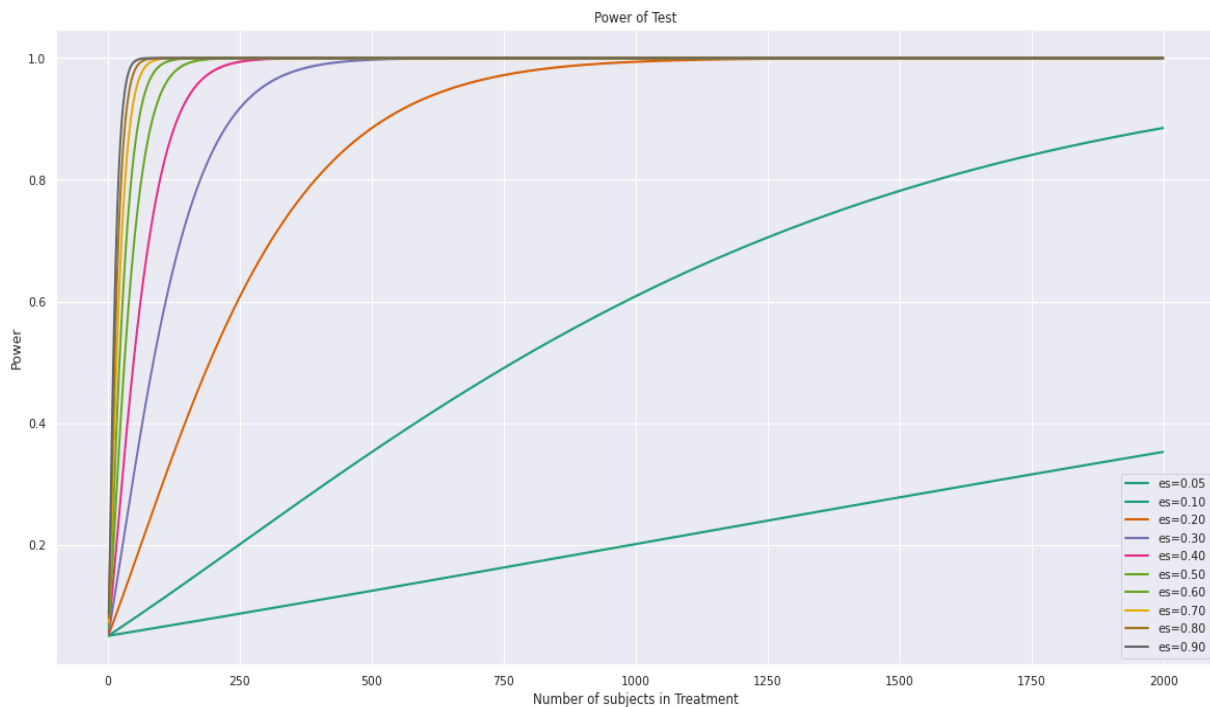
Check for Balance

The first step of this analysis was to check for balance across different study cohorts. This helped us confirm the randomization actually worked as intended.

Main Effect

The Average Treatment Effect was measured as the difference between treatment and control groups with respect to the cumulative quality score (this section identifies and explains the five factors that were chosen to measure product perception). Subjects were asked to rate each factor for each product on a scale from 1 to 10 in which the sum of the four scores became the final composite response variable called the quality score. As per the experiment design, the same questions were asked of each subject twice (pre and post treatment). In this way, the difference between the pre- and post-quality score for each subject captures the impact of our delivered treatment, which we designate as the quality-impact metric. Thus, the average difference in the quality-impact metric between test and control groups represents the main effect of the experiment.

ANALYSIS

Pre-Test: Preliminary Evidence for Rounded Price Hypothesis

Survey Statistics

As discussed before, the purpose of the pretest is to ascertain that the hedonic and utilitarian purchasing contexts for the set of products that were compiled are reflective of the perception of

the broader population, which was done to mitigate "researcher bias". The process began with the selection of four (4) products that were then divided into two (2) hedonic and two (2) utilitarian contexts each. The survey participants were asked whether the purchase decision was primarily driven by "needs" or "desires" on a 10 point Likert scale (1 - strongly disagree and 10 - strongly agree).

In addition, demographic metadata was also collected from participants in order to analyze the results in the right context. The table below summarizes the pretest setup:

| Product / Theme	# Purchase Contexts	# Questions per Purchase Context	# Total Questions
Camera	4	2	8
Bicycle	4	2	8
Headphone	4	2	8
Laptop	4	2	8
Demographic data	NA	NA	5
TOTAL	**16**	**8**	**37**

Convenience sampling was used to conduct the pre-test and forty-two (42) distinct responses were received. The four key demographic dimensions we collected were:
- Age
- Gender
- Highest education level
- Annual household income

As convenience rather than random sampling was used, we did not expect to find a balance across these covariates, which can be observed in the figures below.
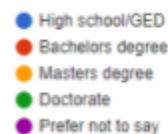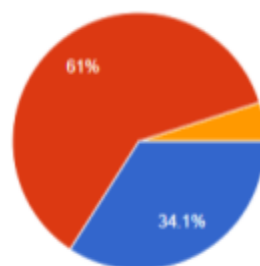
What is your age range?

41 responses

- 0-17
- 18-25
- 26-35
- 36-45
- 46-55
- 56+
- Prefer not to say
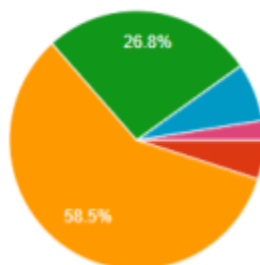
What is the the highest grade or level of school you have completed?

41 responses

- High school/GED
- Bachelors degree
- Masters degree
- Doctorate
- Prefer not to say

What is your gender?

41 responses

- Female
- Male
- Prefer not to

What is your approximate annual household income?

41 responses

- 0 - $100,000
- $100,001 - $200,000
- $200,001 - $300,000
- $300,001 - $500,000
- $500,001 - $1,000,000
- $1,000,000+
- Prefer not to say
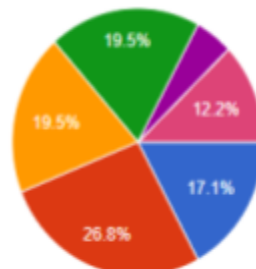
- The 26-35 age range is over-represented in the sample (58.5%) followed by 36 - 45 years (26.8%) with the remaining age buckets being largely underrepresented
- There is a slight skew in gender with a 60/40 split between male and female (2 participants preferred not to disclose)
- In terms of education level, holders of Bachelors and Masters degrees were overrepresented in our sample (containing ~88% subjects)
- Finally, annual household income was more evenly distributed across income buckets, however, the median household income in the $100,000 to $200,000 bracket, which is not representative of the general population in the United States

Please refer to appendix [2] for looking at our actual pre-test survey.

Analytical Framework and Results

In the pretest methodology section, we introduced the MSD, which we rationalize using the following approach:

- Started with four products with two hedonistic and two utilitarian purchase contexts each
- Conducted the pre-test to make ensure that the imaginary scenarios would induce the expected response in the subjects when running the actual test

- Created a summary statistic that measured the strength of the signal to help with the rank-ordering of products-context pairs that were appropriate to employ in the main experiment

Additionally, the <u>absolute value</u> of the MSD for any given product-context pair was used to represent the <u>context clarity</u> with larger values representing more clarity. The below table illustrates the purchase contexts and their computed metrics for both hedonic and utilitarian contexts.

Hedonic Context Clarity

Product	Purchase Context	Scenario	MSD
Camera	Hedonic	Your significant other's birthday is coming up and you decide to buy the pictured Canon digital camera as a gift.	-6
Headphone	Hedonic	You enjoy taking daily walks, especially after dinner. On such walks, you like to put on your headset and listen to music or podcast. While there's nothing wrong with your headset, you want to buy a new pair that's wireless and noise-canceling, so you can filter out any distracting noise.	-6
Camera	Hedonic	Two years after the outbreak of COVID-19, life is almost back to normal, so you decide to book an extended family vacation in Italy. Not wanting to miss a moment, you decide to buy the high-end Canon camera to capture all of the memories and historic landmarks.	-5
Laptop	Hedonic	You love to play PC games and want to have a powerful laptop that you can take anywhere and still game with crisp graphics, no lag, and solid connectivity to WIFI, so you can play with your friends online.	-5
Headphone	Hedonic	You decide to take a long-overdue vacation to another country in which the flight will take 14 hours if there are no delays. While you have other headphones, you decide to buy a noise-canceling one such as the one depicted for the long trip.	-3
Laptop	Hedonic	Because most movie theaters have shut down due to the pandemic, the only way to watch shows or movies is at home. Since you share the TV with 3 kids and a significant other, it's almost impossible to watch your show at the time you want. So you decide to buy a new laptop that you can take anywhere to watch your favorite shows and movies without having to compete for the TV.	-3
Bicycle	Hedonic	To get more exercise, you decide to pull the trigger on the depicted bicycle, since you believe it's an investment in your health.	-2
Bicycle	Hedonic	You recently moved to a new neighborhood in sunny San Diego that's only five miles away from where you work. Due to congestion and traffic, it would take less time to commute to work riding a bicycle than driving. So you decide to purchase the depicted bicycle to commute from and to work.	1

Utilitarian Context Clarity

Product	Purchase Context	Scenario	MSD
Camera	Utilitarian	You are a new self-employed photographer and just booked your first photoshoot for a wedding. You decide to buy a brand new Canon camera as an investment to improve the quality of the pictures and hopefully accelerate the growth of your new venture.	3
Laptop	Utilitarian	You are a freelance graphic designer and have decided to invest in a high-end laptop that will boost your productivity and accelerate the growth of your business.	2
Bicycle	Utilitarian	You recently underwent knee surgery and cannot perform any high-impact aerobic exercise such as jogging or running. So you decide to buy the depicted bicycle so you can ride both indoors and outdoors.	2
Headphone	Utilitarian	Due to the pandemic, you are working from home 100% of the time. As such, video conferencing calls have increased exponentially and dominate your day. To improve your ability to hear what others are saying while blocking out external noises such as leaf blowers or dog barks, you decide to invest in a wireless noise-canceling headset to boost productivity.	1
Laptop	Utilitarian	You recently started to day trade stocks as another way to augment your income. Due to a series of profitable trades, you decided to become a full-time day trader. As such, you also decide to upgrade to a high-end laptop that can take you anywhere but without missing a beat to the market.	0
Camera	Utilitarian	You are an amateur photographer hoping to win this year's National Geographic Photography Contest, so you decide to invest in a high-end Canon digital camera to improve your odds of winning.	-1
Headphone	Utilitarian	You're an avid exercise enthusiast and believe working out with music gives you a significant performance boost. So you decide to pull the trigger on the depicted sweat-resistant headphones, which you can run and lift weights with.	-2
Bicycle	Utilitarian	You've decided to be more active in the new year and recently joined a bicycle club that organizes group rides on weekends. Your first group ride is next weekend, so you decided to buy the depicted bicycle.	-3

Ideally, the MSD should be positive for utilitarian contexts and negative for hedonistic contexts. However, survey participants had more clarity towards the hedonistic relative to the utilitarian scenarios.

Rather than expending additional resources and time on finding a validated set of utilitarian driven products and purchase contexts, we decided to descope the research and focus only on the hedonic hypothesis (i.e. round prices will elicit a higher perception of product quality and anticipated satisfaction from users, if the purchase context is primarily driven by desires), since we had strong signal from the original pretest survey.

Results

Selection of Product-Context Pair

In order to select viable product-context pairs, the threshold to qualify a given product context to be part of the consideration set must be defined given that the maximum difference can be 9 units given the 10 point Likert scale. Thus, we set the threshold to be **absolute(MSD) >= 5**. As stated in the previous section, none of the product-context pairs in the utilitarian group of either pretest passed the threshold. However, there are five pairs that qualify for this definition in the hedonic group. Additionally, because the Camera has two (2) purchase contexts that qualify under the MSD threshold, it was decided that only the scenario that elicited the strongest signal should be used instead of including both.

In summary, we use the following product-context pair in our final experiment (please see the appendix [4] for detailed distribution of individual scores for each product):

Product	Purchase Context	Scenario	MSD
Camera	Hedonic	Your significant other's birthday is coming up and you decide to buy the pictured Canon digital camera as a gift.	-6
Headphone	Hedonic	You enjoy taking daily walks, especially after dinner. On such walks, you like to put on your headset and listen to music or podcast. While there's nothing wrong with your headset, you want to buy a new pair that's wireless and noise-canceling, so you can filter out any distracting noise.	-6
Laptop	Hedonic	You love to play PC games and want to have a powerful laptop that you can take anywhere and still game with crisp graphics, no lag, and solid connectivity to WIFI, so you can play with your friends online.	-5

Main Experiment: Price Roundness and Consumption Goal

Data Collection and Preparation

Through MTurk, we engaged with two-hundred-sixty-three (263) users for our control group and four-hundred-twenty-two (422) users for the treatment group and found the following unique challenges:

- **Repeat responses**: There were seven-hundred-sixty-two (762) responses in the control group and six-hundred-seventy-seven (677) responses in the treatment group. In many instances, multiple responses from a single respondent were received. There were two types of repetitions that were observed:
    - Same respondent answering the same survey multiple times
    - Same respondent answering test and control surveys

- **Homogeneous responses**: We found that in forty-nine (49) cases, the respondent answered all the questions with 10/10 points.

Pricing Type	Original Response count	Deduped count	% share
High (Non-round)	348	119	34%
Low (Non-round)	413	144	35%
Round	677	422	62%
Total	**1438**	**685**	**48%**

To mitigate these unwanted effects, only the first response was taken from each unique user. By deleting the "10/10" responses aims to remove any noise in the data; choosing the first response from each respondent ensures that only the unbiased response is taken.

Summary Statistics

Bin_group	C		T		
Variable	N	Percent	N	Percent	Test
gender	263		422		X2=2.362
... Female	125	47.5%	213	50.5%	
... Male	130	49.4%	201	47.6%	
... Non-binary	5	1.9%	3	0.7%	
... Prefer not to say	3	1.1%	5	1.2%	
education	261		422		X2=5.475
... Bachelors	125	47.9%	220	52.1%	
... Doctorate	9	3.4%	6	1.4%	
... High school/GED	65	24.9%	112	26.5%	
... Masters	60	23%	79	18.7%	
... Prefer not to say	2	0.8%	5	1.2%	
age_group	261		420		X2=2.958
... 18-25	31	11.9%	40	9.5%	
... 26-35	89	34.1%	161	38.3%	
... 36-45	62	23.8%	94	22.4%	
... 46-55	43	16.5%	61	14.5%	
... 56+	36	13.8%	63	15%	
... Prefer not to say	0	0%	1	0.2%	
income_group	263		421		X2=6.371
... $1,000,000+	3	1.1%	8	1.9%	
... $100,001 - $200,000	45	17.1%	72	17.1%	
... $200,001 - $300,000	18	6.8%	26	6.2%	
... $300,001 - $500,000	20	7.6%	15	3.6%	
... $500,001 - $1,000,000	12	4.6%	23	5.5%	
... 0 - $100,000	158	60.1%	266	63.2%	
... Prefer not to say	7	2.7%	11	2.6%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

The four demographic variables (age, income, education and gender) were collected from each respondent and the balance check revealed a balanced dataset. The distribution across treatment and control sets is fairly even and the Chi-Squared statistic is not statistically significant for any of the variables. These factors combined provide confidence that the randomization worked as intended.

Analysis

To calculate the treatment effect, linear regression and the following methodology were used:

- The sum of the pre-treatment exposure (i.e. price) was used as a covariate in order to reduce the variance in the treatment effect
- The demographic features recorded earlier as covariates, which was done with the intention to reduce the standard error of the treatment effect
- The binary column named "Bin_group" as the treatment variable, which is the binary indicator of random group assignment

From each subject, in both treatment and control, information was collected from three different products: (**1**) camera, (**2**) headphone, (**3**) laptop. In order to understand the generalizability of the hypothesis, a separate regression was run for each product.

Our findings show that the treatment effect was not statistically significant for any of the three products, and therefore failed to reject our null hypothesis, which stated that under the hedonic purchase context, the product quality perception is not different across round and non-round pricing.

There could be multiple reasons for this result:

1. <u>The experiment may have been underpowered</u>. There are a few ways this could have happened:
    a. <u>There might be issues with the strength of the treatment dosage.</u> If this is true, our sample may have not been large enough to observe a weaker treatment effect. This would mean our results represent a false negative.
        i.
    b. <u>The total sample size may be too small</u>. Based on the previous power calculation, 80% power with the current sample size would only be achieved if the effect size was >= 0.2. The estimates indicate a smaller effect size, meaning that more data would need to be collected in order to have a sufficiently powered hypothesis test.

2. <u>MTurk respondents may not have answered the questions properly</u>. Roughly forty-nine (49) cases were identified where the respondents marked 10/10 for all the answers. In general, if it is perceived that the quality of responses from MTurk is poor, then the experiment would have to be replicated with samples from a new source. An example of this MTurk quality issue is that around 17% of respondents in both test and control groups marked their income to be >= $200K per annum which seems implausible and indicates potential data quality issues (~2% of respondents clicked $10M+)

3. Finally, it is possible our hypothesis is false and in general, for the products we chose in our consideration set, the desired effects do not exist, meaning, round versus non-round pricing does not have an impact on product quality perception.

```
================================================================================
                                      Dependent variable:
                 ---------------------------------------------------------------
                 Camera - Post treatment  Headphone - Post treatment  Laptop - Post treatment
--------------------------------------------------------------------------------
Treatment                    0.093                    0.599                    0.433
                            (0.347)                  (0.389)                  (0.414)

Previous Score - Camera      0.841***
                            (0.040)

Previous Score - Headphone                            0.907***
                                                     (0.028)

Previous Score - Laptop                                                        0.885***
                                                                              (0.033)

Constant                     5.462***                 3.767***                 1.802
                            (1.848)                  (1.429)                  (2.473)

--------------------------------------------------------------------------------
Age buckets                    Yes                      Yes                      Yes
Gender buckets                 Yes                      Yes                      Yes
Education buckets              Yes                      Yes                      Yes
Income buckets                 Yes                      Yes                      Yes
--------------------------------------------------------------------------------
Observations                   679                      679                      679
R2                            0.624                    0.684                    0.610
Adjusted R2                   0.613                    0.675                    0.598
Residual Std. Error (df = 658) 4.474                    4.864                    5.303
F Statistic (df = 20; 658)   54.615***                71.265***                51.471***
================================================================================
Note:                                                        *p<0.1; **p<0.05; ***p<0.01
```

Heterogeneous Treatment Effect

Looking at the exploration of heterogeneous treatment effects, treatment control interaction was used to attempt to understand if the Confounder Adjusted Testing and Estimation (CATE) is significant for any specific sub-group. It is important to take into consideration the multiple comparison issues and take a conservative approach when using Bonferroni Correction.

In the above table, the p-value of F-statistic for models with control treatment interactions is noted. It seems that for only one product (headphone) statistical significant heterogeneity for different education levels can be observed. Since the main effect is not statistically significant replication is needed before declaring success.

	Camera	Headphone	Laptop	Bonferroni corrected threshold
Age	0.423	0.546	0.978	0.01
Income	NA	0.263	0.024	0.0083
Gender	0.578	0.132	0.128	0.0167
Education	0.036	0.0118	0.035	0.0125

Clustered Assignment

Since the experiment is designed to have three (3) responses across three (3) different products per person, a clustered design can be utilized to improve the standard errors and assess if statistical significance can be achieved.

After running the clustered assignment setup, we see that we still get a statistically insignificant treatment effect of **0.396 (0.278)**.

```
===============================================================
                                  Dependent variable:
                          -------------------------------------
                          Clustered Assignment - Post treatment value
---------------------------------------------------------------
Treatment                                0.396
                                        (0.278)

Pre-treatment average score              0.881***
                                        (0.022)

Constant                                 0.809
                                        (1.052)


---------------------------------------------------------------
Age buckets                               Yes
Gender buckets                            Yes
Education buckets                         Yes
Income buckets                            Yes
---------------------------------------------------------------
Observations                             2,055
R2                                       0.638
Adjusted R2                              0.634
Residual Std. Error              4.878 (df = 2031)
F Statistic               155.925*** (df = 23; 2031)
===============================================================
Note:                         *p<0.1; **p<0.05; ***p<0.01
```

CONCLUSION

As noted above, balanced data across multiple demographic covariates exists for all three products. However, the treatment effect is not statistically significant. The experiment results, therefore, fail to reject the null hypothesis that product quality perception is not different across round and non-round pricing, under the hedonic purchase context.

The general consensus is that the test conducted was underpowered (i.e. we were not able to collect large enough samples to detect a potentially more subtle effect) and there were enough issues detected with MTurk data quality that it warrants replication of the experiment in a different setting.

LIMITATIONS AND FUTURE WORK

There are several limitations related to this study that could fuel further research in this domain.

Randomization

The randomization component of this study was impacted by negative factors such as platform (i.e. MTurk), budget and time constraints. Given the limitations that these constraints inherently introduced, below are some of the recommended techniques, which might help mitigate such risks in future studies:

- Schedule the design phase as a *priority* in the action plan and begin administering the survey to subjects at an *earlier time* in the experimental cycle
- Implement *time-series randomization* of survey by incorporating a "release-and-withdrawal" approach to help focus on the time frames when it is assumed certain demographics are generally engageable

Generalizability

Because the survey hinged on specific product and purchase contexts, the results may not be generalizable to other products. This can be even more problematic within the same family product, but different brands. More specifically, luxury brands tend to be seen as more hedonic even though the underlying product may be viewed as utilitarian and functional (i.e., Rolex watch). In our study, we tried to pick non-luxury brands to minimize the "brand effect." Some ways to improve the generalizability of the findings to other product and purchase contexts includes:

- Categorize products into homogeneous data buckets to facilitate the evaluation process of the results to allow an apple-to-apple comparison at the category level
- Aggregating the results based on general categories will also help with generalizability and transferability of similar products within a category. Thus, facilitating cross-domain comparisons for evaluation and study

Spillover

Because three separate surveys were launched on MTurk, the responses were subject to spillover effects. Since there is no mechanism in MTurk to restrict workers from taking multiple surveys published under the same publisher, a worker could technically take both the round and non-round surveys. To mitigate such effect, the surveys were published on different times and days while titles and keywords were also altered with each publication in an attempt to attract a different participant base. Additional mechanisms that could help facilitate the outreach to a more diverse audience include:

- Use a platform that allows for a "pull" model in which survey publishers have control over randomization
- Cross-examine MTurk IDs across the 3 surveys and remove duplicate responses

Pricing Convention

We defined round prices in this study as price numbers with two zeros after the decimal point (i.e. $20.00). However, there could be other operational definitions of rounded prices such as price numbers with no decimals (i.e. $20). While we expect such price numbers to yield the same pattern of results found in this paper, it is an empirical question that needs to be examined in future research. Furthermore, when the magnitude of the price increases (i.e. when buying a high-priced product such as a car), a price such as $34,973 could be perceived as non-rounded as compared with a price of $25,000. This could mean that the concept of roundedness can be relative and further experimentation is required to fully understand the implications.

Risk-Seeking Behavior

It would also be interesting to investigate the impact of round numbers on risk-seeking behaviors. For example, would consumers become more risk-seeking when the stakes are presented in rounded vs. unrounded numbers? Interestingly, it could potentially help explain a phenomenon in the stock market in which investors have a strong tendency to purchase stocks when its price coincides with a round number [5]. Examining these issues will provide researchers and marketers with richer insights into the factors that influence the effects of price numbers on consumer decision-making.

Ethical Considerations

Lastly, exploring how the study's findings can be used to price products in an economic and ethically sound manner would be a worthy topic for research. Specifically, the findings of this study can be used to better inform research in the area of physiological and value-based pricing schemes; the price of the product is not solely based on the cost to produce and market value, but rather on the perceived value by the consumer. It's important to understand that this perceived consumer value can be manipulated to elicit false-positive responses and studies that focus on understanding the harmful effect of these practices would help to protect consumers from predatory pricing behavior in the real world scenarios.

References

[1] Manoj, T. and Morwitz, V. (2005). Penny Wise and Pound Foolish: The Left Digit Effect in Price Cognition. *Journal of Consumer Research.* 32. 54-64.

[2] Robert M. Schindler, Thomas M. Kibarian (1996), Increased consumer sales response though use of 99-ending prices, *Journal of Retailing*, Volume 72, Issue 2, 187-199

[3] Wadhwa, M., & Zhang, K. (2015). This Number Just Feels Right: The Impact of Roundedness of Price Numbers on Product Evaluations. *Journal of Consumer Research*, 41(5), 1172-1185

[4] Robert M. Schindler & Thomas M. Kibarian (2001). Image Communicated by the Use of 99 Endings in Advertised Prices, Journal of Advertising, 30:4, 95-99

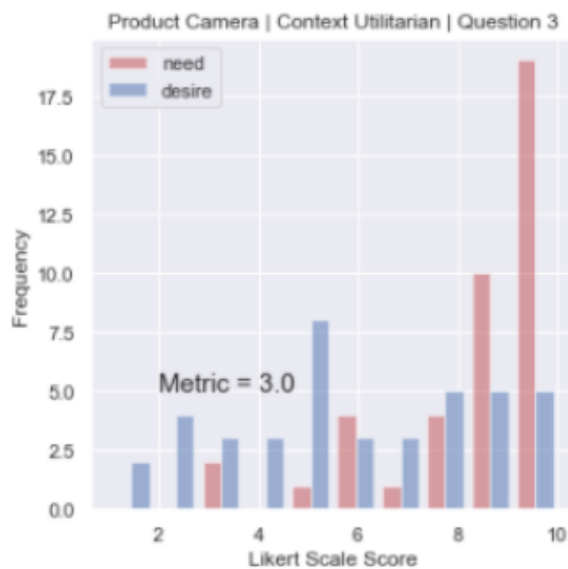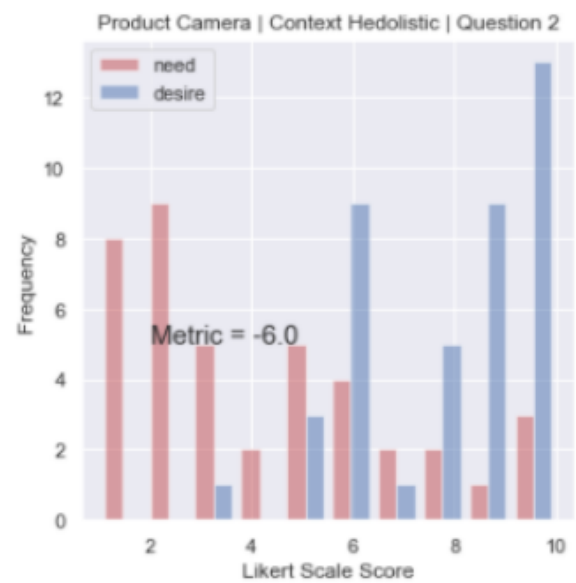[5] Weiss, Richard E. (2009), "Forex Trading: The Power of Round Numbers," http://ezinearticles.com/?forex-trading-the-powerof-round-numbers&idp1851400/.
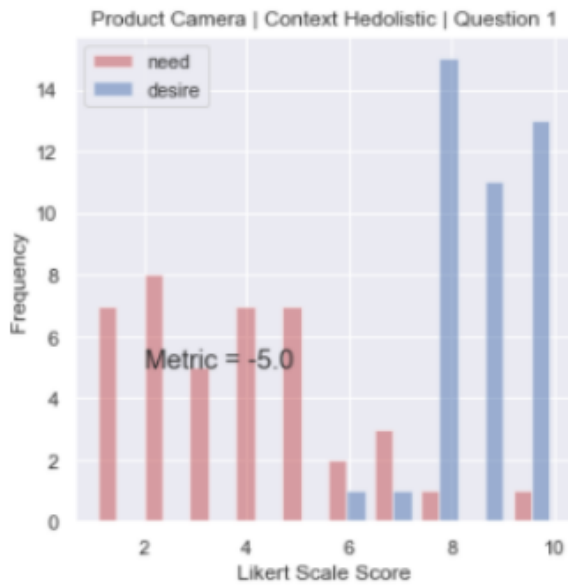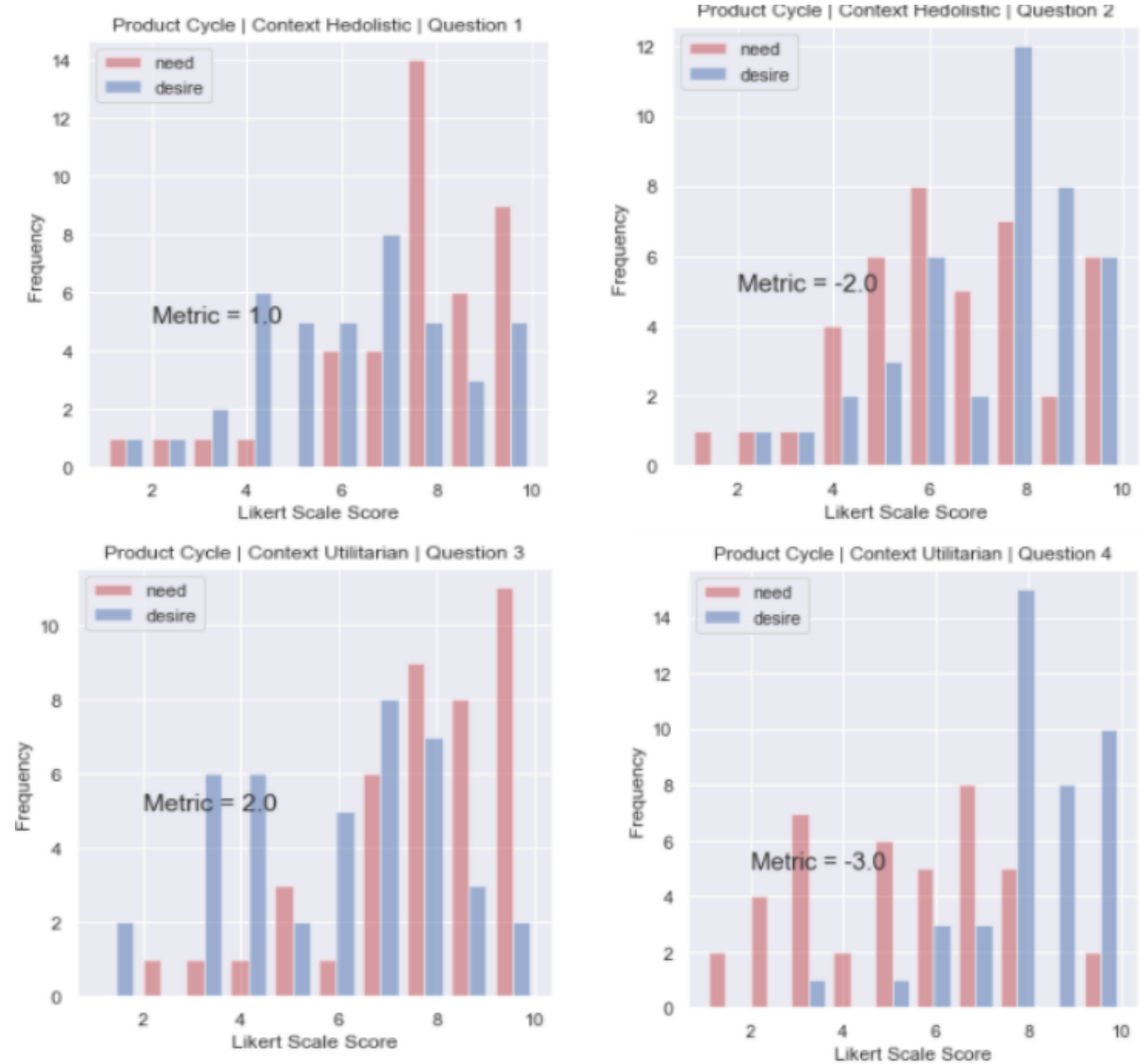
Appendix

[1] Study proposal
[2] Pre-test survey
[3] Product wise score distribution [pre-test]

Camera

Bicycle

Headphone

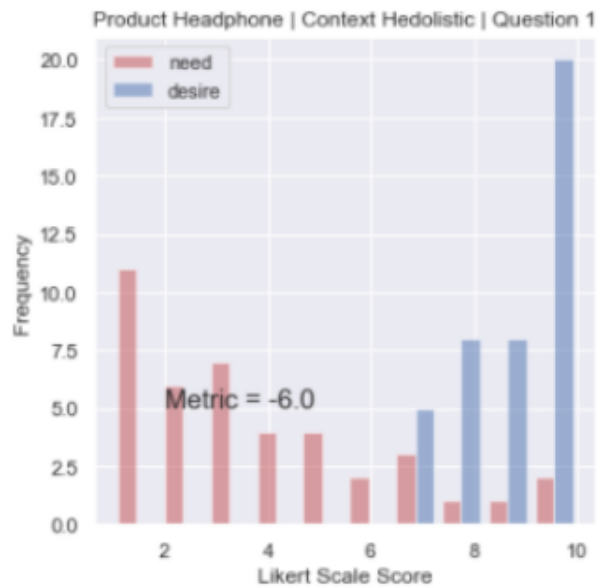Product Headphone | Context Hedolistic | Question 1
Metric = -6.0

Product Headphone | Context Utilitarian | Question 2
Metric = 1.0

Product Headphone | Context Utilitarian | Question 3
Metric = -2.0

Product Headphone | Context Hedolistic | Question 4
Metric = -3.0

Laptop

Product Laptop | Context Hedolistic | Question 1
Metric = -5.0

Product Laptop | Context Utilitarian | Question 2
Metric = 2.0

Product Laptop | Context Hedolistic | Question 3
Metric = -3.0

Product Laptop | Context Utilitarian | Question 4
Metric = 0.0