

24 August 2015

How to Nickel-and-Dime Your Customers



W241-3: Field Experiments
Summer 2015
Final Project
Tuhin Mahmud | Marguerite Oneto

Table of Contents

1. Introduction
 - 1.1 The Literature
 - 1.2 The Hypothesis
2. The Experiment
 - 2.1 The Context
 - 2.1.1 Skin in the Game: Spending Monetary Units
 - 2.1.2 Attention to Detail: The Final Quiz
 - 2.2 The Treatments
 - 2.2.1 Control
 - 2.2.2 Treatment 1
 - 2.2.3 Treatment 2
 - 2.3 The Subjects
 - 2.4 The Randomization Process
 - 2.4.1 Randomized Block Design
 - 2.4.2 Using Qualtrics to Randomize
 - 2.4.3 Covariate Balance
 - 2.4.4 Manipulation Test
 - 2.5 The Outcome Measures
 - 2.5.1 Satisfaction and Loyalty
 - 2.5.2 The Likert Scale
 - 2.5.3 Measuring the Average Treatment Effect (ATE)
3. The Results
4. Discussion
 - 4.1 Interpretation of results
 - 4.2 Challenges
 - 4.2.1 The Lab Environment
 - 4.2.2 Generalizability
 - 4.2.3 Click-Through
 - 4.2.4 Excludability
 - 4.2.5 Non-Interference
5. Conclusion
 - 5.1 Summary of Findings
 - 5.2 Future Work
6. Appendix
 - 6.1 Additional Materials
 - 6.2 Example of the Purchasing Experience: Treatment 2
 - 6.3 Distribution of Likert Scores by Treatment Group

1. Introduction

Many companies that compete on price advertise low base prices in order to draw in customers, but then they hit those customers with add-ons and additional fees once the purchasing process has started. Customers get “nickeled-and-dimed.” This strategy is called price-partitioning, and it can alienate customers, decrease repeat sales, and damage a company’s brand. When is it in a company’s best interest to take such an approach to pricing, and is there a way to do it that protects the company’s reputation?

1.1 The Literature

In "[When Should You Nickel-and-Dime Your Customers?](#)"¹, Rebecca W. Hamilton, et. al., provide definitive rules for when price partitioning is the revenue-maximizing strategy. Based on research conducted in their consumer lab, they develop seven separate rules for deciding whether to charge separately for add-ons or to charge one all-inclusive price. They present their results in a decision tree. The root starts with the question, "Is the consumer committed to making the purchase from you [or will they comparison shop]?" Further questions consider characteristics of the product being sold, what the competition is doing, as well as the psychological state of the consumer.

Now assume that a company has traversed this tree and come to the conclusion to pursue price partitioning. What next? How should they best present their pricing scheme to protect their brand? Several Wharton professors discuss some factors to take into consideration in "[Nickled and Dimed: Is It Possible to 'Over-fee' Consumers?](#)"² Most of these factors are psychological. As the article points out, "What businesses perceive as a 'reasonable' fee becomes an intolerable fee in the eyes of consumers when the psychological pain becomes too great."

Marketing professor David Bell says there are three things that consumers must have in order to be comfortable with additional fees. First, they must have an understanding of the need for the fees. Second, they must have the perception that the fees are fair. Third, they must have the feeling that they receive some benefit from the fee. When a company puts its price partitioning schedule into practice, how can it take these concerns into consideration so it can avoid that psychological breaking point?

The Wharton article states, "What is fair ... is determined simply by how a company can spin it." In this research, we will examine ways to "spin" fees that might address the three 'must-haves' for consumers. We will measure how addressing these consumer concerns influences the consumer's satisfaction with the buying experience, as well as the consumer's

¹ Hamilton, R. W., Srivastava, J., & Abraham, A. T. (2010). When Should You Nickel and Dime Your Customers? *MIT Sloan Management Review*.

² Nickled and Dimed: Is It Possible to 'Over-fee' Consumers? *Knowledge@Wharton*. (2011). Retrieved on June 2, 2015 from <http://knowledge.wharton.upenn.edu/article/nickled-and-dimed-is-it-possible-to-over-fee-consumers/>.

perception of the company's brand.

1.2 The Hypothesis

We hypothesize that customers who are given an explanation of add-on fees will have higher satisfaction with the purchasing experience and will be more likely to recommend the company to others compared to those who receive no such explanation. The particular explanation would address the three consumer must-haves. It would include:

1. a company's reasons for charging the fee,
2. justification of the amount of the fee, and
3. what additional benefit the customer receives from paying the extra fee.

We also hypothesize that customers who are charged one all-inclusive fee have higher satisfaction than those who are nickel-and-dimed. We measure how much of the lost satisfaction is regained when pricing explanations are given.

2. The Experiment

To test the posed hypothesis, we conducted a randomized experiment where subjects had an online buying experience with varying presentations of the seller's price partitioning scheme. Sometimes subjects were charged one all-inclusive price. Other times subjects were given a base price and then had additional fees added on during the purchasing process. For those subjects who faced add-ons, sometimes they received explanations of the fees and sometimes they did not. At the end of the process, all subjects were given a survey asking them to rate their satisfaction with the experience and how they felt about the company.

2.1 The Context

In particular, subjects went through a simulation of purchasing a plane ticket online. First, subjects were given a base price for the plane ticket. Next, they had to walk through the steps of choosing a seat on the departing flight and choosing a seat on the returning flight. Finally, they had to go through the motions of choosing whether to check their luggage or carry it on the plane for both the departing and return flights.

We used the survey service from [Qualtrics](#) to build the purchasing experience. Please see Appendix 6.2 for an example of the screens that the subjects walked through during the simulation.

2.1.1 Skin in the Game: Spending Monetary Units

To ensure that the subjects cared whether they were being nickel-and-dimed, we had to force them to spend their own money. First, we had to pay all subjects for going through our simulation. Second, we tied their final payments to the amount of money they spent buying their plane tickets. The more money they spent on the ticket, the less money they would be paid at the end of the simulation.

At the start of the task, each subject was given 600 monetary units (MU\$) to spend. They were

then told that the final amount they would be paid depended on how many monetary units they had left over after purchasing the plane ticket. Subjects were given an exchange rate of MU\$1 = US\$0.01. Thus, for every 100 monetary units subjects had remaining at the end of the task, they would be paid US\$1. To be fair, the final price paid for the plane ticket was the same for all subjects: MU\$450. This was true whether they were charged an all-inclusive price or were nickel-and-dimed. Given the exchange rate, this meant everyone was paid \$1.50 for completing the task.

2.1.2 Attention to Detail: The Final Quiz

We wanted to encourage our subjects to pay attention to the details of our task, especially to the questions in our survey asking them their opinions of the simulation. To do this, we offered them a chance to earn a bonus at the end of the task. We provided them with a quiz to test their recall of certain details of the simulation, offering them an extra \$0.05 for each correct answer. We asked such questions as, "What was the starting price of the ticket?" and "What was the MU\$ to US\$ exchange rate?" We hoped that this would keep subjects from quickly clicking through the survey without giving the questions due consideration.

2.2 The Treatments

The experiment had two treatment conditions and one control condition. The control group was not nickel-and-dimed. Both treatment groups were. Treatment Group 1 did not receive any explanations for the add-on fees. Treatment Group 2 did receive the explanations, which were meant to address the three consumer 'must-haves' posited by Prof. Bell.

2.2.1 Control

Control group subjects were presented an all-inclusive fee at the start of the simulation. They were quoted a price of MU\$400. They then went through the process of choosing seat and baggage options without having to pay extra for them. In the final tally, they were charged another MU\$50 in taxes, for a total ticket price of MU\$450. Throughout the simulation, the control group saw exactly the same screens as Treatment Group 1, except the ticket price was the same flat fee on each screen -- MU\$400. The screens that Treatment Group 1 walked through had different prices on each screen and are described in detail below.

2.2.2 Treatment 1

Treatment Group 1 was presented with a starting price of MU\$240. The subjects were then shown a screen that asked them to choose seats for their departing and returning flights. The next screen had a map of a plane, where open seats were labeled \$30, and subjects were able to submit their seat choice for the departing flight. After making their choice, they were moved to the next screen where they were again provided with a map of a plane where open seats were labeled \$30. The displayed ticket price now showed MU\$270, reflecting the MU\$30 spent on the seat for the first flight. On this screen, the subjects submitted their seat choice for the returning flight. At this point, the ticket price had reached MU\$300.

The subjects were next passed to a screen that asked them to choose their baggage options.

The following screen showed the ticket price of MU\$300 and allowed the subjects to choose between checking a bag or carrying it on for the departing flight. Both choices cost MU\$50, taking the ticket price up to MU\$350. The next screen was similar to the previous one, allowing the subjects to make a baggage choice for the returning flight, and it displayed the ticket price of MU\$350. Once the subjects had made their returning flight baggage selection, the price had reached MU\$400. The subjects were finally directed to a screen showing a summary of all the charges (see Screen 10 in Appendix 6.2):

- MU\$120 base price for the departing flight,
- MU\$120 base price for the returning flight,
- MU\$60 in seat fees,
- MU\$100 in luggage fees,
- MU\$50 in taxes and other fees.

With the addition of taxes and other fees, the final price of the plane ticket reached MU\$450. Thus, Treatment Group 1 was nickeled-and-dimed and given no explanations for the extra fees.

A summary of the screens seen by Treatment Group 1 during the simulation follows:

1. “Please choose a seat for both your departing and return flights.”
2. Map of plane for departing flight. Ticket price = MU\$240
3. Map of plane for returning flight. Ticket price = MU\$270
4. “Please choose your baggage options for both your departing and return flights.”
5. Baggage choice screen for departing flight. Ticket price = MU\$300
6. Baggage choice screen for returning flight. Ticket price = MU\$350
7. Summary of Charges. Ticket price = MU\$450

2.2.3 Treatment 2

The simulation that Treatment Group 2 went through can be found in Appendix 6.2. It was exactly the same as that experienced by Treatment Group 1 except for two screens. For Treatment Group 2, the screen “Please choose a seat for both your departing and return flights” had an additional note, which gave the subjects three explanations for the added-on seat fee of MU\$30. It said,

Please note:

- The airline charges for a seat in order to cover its costs for cleaning them.
- The airline charges MU\$30 for a seat because that is what it costs to pay the cleaners to clean them after each flight.
- The benefit you receive for paying for your seat is a guaranteed clean seat when you board the plane.

This note attempted to address the three ‘must-haves’ of consumers.

The screen “Please choose your baggage options for both your departing and return flights” had the following additional wording, which explained the added-on luggage fee of MU\$50:

Please note:

- The airline charges for each bag you bring on the flight in order to cover its additional fuel costs for carrying the extra weight and to pay baggage handlers.
- The airline charges MU\$50 for each bag you bring on the flight because that is what it costs the airline, on average, in additional fuel costs and for salaries for the baggage handlers.
- The benefit you receive from paying for your baggage is that the airline's baggage handlers will load and unload your baggage from the plane for you.

This note was also an attempt to address the three ‘must-haves’ of consumers. Treatment Group 2 was thus nickeled-and-dimed, but they were also given explanations for the add-on fees.

2.3 The Subjects

Our subjects were recruited through [Amazon’s Mechanical Turk](#). We limited our workers to the US. This means that we should only consider our results generalizable to US residents. We did not limit ourselves to Master workers, but workers did have to meet the criteria that their Human Intelligence Task (HIT) approval rate was greater than 95% and the number of HITs they had completed to date was greater than 1,000. We wanted highly rated, experienced workers to ensure that our task was done with diligence and do not believe that this additional criteria is correlated with our experimental outcomes, described in Section 2.5.1 below.

We had no attrition from our study. All subjects answered all of our survey questions at the end of the simulation.

We had no non-compliance amongst our subjects. With our experimental design, subjects had no choice whether or not to comply with the treatment. They could not choose which screens to observe or not observe during the simulation.

We suspected that the variables age, income, and employment status might predict outcomes and so collected this demographic information after the task was completed. We collected it after the simulation because we did not want to disrupt the simulation flow with too many questions and because we did not believe the treatment would affect the reported values of these variables.

Before treatment, we also collected information on subjects’ travel habits. We were interested in whether they had recently traveled on a discount airline and whether they were more likely to check their luggage or carry it on. In particular, we asked them what airlines they had flown in the last 12 months and whether or not they checked their bags. In the end, we did not use these covariates in our analysis because we did not have time to parse the open-ended answers. They should be considered in future work.

2.4 The Randomization Process

2.4.1 Randomized Block Design

Due to our prior beliefs about airline travelers, we used a randomized block design to allocate subjects into our control and treatment groups. In particular, we believed that people flying

more regularly would likely have purchased more tickets online, thus making them more inured to the nickel-and-diming process that many airlines currently employ. We also believed that people traveling for personal or leisure reasons, who do not get business reimbursement, are more sensitive to price differences. Therefore, we divided our subjects into business versus leisure travelers and frequent versus infrequent fliers. We thus had four different blocks: 1) infrequent business travelers; 2) frequent business travelers; 3) infrequent leisure travelers; 4) frequent leisure travelers. The blocks were defined by answers to the following two questions:

Question 1: How many times have you purchased a plane ticket online in the past 12 months?

Answer Choice:

- a. 0 to 3
- b. 4 or more

Question 2: Were the majority of those tickets bought for business travel or personal travel?

Answer Choice:

- a. Business
- b. Personal OR "I have not purchased any tickets online in the past 12 months."

Within each of the four blocks, subjects were randomly assigned to one of the three groups (one control, two treatments) using simple randomized assignment. Each subject had a $\frac{1}{3}$ chance of ending up in each of the three groups.

2.4.2 Using Qualtrics to Randomize

Qualtrics has a very intuitive interface for setting up randomization flows for surveys. We used Survey Flow and the Randomizer's "Evenly Present Elements" option for each of the four blocks.

The screenshot shows the Qualtrics Survey Flow editor. At the top, there is a 'Then Branch If:' condition block. It contains two 'If' statements: 'How many times have you purchased a plane ticket online in the past 12 months? 0 to 3 Is Selected' and 'And Were the majority of those tickets bought for business travel or personal travel? Personal OR "I have not purchased any tickets online in the past 12 months." Is Selected'. Below this is a 'Randomizer' element. It is set to 'Randomly present 3 of the following elements' and has the 'Evenly Present Elements' checkbox checked. There are buttons for 'Move', 'Duplicate', 'Options', 'Collapse', 'Delete', 'Add Below', 'Move', 'Duplicate', 'Expand', and 'Delete'.

We did a test run of the survey with a small number of subjects to measure the distribution

among the different treatment groups within each block. We found that the subjects were evenly distributed among the three groups, which gave us confidence in our randomization process.

The breakdown of our final sample by block and treatment is as follows:

Block	Treatment Group		
	Control	Treatment 1	Treatment 2
Infrequent Business Traveler	6	9	5
Frequent Business Traveler	2	1	1
Infrequent Leisure Traveler	51	46	50
Frequent Leisure Traveler	6	0	3

Unfortunately, although not unexpectedly, non-frequent leisure travelers dominated the sample, and so we might consider limiting the generalization of our results to this group.

2.4.3 Covariate Balance

We also tested for covariate balance and found the distribution of subjects across age, income, and employment status consistent with our randomized design:

Age	TREATMENT GROUP		
	CTRL	Treat 1	Treat 2
15-20	1	2	0
21-30	29	24	29
31-40	22	22	18
41-50	9	5	8
51-60	1	1	0
60+	3	2	4

Income	TREATMENT GROUP		
	CTRL	Treat 1	Treat 2
0 - 10K	9	6	10
10 - 50K	45	30	34
50 - 75K	7	10	8
75 - 100K	3	5	4
100 - 150K	1	2	3
150 - 200K	0	2	0

Status	TREATMENT GROUP		
	CTRL	Treat 1	Treat 2
Employed	52	42	47
Unemployed	6	8	4
Retired	1	0	0
Student	4	3	2
Other	2	3	5

More formally, we ran a logistic regression with treatment group as the dependent variable and age, income, and employment status as the independent variables. None of the covariates -- age, income, or employment -- were predictive of the treatment group. All p-values of the coefficients were greater than 0.05:

Call:

```
glm(formula = as.factor(treat) ~ as.factor(age) + as.factor(income) +
  as.factor(employment), family = "binomial", data = data_to_date)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.51356	1.35212	0.380	0.704
as.factor(age)3	-0.20046	1.32920	-0.151	0.880
as.factor(age)4	-0.41635	1.36160	-0.306	0.760
as.factor(age)5	-0.51471	1.38702	-0.371	0.711
as.factor(age)7	15.38087	2399.54513	0.006	0.995
as.factor(age)8	-0.03184	1.51168	-0.021	0.983
as.factor(income)2	0.12072	0.54700	0.221	0.825

as.factor(income)3	0.67164	0.71331	0.942	0.346
as.factor(income)4	0.83192	0.84772	0.981	0.326
as.factor(income)5	1.32288	1.21751	1.087	0.277
as.factor(income)6	16.36224	1692.92852	0.010	0.992
as.factor(employment)2	0.31981	0.58144	0.550	0.582
as.factor(employment)3	-32.58121	3393.46872	-0.010	0.992
as.factor(employment)4	-0.18670	0.73074	-0.255	0.798
as.factor(employment)5	1.10452	0.92192	1.198	0.231

2.4.4 Manipulation Test

We ran a manipulation test throughout the study to ensure that the intended treatment was actually delivered. At the end of the purchasing simulation, we asked the subjects the following two survey questions:

1. During the task, were you charged separately for your seats and luggage?
2. During the task, did you have an opportunity to read explanations of the airline's seat and luggage charges?

Negative answers to both questions indicate that the subject was allocated to the control group. Affirmative answers to the first question with negative answers to the second question indicate that the subject was allocated to Treatment Group 1. Finally, affirmative answers to both questions indicate that the subject was allocated to Treatment Group 2.

From the results, we are confident that the correct treatment was delivered since most of the subjects' perceptions matched with the intended treatment.

The results from the manipulation test are shown in the table below:

		Actual Treatment Group		
Perceived Treatment Group		Control	Treatment 1	Treatment 2
Control	97%	3%	0%	
Treatment 1	36%	58%	6%	
Treatment 2	12%	19%	69%	

2.5 The Outcome Measures

2.5.1 Satisfaction and Loyalty

To assess whether nickel-and-diming and the explanation of fees had an effect on the subjects, we measured both customer satisfaction and customer loyalty using standard marketing techniques. At the end of the purchasing process, we measured customer satisfaction by asking the following three questions:

1. Overall, how satisfied or dissatisfied are you with this buying experience?
2. Overall, how satisfied or dissatisfied are you with paying separately for a seat on the plane?
3. Overall, how satisfied or dissatisfied are you with paying separately for your luggage?

The responses were gathered on a 5-point Likert scale: 1) Very Dissatisfied; 2) Somewhat Dissatisfied; 3) Neither Satisfied Nor Dissatisfied; 4) Somewhat Satisfied; 5) Very Satisfied.

We gauged customer loyalty by asking the following two questions:

1. How likely are you to purchase from this airline again?
2. How likely are you to recommend this airline to a friend or colleague?

The responses were gathered on the following 5-point Likert scale: 1) Very Unlikely; 2) Somewhat Unlikely; 3) Neither Likely Nor Unlikely; 4) Somewhat Likely; 5) Very Likely.

Not all subjects received the same questions. Those in the Control group were only asked the first satisfaction question and the two loyalty questions. Since they paid a flat fee up front and were not asked to pay separately for their seats and luggage, they did not answer the last two customer satisfaction questions. Those in both treatment groups were asked all five questions.

2.5.2 The Likert Scale

Each of the five questions listed above is a Likert item. The Likert scale is calculated by adding up the scores of the Likert items. In our case, a response of "Very Dissatisfied" corresponds to a score of 1, "Somewhat Dissatisfied" corresponds to a score of 2, "Neither Satisfied Nor Dissatisfied" corresponds to a score of 3, etc. The same holds true for the loyalty questions: "Very unlikely" corresponds to a score of 1, "Somewhat Unlikely" corresponds to a score of 2, etc.

We calculated two different Likert scales: a 3-item scale ("Likert Scale 3") and a 5-item scale ("Likert Scale 5"). Likert Scale 3 was calculated for all three groups using the following questions:

1. Overall, how satisfied or dissatisfied are you with this buying experience?
2. How likely are you to purchase from this airline again?
3. How likely are you to recommend this airline to a friend or colleague?

Likert Scale 5 was calculated only for the two treatment groups using all five questions. Likert Scale 3 has a range of [3, 15], while Likert Scale 5 has a range of [5, 25].

Because Likert scores are ordinal and not interval measures, there is controversy over whether comparing the means of Likert scales across groups is a valid way to measure average treatment effects. Nevertheless, we take the common academic view that there exists a latent continuous variable whose value characterizes the subjects' opinions and that the Likert score approximates this continuous variable. This, along with the Central Limit Theorem will allow us to compare the means of the Likert scales (the sums of the scores) in order to assess whether there is a treatment effect.

Additional controversy exists over whether we can evaluate differences of means on individual Likert items (i.e. individual questions). We will invoke the Central Limit Theorem again to overcome this objection.

2.5.3 Measuring the Average Treatment Effect (ATE)

To measure our average treatment effect, we used regression with covariate adjustment. We did this for Likert Scale 3, Likert Scale 5, and each of the Likert item questions. The regression equation is

$$Y_i = \beta_0 + \beta_1 d_i + \beta_2 BLOCK_{2i} + \dots + \beta_4 BLOCK_{4i} + \beta_5 AGE_{2i} + \dots + \beta_{10} AGE_{7i} + \beta_{11} INCOME_{2i} + \dots + \beta_{17} INCOME_{8i} + \beta_{18} STATUS_{2i} + \dots + \beta_{21} STATUS_{5i}$$

with indicator variables $BLOCK_{ki}$ indicating the block k to which subject i belongs, AGE_{ki} indicating the age bracket k to which subject i belongs, $INCOME_{ki}$ indicating the income bracket k to which subject i belongs, and $STATUS_{ki}$ indicating the employment status of subject i .

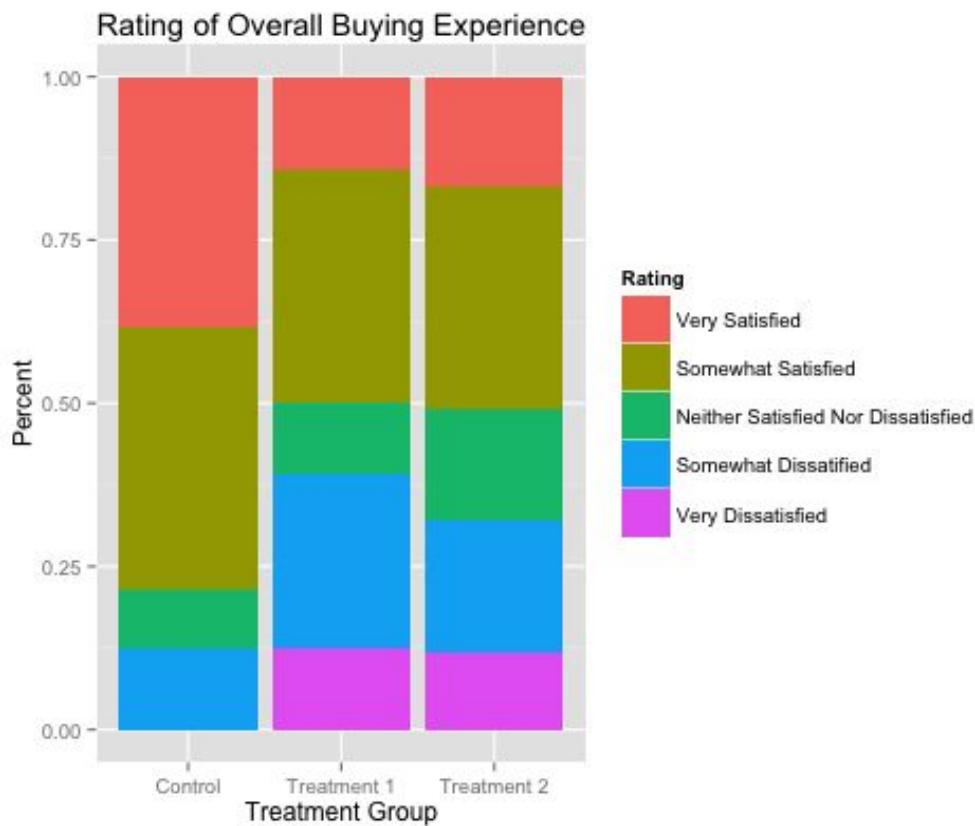
3. The Results

To begin, we present the means and standard deviations of the experiment's outcomes below:

Likert Scale	Control	Treatment 1	Treatment 2
3-Item	11.831 (2.781)	8.429 (3.642)	8.898 (3.502)
5-Item	NA	13.714 (5.352)	14.085 (5.778)

Question	Control	Treatment 1	Treatment 2
Overall, how satisfied or dissatisfied are you with ...			
1. ... this buying experience?	4.046 (0.991)	3.125 (1.308)	3.237 (1.291)
2. ... paying separately for a seat on the plane?	NA	2.714 (1.217)	2.576 (1.329)
3. ... paying separately for your luggage?	NA	2.571 (1.173)	2.610 (1.377)
How likely are you to ...			
4. ... purchase from this airline again?	4.015 (0.944)	2.786 (1.289)	2.949 (1.279)
5. ... recommend this airline to a friend or colleague?	3.769 (1.072)	2.518 (1.321)	2.712 (1.204)

We can already see that the outcomes for Treatment Group 1 and Treatment Group 2 look very similar and that they both look distinctly different from the control group. The following example, the distribution of the answers to the Likert item question, “Overall, how satisfied or dissatisfied are you with this buying experience?” is typical of all other Likert items (see Appendix 6.3 for additional graphs):



As expected from the preview above, we get statistically significant differences in the ATE between the control group and each treatment group, while the ATE between the two treatment groups is not significantly different from zero:

Average Treatment Effect

Likert Scale	Treatment 1 vs. Treatment 2	Control vs. Treatment 1	Control vs. Treatment 2
ATE			
3-Item	0.7952 (0.7156)	-3.1850*** (0.6209)	-2.58400*** (0.5890)
5-Item	0.8477 (1.1361)	NA	NA

*significance at the 95% level **significance at the 99% level ***significance at the 99.9% level

Question	Treatment 1 vs. Treatment 2	Control vs. Treatment 1	Control vs. Treatment 2
Overall, how satisfied or dissatisfied are you with ...			
1. ... this buying experience?	0.2390 (0.2662)	-0.8527 *** (0.2293)	-0.6850 *** (0.2141)
2. ... paying separately for a seat on the plane?	-0.1392 (0.2616)	NA	NA
3. ... paying separately for your luggage?	0.1918 (0.2637)	NA	NA
How likely are you to ...			
4. ... purchase from this airline again?	0.2483 (0.2589)	-1.1990*** (0.2181)	-0.9449*** (0.2128)
5. ... recommend this airline to a friend or colleague?	0.3078 (0.2491)	-1.1333*** (0.2253)	-0.9541*** (0.2099)

*significance at the 95% level **significance at the 99% level ***significance at the 99.9% level

4. Discussion

4.1 Interpretation of results

Our goal was to test whether providing explanations of add-on fees when a customer is being nickel-and-dimed would improve customer satisfaction and loyalty. We found no statistically significant evidence of this. Using either of the Likert Scales to measure satisfaction, customers who received explanations of the fees had higher satisfaction than the customers who saw no such explanations. That is, the ATE was positive on both Likert Scales. But the standard errors were large, making it impossible for us to conclude that these ATEs were not actually equal to zero.

When looking at individual Likert items (the satisfaction and loyalty questions), the results were mixed. Sometimes satisfaction was higher with explanations and sometimes it wasn't. Again, none of these outcomes were statistically significant. At this juncture, we were not able to conclude that explanations to consumers increase their satisfaction and loyalty.

On the other hand, going from an all-inclusive fee to nickel-and-diming brought statistically significant decreases in customer satisfaction and loyalty. The decrease in satisfaction between an all-inclusive fee (Control) versus nickel-and-diming with no explanations (Treatment 1) was 0.85 points, from 4.05 down to 3.20. When explanations of the add-on fees were added (Treatment 2), the decrease wasn't as drastic: 0.69 points, from 4.05 down to 3.36.

For loyalty, the decrease in the likelihood of repeat business with the airline for Treatment Group 1 was 1.2 points, from 4.02 down to 2.82. The decrease for Treatment Group 2 was again smaller, 0.94. When it came to recommending the airline, Treatment Group 2 was 1.1 points less likely to do so than the control group. Treatment Group 1, i.e. those who received the explanations, was only 0.95 points less likely to recommend the airline than the control group.

Overall, when comparing the two treatment groups to the control, providing explanations seems to have bought back some customer goodwill. This is somewhat contradictory to the results of our direct comparisons between Treatment Groups 1 and 2. Our sample size may have been too small to give us the statistical power to detect a benefit from fee explanations when comparing Treatment Groups 1 and 2 directly. We think additional experiments are warranted.

4.2 Challenges

There are several issues that may call into question our results. They are addressed here.

4.2.1 The Lab Environment

There is an argument to be made that a simulation may not generalize to the field. But the fact that flat fees versus nickel-and-diming was found to be preferable, as we would expect, gives us evidence that our experiment set-up is valid in this context.

4.2.2 Generalizability

Because we deliberately chose Mechanical Turk workers located in the US, we only claim that these results generalize to US residents. Furthermore, because of the distribution of workers across our blocks, we should consider that these results may only generalize to infrequent leisure travelers. But this is the group that we were most interested in because we expected them to be the most affected by the treatment. We hypothesized that those who had not bought airline tickets recently might be surprised by the existence of add-on fees and be most put off by them. Thus we thought that they would benefit most from the explanations of the fees.

4.2.3 Click-Through

The same evidence can be used to allay the fear that subjects just clicked through the survey without seriously considering the questions. We would not have gotten such differential results if the subjects did not provided their true opinions. We believe the bonuses paid did entice subjects to give full attention to the task.

4.2.4 Excludability

We argue that our experimental design satisfactorily met the excludability assumption. All subjects who received the treatment saw exactly the same screens giving the explanations of the add-on fees (see Screens 4 and 7 in Appendix 6.2). The survey was conducted over a short 7-day period. But we cannot guarantee that subjects received the treatment under similar

conditions. We do not know under what conditions our Mechanical Turk workers undertook our task, e.g. whether they were at home on their laptops or out-and-about on their mobile phones. We do not believe that any such differences in conditions had a material effect on how workers felt about being nickeled-and-dimed.

4.2.5 Non-Interference

There is some concern that the Mechanical Turk workers may have discussed our survey with each other online through the very active Mechanical Turk forums. Those who had already taken our survey may have tipped off others about what kinds of treatments they received. This could have affected how new survey takers felt about the experience. For example, if a potential survey-taker heard that someone else was charged a flat fee but they were nickeled-and-dimed, then their opinion might be more negative than if they hadn't known that a flat fee was possible. This is a valid concern and one that we were not able to address in our experimental design. We suggest that in future work non-interference be addressed more directly and mitigated as much as possible.

5. Conclusion

5.1 Summary of Findings

We draw two main conclusions from our study:

1. Consumers who are nickeled-and-dimed are more dissatisfied and less loyal than those who are charged all-inclusive fees up front.
2. Within the group of consumers who are nickeled-and-dimed, there is no difference in customer satisfaction or loyalty between those who are given explanations of the add-on fees and those who are not.

We did have statistically significant evidence that Treatment Group 1's results (those who were not given explanations) always showed lower satisfaction than Treatment Group 2's results (those who were given explanations) when compared to the control group. We believe this is enough evidence to warrant further study of whether explanations help to buy back some customer goodwill.

5.2 Future Work

The highest priority for future work would be to deploy this study in the field. One experiment we considered was selling silkscreened t-shirts online. In such a context, we would have had a base price for the t-shirt. The first add-on fee would have been the cost of actually printing the image on the t-shirt. The second add-on fee would have been the shipping charges. The main barrier to implementing this study was our inability to drive an adequate amount of traffic to our t-shirt website in time to collect enough data before the semester deadline. Such a limitation has probably plagued other students' potential experiments in W241. It might be worthwhile for the MIDS program to consider setting up some infrastructure for driving traffic to student-created websites that would help to mitigate this problem.

Another avenue we considered was to convince a real airline to allow us to implement our treatments directly on its website. Given the short time-frame for setting up our study, getting corporate buy-in was beyond our capacity. The nature of the results of our current experiment would compel us to pursue such a strategy in the future.

We also considered that if we found explanations of add-on fees to be effective, then we should study which of the three explanations is the most effective: 1) giving reasons for the existence of the fee; 2) justifying the exact dollar amount of the fee; or 3) explaining the benefit obtained from paying the fee. Furthermore, we could consider what specific styles of language or delivery mechanisms are most effective for engendering customer goodwill. For example, we could test whether specifying all the explanation together at the beginning of the experiment is more effective than having the explanations given at each phase of the treatment.

6. Appendix

6.1 Additional Materials

- Final Dataset Archive:
https://github.com/BerkeleyMIDS/w241-final_project-berkeley-MIDS
- R Code:
<http://rpubs.com/tuhinm/w241MahmudOneto>

6.2 Example of the Purchasing Experience: Treatment 2

<u>Screen 1</u>	<u>Screen 2</u>
<p>In this task, you will simulate buying a plane ticket online.</p> <p>We will give you 600 monetary units (MU\$) that you will use to buy the ticket. At the end of the task, you will be paid according to the number of MU\$s you have left over. The conversion rate from MU\$s to US\$s will be MU\$1 = US\$0.01.</p> <p>For example, if you have MU\$200 left over after you purchase the plane ticket, then you will be paid US\$2.00 for completing this task and filling out a survey.</p> <p>As another example, if you have MU\$150 left over after you purchase the plane ticket, then you will be paid US\$1.50 for completing this task.</p> <p>It will never happen that you will have no MU\$ left over. You will be paid a minimum of US\$1.50.</p>	<p>When you have completed the task, you will be required to fill out a survey asking what you thought of the buying experience.</p> <p>After the survey, there will be a quiz with several questions about the task. You will be paid an additional MU\$5 (US \$0.05) for each correct answer.</p> <p>Please click the >> button below to begin.</p>

Screen 3

At the start of the task, you will see a base price for a plane ticket. You will then be asked to choose a seat on the departing and return flights.

Next, assume you are planning to bring one piece of luggage on your trip. You will be asked to choose whether to check in or carry your luggage onto the plane.

Screen 4

Select Your Seats

SFO ➤ DEN

Please note:

- The airline charges for a seat in order to cover its costs for cleaning them.
- The airline charges MU\$30 for a seat because that is what it costs to pay the cleaners to clean them after each flight.
- The benefit you receive for paying for your seat is a guaranteed clean seat when you board the plane.

Please choose a seat for both your departing and return flights.

Screen 5

Select Your Seats

SFO ➤ DEN

18 06 JUL 2015 08 JUL 2015

TRAVELING DEPARTING RETURNING

MU\$240.00

BALANCE DUE

Departing

San Francisco → Denver

Passenger Selected Seat Cost

Test Subject -- --

TOTAL

MU\$0.00



Screen 6

Returning

Denver → San Francisco

Passenger Selected Seat Cost

Test Subject -- --

TOTAL

MU\$0.00

SFO ➤ DEN

18 06 JUL 2015 08 JUL 2015

TRAVELING DEPARTING RETURNING

MU\$270.00

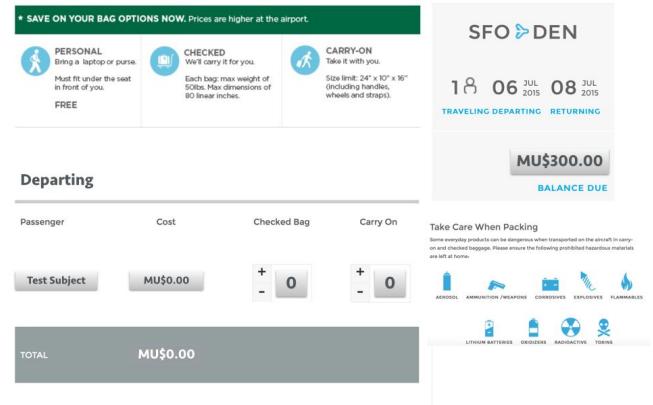
BALANCE DUE



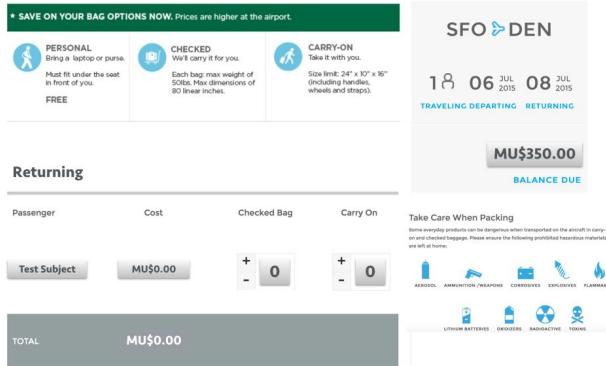
Screen 7



Screen 8



Screen 9



Screen 10

SFO ➔ DEN

18 06 JUL 08 JUL
TRAVELING DEPARTING RETURNING

Departure
Mon Jul 06, 2015
10:10am MU\$ 120.00
San Francisco to Denver Economy

Return
Wed Jul 08, 2015
07:45pm MU\$ 120.00
Denver to San Francisco Economy

Travelers
1 Adult

Options
Test Subject MU\$ 60.00
Seat Fee MU\$ 100.00
Baggage Fee

Summary
Airfare MU\$ 240.00
Options MU\$ 160.00
Taxes and Fees MU\$ 50.00
GRAND TOTAL MU\$ 450.00

6.3 Distribution of Likert Scores by Treatment Group

