The Effect of Showing Restaurant Ratings on Customer Choice

Risheng Guo, Peixuan Liu, Yu Shu, Ta-Wei Wang

12/9/2021

Introduction

The Covid-19 pandemic has greatly hit the restaurant industry. Many local restaurants closed permanently. As chain restaurants keep investing a large amount of money in advertisement, which makes the market become even more competitive and hard to survive for local restaurants. Due to the quarantine policy, people are forbidden to dine-in in restaurants and are forced to order through delivery apps such as Doordash, UberEats, GrubHub, Postmates, etc. We then are interested in whether showing the restaurant's rating on food delivery app UI will affect people's decision when choosing between relatively higher rating, lower recognition local restaurants and lower rating, higher recognition chain restaurants. We expected that without displaying the ratings, people will tend to choose chain restaurants. When ratings were displayed, people would then tend to purchase from higher rated local restaurants but well-known chained restaurants.

Methodology

Survey

To observe the effect of showing restaurant ratings on consumer choice, we decided to choose an online survey because of its simplicity, efficiency, and accuracy. The survey was distributed electronically via social media platforms, such as Instagram and Facebook. We also posted the survey link on the "Boston University Questrom MSBA 2022" Slack channel. Those who had access to the survey and were interested in the main topic of the survey were able to voluntarily participate in the survey. Hence, the sample was selected depending on convenience, availability, and voluntary.

Participants were asked to fill out the Qualtrics survey detailing the experiment and also their gender and age range. The online questionnaire contained 5 multiple choice questions, 3 on different categories of food delivery restaurant: Burger, Coffee, Fried Chicken, and 2 on participants' demographics. We took the screenshots from the real-world food delivery app Doordash and asked people to select their preferred restaurant based on a certain scenario. At the same time, we also created different surveys on Qualtrics containing the same questions about restaurant choice. The only difference between treatment and control was whether the restaurant ratings were shown or not shown on the screenshots.

The survey lasted for 10 days and received a total of 70 valid responses, with the age distribution selected as under 18, between 18 and 24, and over 25. Of these 70 people, 33 identified as male and 37 identified as female.

Randomization

To get a greater statistical power, we decided to randomize the unit of participants. With the help of the online research tools provided by Qualtrics, we were able to randomize participants to be evenly distributed into treatment and control groups. After doing so, those in the control group would answer multiple-choice questions about which restaurant to choose without showing restaurant ratings, while those in the treatment group would experience the same questions given the restaurant ratings.

Pre-Experiment Randomization Check

After randomizing participants into the treatment and control groups, we needed to check if the randomization was done properly. We did this by conducting prop. test to examine whether the randomization proportion was intended, and t.test function to check whether there were differences between treatment and control for gender and age. The results are as below:

```
# proportions between treatment and control
prop.test(data[treatment == 1, .N], data[, .N], p = 0.5)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  data[treatment == 1, .N] out of data[, .N], null probability 0.5
## X-squared = 0.7, df = 1, p-value = 0.40278
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.43389786 0.67408711
## sample estimates:
##          p
## 0.55714286
```

Based on the proportion test, we can't reject the null hypothesis of no difference between two proportions because the p-value is larger than 0.05 and the proportion of 0.5 is within the confidence interval.The treatment and control have the same proportion.

```
# gender difference test treatment vs control
t.test(data[treatment == 1, gender], data[treatment == 0, gender])
```

```
##
##  Welch Two Sample t-test
##
## data:  data[treatment == 1, gender] and data[treatment == 0, gender]
## t = 0.659601, df = 64.16, p-value = 0.51187
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.16275387  0.32321707
## sample estimates:
##  mean of x  mean of y
## 0.56410256 0.48387097
```

```
# age difference test treatment vs control
t.test(data[treatment == 1, age], data[treatment == 0, age])
```

```
##
##  Welch Two Sample t-test
##
## data:  data[treatment == 1, age] and data[treatment == 0, age]
## t = 0.0353506, df = 64.4591, p-value = 0.97191
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.22954543  0.23781673
## sample estimates:
##  mean of x  mean of y
## 0.35897436 0.35483871
```

According to the above t-test results, we have 95% confidence that we can't reject the null hypotheses of no bias in gender and age between treatment and control since the p-values are above 0.05 and the confidence intervals contain 0. We are able to conclude that the randomization proportion was done properly.

Data Analysis

Our collected dataset includes 70 rows and 6 columns. Below are the variable definitions:

```
knitr::kable(variable_data, caption = "Variable Definitions")
```

Table 1: Variable Definitions

question_id	question_content
treatment	'0' for participants in the control group, '1' for participants in the treatment group.
Q1	Which burger restaurant would you choose? '0' for Wendy's, '1' for Boston Burger Company
Q2	Which coffee restaurant would you choose? '0' for Caffe Nero, '1' for Espresso Yourself
Q3	Which fried chicken restaurant would you choose? '0' for KFC, '1' for Coreanos
gender	'0' for male participants, '1' for female participants
age	'0' for participants aged between 18-24, '1' for participants aged above 24

With our data, we can run a number of regressions to analyze the performance of our participants in different ways. We start with a simple regression to analyze the main effects we are looking at the effect of being in treatment (showing ratings). Therefore, we ran the regressions as followed:

Regression: Restaurant Choice on Treatment

```
# Q1, Q2, Q3 treatment vs control
Q1_reg <- feols(Q1~treatment,data = data, se = 'hetero')
Q2_reg <- feols(Q2~treatment,data = data, se = 'hetero')
Q3_reg <- feols(Q3~treatment,data = data, se = 'hetero')
# accumulative percentage
Accumulative_Prob <- feols((Q1+Q2+Q3)/3~treatment,data = data, se = 'hetero')
# individuals choosing all local restaurants
All_Local <- feols(Q1*Q2*Q3~treatment,data = data, se = 'hetero')
etable(Q1_reg, Q2_reg, Q3_reg, Accumulative_Prob,All_Local)
```

```
##                             Q1_reg             Q2_reg             Q3_reg
## Dependent Var.:                 Q1                 Q2                 Q3
##
## (Intercept)     0.5161*** (0.0911) 0.4516*** (0.0907) 0.4194*** (0.0899)
## treatment          0.1505 (0.1190)    0.0099 (0.1216)   0.2473* (0.1181)
## _______________ __________________ __________________ __________________
## S.E. type       Heteroskedas.-rob. Heteroskedas.-rob. Heteroskedas.-rob.
## Observations                    70                 70                 70
## R2                         0.02330             9.8e-5            0.06116
## Adj. R2                    0.00893           -0.01461            0.04736
##                  Accumulative_Prob        All_Local
## Dependent Var.:        (Q1+Q2+Q3)/3         Q1*Q2*Q3
##
## (Intercept)     0.4624*** (0.0454)  0.0645 (0.0448)
## treatment         0.1359. (0.0682) 0.1919* (0.0839)
## _______________ __________________ ________________
## S.E. type       Heteroskedas.-rob. Heterosked.-rob.
## Observations                    70               70
## R2                         0.05261          0.06396
```

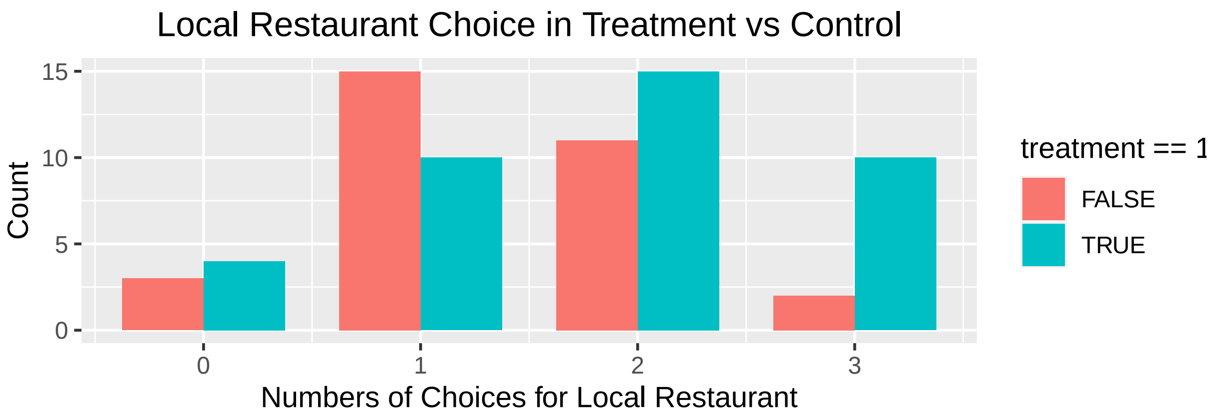
```
## Adj. R2                        0.03868          0.05020
```

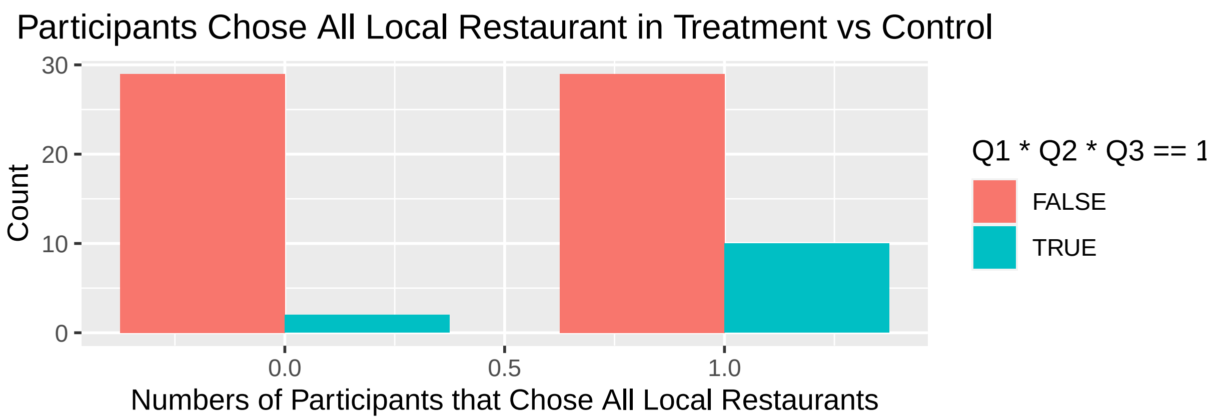
Based on the above results, we can see that without showing the ratings, 52% of people in the control group selected the local restaurant for Q1, 45% of people selected the local restaurant for Q2, and 42% of people selected the local restaurant for Q3. The coefficient of treatment with Q1 is 15%, which means 67% participants chose to order at Boston Burger Company, while the coefficient of treatment with Q2 is around 1%, and the coefficient of treatment with Q3 is 25%. From Q3 regression, we can see that the p-value is lower than 0.05 and is statistically significant; it is strong evidence against the null hypothesis. Ratings on local restaurants have influenced the choice of Q3.

Overall, we have a 13% increase in the choices of local restaurants for all restaurant questions on average. Moreover, if we look at people who chose all the local three restaurants, the control group only has 6% while the treatment group has 20%, which is a huge jump. And from the last regression results, we can see that it is statistically significant, which indicates that ratings have an effect on the outcome. These two regressions are demonstrated with the following two graphs:

```
g1 <- ggplot(data, aes(fill=treatment==1, x=Q1+Q2+Q3)) +
        geom_bar(position="dodge", stat="count", width = 0.75) +
        xlab('Numbers of Choices for Local Restaurant') +
        ylab('Count') +
        ggtitle('Local Restaurant Choice in Treatment vs Control') +
        theme(plot.title = element_text(hjust = 0.5))

g2 <- ggplot(data,aes(fill=Q1*Q2*Q3==1,x=treatment)) +
        geom_bar(position="dodge", stat='count', width = 0.75) +
        xlab('Numbers of Participants that Chose All Local Restaurants') +
        ylab('Count') +
        ggtitle('Participants Chose All Local Restaurant in Treatment vs Control') +
        theme(plot.title = element_text(hjust = 0.5))
grid.arrange(g1, g2, ncol=1)
```





Conditional Average Treatment Effect: Gender

```
# gender difference in treatment and control groups
Q1_gender_reg <- feols(Q1 ~ treatment*gender,data = data, se = 'hetero')
Q2_gender_reg <- feols(Q2 ~ treatment*gender,data = data, se = 'hetero')
Q3_gender_reg <- feols(Q3 ~ treatment*gender,data = data, se = 'hetero')

# regression of accumulative percentage for all three questions on gender
All_gender_reg <- feols((Q1+Q2+Q3)/3 ~ treatment*gender,data = data, se = 'hetero')

etable(Q1_gender_reg,Q2_gender_reg,Q3_gender_reg,All_gender_reg)
```

```
##                        Q1_gender_reg      Q2_gender_reg     Q3_gender_reg
## Dependent Var.:                   Q1                 Q2                Q3
##
## (Intercept)        0.5625*** (0.1277) 0.5000*** (0.1287) 0.3750** (0.1246)
## treatment             0.0257 (0.1773)   -0.2059 (0.1718)   0.2132 (0.1751)
## gender               -0.0958 (0.1842)   -0.1000 (0.1831)   0.0917 (0.1820)
## treatment x gender    0.2349 (0.2420)    0.3968 (0.2411)   0.0474 (0.2404)
## ------------------ ------------------ ------------------ -----------------
## S.E. type          Heteroskedas.-rob. Heteroskedas.-rob. Heteroskeda.-rob.
## Observations                       70                 70                70
## R2                            0.03856            0.05318           0.07566
## Adj. R2                      -0.00514            0.01014           0.03365
##                       All_gender_reg
## Dependent Var.:           (Q1+Q2+Q3)/3
```

```
##
## (Intercept)        0.4792*** (0.0604)
## treatment            0.0110 (0.0930)
## gender              -0.0347 (0.0924)
## treatment x gender   0.2263. (0.1347)
## ___________________ __________________
## S.E. type          Heteroskedas.-rob.
## Observations                       70
## R2                            0.11221
## Adj. R2                       0.07186
```

We then ran a regression on treatment where gender is an interaction term. For example, 56% male participants in the control group chose to order at Boston Burger Company. The correlation coefficient of treatment is positive as 2.6% which implies that male participants increased their willingness to purchase at local restaurants after displaying the ratings. When looking at the whole model's summary, we can conclude that women would be more likely to be influenced by the ratings and change their choices than men. The p-value of each gender regression is all higher than 0.05 which means that we cannot reject the null hypothesis. The high p-values indicate that the results might not be significant.

Conditional Average Treatment Effect: Age

```
# age difference in treatment and control groups
Q1_age_reg <- feols(Q1 ~ treatment*age,data = data, se = 'hetero')
Q2_age_reg <- feols(Q2 ~ treatment*age,data = data, se = 'hetero')
Q3_age_reg <- feols(Q3 ~ treatment*age,data = data, se = 'hetero')

# regression of accumulative percentage for all three questions on age
All_age_reg <- feols((Q1+Q2+Q3)/3 ~ treatment*age,data = data, se = 'hetero')

etable(Q1_age_reg,Q2_age_reg,Q3_age_reg,All_age_reg)
```

```
##                          Q1_age_reg         Q2_age_reg        Q3_age_reg
## Dependent Var.:                  Q1                 Q2                Q3
##
## (Intercept)     0.5500*** (0.1146) 0.4500*** (0.1146) 0.3000** (0.1055)
## treatment          0.1300 (0.1495)    0.0300 (0.1540) 0.4600** (0.1374)
## age               -0.0954 (0.1924)    0.0045 (0.1924)  0.3364. (0.1829)
## treatment x age    0.0583 (0.2523)   -0.0560 (0.2572) -0.5964* (0.2452)
## _______________ __________________ __________________ _________________
## S.E. type       Heteroskedas.-rob. Heteroskedas.-rob. Heteroskeda.-rob.
## Observations                    70                 70                70
## R2                         0.02788            0.00147           0.14278
## Adj. R2                   -0.01630           -0.04392           0.10381
##                        All_age_reg
## Dependent Var.:       (Q1+Q2+Q3)/3
##
## (Intercept)     0.4333*** (0.0600)
## treatment         0.2067* (0.0834)
## age                0.0818 (0.0905)
## treatment x age   -0.1980 (0.1444)
## _______________ __________________
## S.E. type       Heteroskedas.-rob.
## Observations                    70
## R2                         0.08042
```

```
## Adj. R2                      0.03862
```

From the regression results above where age is the interaction term, we found that the treatment has a much higher effect on the younger age group (under 24 years old). For participants who are under 24 years old in the treatment, the willingness increased by 21%. In comparison, for those who are above 24 years old in the treatment, the willingness decreased. From the last regressions, we can see the p-value is lower than 0.05 so it is statistically significant. We can reject the null hypothesis and conclude that participants who are under 24 years old would be more affected by ratings of restaurants.

Limitations

A primary limitation of the study was the sample used. The sampling method we used was convenience sampling, which means the survey was taken by participants who were available and easy to reach. Even though we randomly assigned participants to the control and treatment groups, the respondents of the study were not able to represent all the target audiences due to a small sample size.

In addition, the type and number of restaurants in our survey can also be more diverse. We only included three categories: Burger, Coffee, Fried Chicken, and a total of 6 restaurants in our survey. Thus, the external validity of our result can be improved in future research if we include more types and numbers of restaurants in the survey. Moreover, some chain restaurants we choose in our survey are not well known to all of our participants due to participants' different backgrounds. For example, even though Caffe Nero is famous in the Boston area, participants from New York may not know Caffe Nero is a chain store and treat it evenly with un-well known local restaurants.

Conclusion

In conclusion, the result meets our expectations. We didn't find a statistically significant effect of ratings on restaurant choice. However, the results suggest that when the ratings are hidden, people tend to choose famous chain restaurants. When ratings are shown, people then are more likely to purchase from highly-rated local restaurants instead of well-known chain restaurants. Moreover, we found that the treatment is more effective for females and the younger generation. That is, the gap between treatment and control groups for women is bigger than for men. Individuals aged under 24 also yield better results than others, possibly because they look at more information on the application when making decisions of ordering food. Due to the limitations of this study, future research is needed to further analyze the effect of the restaurant ratings.

Appendix

```r
knitr::kable(questions, caption = "Survey Questions")
```

Table 2: Survey Questions

ID	Question
Q1	If you want to order burgers, which restaurant would you choose?
Q2	If you want to order coffee, which restaurant would you choose?
Q3	If you want to order fried chicken, which restaurant would you choose?
Q4	What is your gender?
Q5	What is your age range?

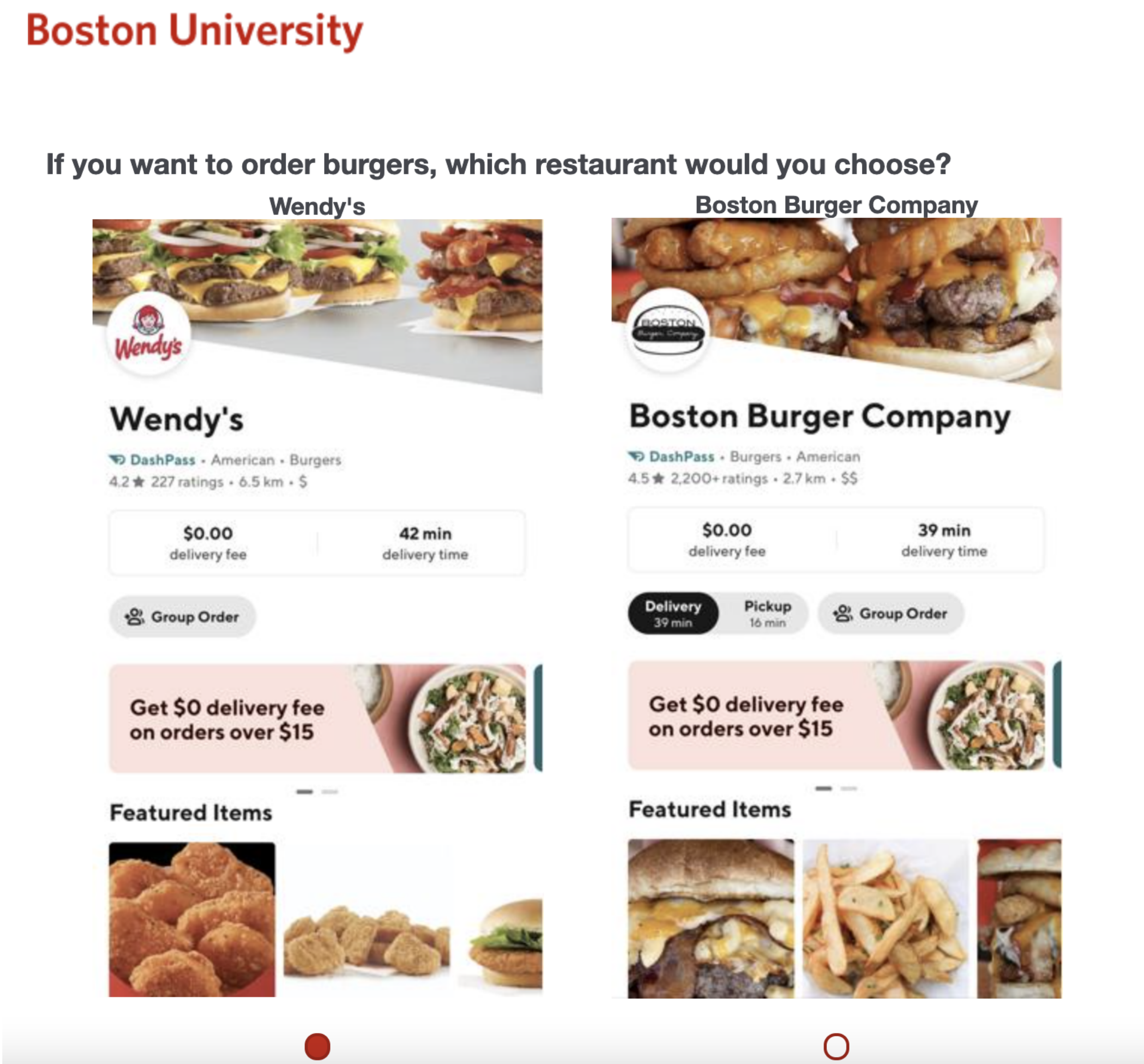


Figure 1: Treatment Group Survey

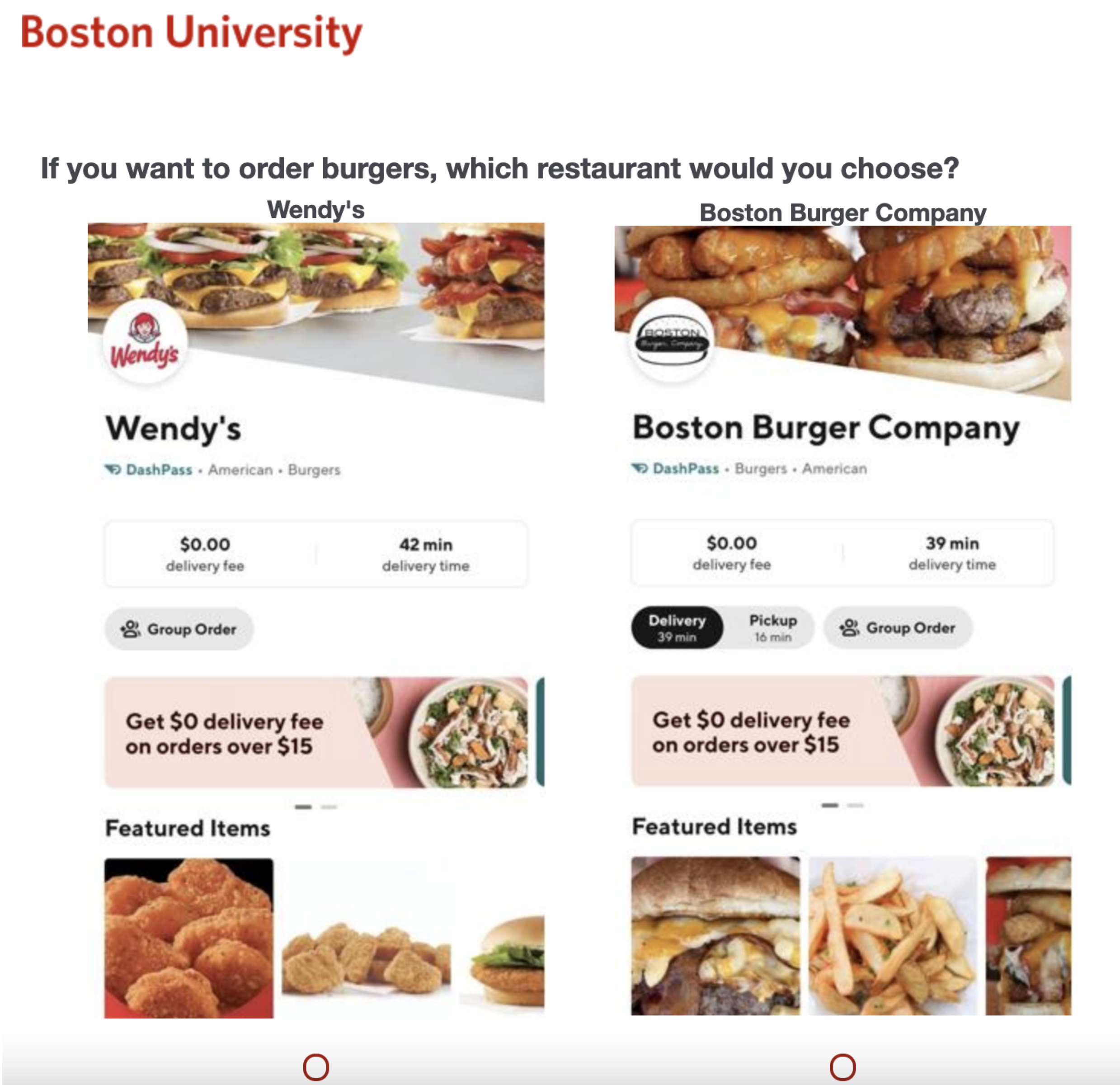


Figure 2: Control Group Survey