

# The Effects of Branded Coffee on Perceived Taste Satisfaction

Fall 2024 - W241 Final Project

Amy Zhang, Chris John, Jenna Farac, Simran Gill

## Abstract

Coffee is a consumable that tends to be associated with loyal and discerning consumers. While efforts have been made to democratize the scoring of coffee beans and provide transparency on bean quality to consumers, the coffee industry's largest players still invest heavily into branding. Does branding on a coffee's package actually improve a consumer's perceived sense of taste satisfaction? In order to evaluate this, we study how a participant scores the taste of a generic and specialty coffee with and without labels presented.

We do this by splitting our colleagues, friends, and family into a control and treatment group to share their score for the taste of the two different coffee brands with the treatment being branding of the coffee. Though we observed that the specialty brand had an increase in satisfaction, while the generic brand had a decrease in satisfaction after the consumer was exposed to branding information, our results were not statistically significant meaning that showing the brand does not have a significant impact on taste scoring in this case. For the day to day consumer, this study's findings may encourage them to revisit any pre-existing brand loyalties and for the brand these findings may encourage them to invest in unique and memorable branding.

## Introduction

### Previous Research

The treatment in our experiment involves presenting branded coffee to our participants and expecting a change in taste score in comparison to coffee that is presented as being non-branded. We expect to see change in the behavior of participants based on well-established studies of marketing and psychology, particularly in relation to the expectation-confirmation theory (Oliver, 1980), that suggested consumers develop expectations based on the available information provided to them before consuming a product, allowing these expectations to influence their perceived quality of that particular product. If we are allowing participants in our study to have exposure to the branding of said coffee, particularly involving well-known, high quality brands, the expectation of quality is expected to increase a participant's perceived taste in comparison to a coffee that appears to be non-branded.

### Research Question:

Each coffee brand chooses to market itself in different ways, and invest a varying range of money into this effort. While some brands are more well known than others, all brands must decide on how to stand out on a shelf in a grocery store to a consumer. The packaging of the product will typically display

objective details about the product, but also be designed in an intentional way to differentiate itself and provide a unique user experience. For example, as we will see in this study, the packaging for coffees can be entirely different materials (glass, paper, plastic), and include aesthetic graphics.

The goal of our field experiment is to identify the impact of branding on a consumer's taste preferences, specifically for cold brew coffee. The scope of this impact is not intended to be limited to any specific consumer group, but be broadly applicable.

## Hypothesis

Null Hypothesis: The treatment group will yield the same taste test score for branded coffee in comparison to non-branded coffee.

Alternative Hypothesis: The treatment group will yield a difference in taste test score for branded coffee in comparison to non-branded coffee

The prediction of increased taste satisfaction for those in the treatment group comes from the relationship people have with brands. Once people are aware of the brand, then they might be likely to use other factors such as their loyalty, brand message, and design when scoring the taste. Because of the additional information participants have, we are expecting a change in taste scores.

## Experimental Design

### Experiment Overview

To test this hypothesis, the experiment implemented will follow between-subjects design, for which people recruited will see either treatment *or* control. We will then measure the impact for each of the two brands separately, and observe if the scores with and without branding are different. No scores will be measured before we administer treatment. After the treatment is administered, we will measure the participants' satisfaction with the coffee by asking them to score it on a scale of 1-7, as well as collect other covariates for our analysis. All participants will have their scores noted at this time, and no further details are required at a subsequent time.

### Treatment

The treatment of this experiment revolves around the exposure of branding during a coffee taste test. Participants in the treatment group were presented two coffee samples, the lesser brand quality being the "Good & Gather" coffee and the higher branded quality being the "Chameleon" coffee. The control group, on the other hand, received the same coffee samples as the treatment, just with generic labels on them instead. "1" represented the Good & Gather brand coffee, while "2" represented the Chameleon brand coffee. This manipulation is designed to measure the impact of branding on the individual's satisfaction scores for the two types of coffee.

In regards to potential deviations that were present in our experiment, we believe certain participants being unaware of the branding significance between Chameleon coffee vs. Good & Gather coffee played a part in our expectations for differences in scores. If an individual isn't familiar with either brand, they could attribute both brands having similar effects on their perceptions of taste, which can definitely play a role in how much we estimate it to play a part in their scoring of the coffees. To account for this, we asked participants whether they knew each of the previous brands, and if so, what are their pre-existing opinions of them.

We also acknowledge that potential spillover may have occurred throughout the sampling process. Many of our participants had relationships with each other, whether that be from the same workplace, same friend group, etc. Understanding that opinions about each score may have been passed along to individuals that may have not participated in the experiment is integral to our assumptions process. Comments about the taste of the coffee prior to a participant tasting the coffee for themselves is an example of how spillover could have potentially influenced a number of our scores received.

## Project Timeline

Due to practical constraints, the participants in our experiment were limited to people the team encounters naturally. For each team member, we stratified our sample collection so each team member would collect at least 20 responses. Depending on the member, they identified a suitable occasion to execute the experiment where a large enough sample size would be possible. For example, some team members executed the experiment at the office, while others hosted events in their home.

For each team member, all participants at each event were mandated to participate, and as the participant list was available in advance, the treatment and control groups were pre-determined through random number generation.

10/23/24	10/24/24-11/10/24	11/11/24 - 11/28/24	11/29/24 - 12/8/24	12/20/24
<b>Project Setup:</b> Determining what data we wanted to collect and setting up our Google Form	<b>Recruitment and Randomization:</b> Reaching out to potential participants and planning for when we would collect data. Randomizing based on our attendees	<b>Data Collection:</b> Hosting in office or at home coffee sampling in person	<b>Data Analysis:</b> Analyzing our results	<b>Participant Debrief:</b> Sharing our findings with our participants

## Measurement tool - Survey/Questions

To conduct our experiment, we used Google Forms to collect each participant's responses. After collecting our data, the results were formatted into a Google Sheet which we were easily able to upload and clean upon starting our analysis.

We had two forms for this experiment, one for the treatment group and one for the control group. All of the questions included in the form were the same for both groups, minus the first question where participants were asked to rank the taste of the coffees. For this question, the control group referred to the coffees as “Coffee 1” and “Coffee 2” while the treatment group referred to them as the “Good & Gather Coffee” or the “Chameleon Coffee.” We kept most of the questions short and as multiple choice questions to avoid any participants dropping out due to the experiment taking too long. Participants were asked to try the coffees, and answer the first three questions before scrolling down any further to prevent the control group from seeing the brands utilized to conduct the experiment before rating the coffees. The questions included in our survey were:

1. What is your name?
2. How would you score the taste for Coffee 1/the Good & Gather coffee? (7 being the best, 1 being the worst)
3. How would you score the taste for Coffee 2/he Chameleon coffee? (7 being the best, 1 being the worst)
4. What is your age?
5. What is your gender?
6. How often do you drink coffee?
  - a. Options provided were: Never, Occasionally (Up to 1 time a week), Sometimes (a few times a week), Often (almost everyday)
7. If you drink coffee, do you usually take it hot or cold?
  - a. Options provided were: hot, cold, no preference
8. If you drink coffee, do you usually prefer it sweet or not sweet?
  - a. Options provided were: sweet, not sweet
9. Do you have any medical conditions that may impact your sense of taste?
  - a. Options provided were: yes or no
10. Are you aware of the brand Good & Gather? If yes, do you view them positively, negatively, or neutral?
  - a. Options provided were: Yes & positively, Yes & neutral, Yes & negative, or No
11. Are you aware of the brand Chameleon? If yes, do you view them positively, negatively, or neutral?
  - a. Options provided were: Yes & positively, Yes & neutral, Yes & negative, or No

## Distribution of Coffee

To present our two coffee samples to our participants, we first mixed the cold brew concentrate as instructed on the bottles: one part concentrate and one part water, and more specifically 1.5 oz of each. We opted to use water as a mixer instead of milk to avoid any potential allergies and to focus on the taste of the coffee consistently rather than allowing the milk to affect the participants ranking. For example, oat milk would make the coffee creamier or cashew milk might provide a nutty taste. Both coffees were presented in small standard paper cups marked as “Coffee 1” and “Coffee 2”. For the experimental group, the presentation of the coffee remained the same with the addition of the bottle of concentrate being placed behind the paper cups so the participants could take note of the branding. More specifically, they may notice that Good & Gather had a more basic, dark brown, plastic bottle while Chameleon had a green glass bottle with a chameleon on the bottle.

Participants were asked to come into a designated room one by one, in order to avoid any discussion amongst participants and potential spillover effects regarding the branding. We asked participants to try Coffee 1, then take a sip of water to cleanse their palate, try a sip of Coffee 2, and then score the coffees. Upon completion of the survey, participants were asked not to discuss the experiment with anyone else present in order to preserve the integrity of the experiment. When determining the order of participants, the control group was tested first. This was important to ensure that if participants did discuss the experiment, they would not discuss specific traits of the branding shown because they had not yet seen it.

## Enrollment and Recruitment Process

Our enrollment process was very personalized and informal, recruiting individuals by word of mouth, text messages, and phone calls. Some team members strategically decided to recruit participants at large gatherings (family events, Thanksgiving parties, workplace meetings, etc.) to recruit bunches of participants at once. This method allowed for efficient sample collection.

Our participant pool consisted of primarily family members, friends, and co-workers, leveraging existing relationships to encourage participation. This approach allowed us to achieve participation commitments rather easily, as recruiting from known social circles helped reduce attrition and potential no-shows.

Although convenient, we do acknowledge the potential bias that may exist from our approach. Participants are not stratified from multiple samples, and we understand that recruiting from many similar circles will introduce a lot of similar results. Similar demographics from this process will influence our results, and looking back, possibly conducting a more balanced, diverse recruitment process may have helped mitigate some of that selection bias that may have been introduced.

## Script in our Experimental Process

We will rely on self-reported satisfaction scores that are provided by the participants for both coffee brands. Using a scale between 1 - 7, we will ask each of our participants to provide a score of their perceived taste. At the beginning of the trial, a team member will ask the participant: “We’re conducting a taste test for two different coffees. Please rate the taste of these two coffee samples on a scale of 1 to 7. A score of 1 will indicate you were not at all satisfied with the taste, while a score of 7 will indicate you were extremely satisfied.” After the participant has tried each of the coffees, the team member will ask the following depending on whether they are in the Control or Treatment group:

<u>Control Group Script</u>	<u>Treatment Group Script</u>
<ul style="list-style-type: none"><li>- “For Coffee A, what’s your score?”</li><li>- “For Coffee B, what’s your score?”</li></ul>	<ul style="list-style-type: none"><li>- “For Good&amp;Gather, what’s your score?”</li><li>- “For Chameleon, what’s your score?”</li></ul>

Translation was provided based on the participants' requests. Some of our participants requested translation to better understand the task and what was required.

## Power Calculation

Our pre-experiment assumption was that the effect size would be in the range of  $(-0.63, 0.581)$  based on a previous experiment done for wine, which would require an experiment size of 100 to 80, respectively, to obtain a 80% power level with an additional assumption of a standard deviation of 1.

Through our convenience sample, we were able to collect 91 samples (39 in treatment, 52 in control). Given this, we anticipate that this experiment will be able to generate a power level of 78.3% for an increase in ATE of 0.581, and a power level of 84.5% for a decrease in ATE of -0.63. These calculations are done using the same assumptions as in our previous power analysis.

## Randomization

To conduct our randomization, we first started with our intended list of participants. For each participant we generated a random number between 0 and 1. We then sorted our list in ascending order and divided the list in half, assigning the first half of participants to the control group, and the second half of participants to the treatment group.

We opted not to stratify our sample for two reasons, the first being that we did not know the age, coffee preferences, and brand awareness of all of our participants. Additionally, we were not confident about the impact that gender would have on the range of coffee scores, but opted to collect the data to explore a potential relationship. The second reason we opted not to stratify our sample was because we conducted a convenience sample based on people we were able to interact with in person and therefore opted to maximize our participants rather than narrow our scope to stratify the sample.

## Data

The data consists of responses from the two Microsoft Form Surveys who were asked to participate in the taste test. The two outcome of interest measures produced by the experiment are:

Outcome of Interest Variables	Definition
<b>Good &amp; Gather Score</b>	Represents the taste test score for the “Good & Gather” coffee brand, ranging from 1 (lowest score) to 7 (highest score)
<b>Chameleon Score</b>	Represents the taste test score for the “Chameleon” coffee brand, ranging from 1 (lowest score) to 7 (highest score)

The participants are provided the option to only select one number from 1 to 7, so there are no floating points. The difference in scores for the control and treatment group will help assess if branding has any effect on the perceived taste for the participants. We are choosing this outcome because it is an important factor when it comes to consumer decision making.

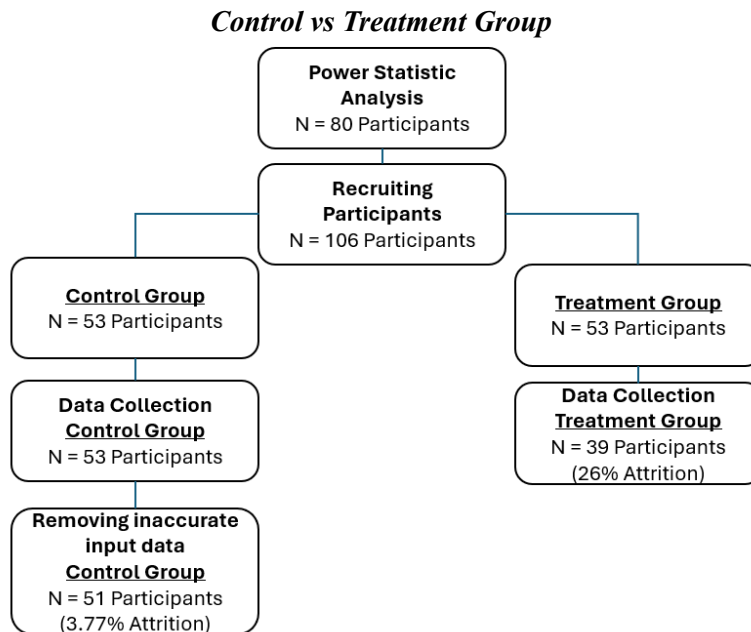
When conducting the surveys, we also asked participants to share some key covariant variables that we can use for this analysis and modeling. This helped us get a better understanding of the participants in terms of demographics and general opinion on coffee.

<b>Covariant Variables</b>	<b>Definition</b>
<b>Age</b>	Age of the participant
<b>Gender</b>	Gender of participant (Female, Male, Other)
<b>Medical Condition</b>	If the participants has any medical condition that affects their taste
<b>Frequency of drinking coffee</b>	Tells is the frequency of drinking coffee:"Never", "Occasionally (up to 1 time a week)", "Sometimes (a few times a week)", "Often (almost every day)", "Every day"
<b>Hot or Cold</b>	If participant prefers or or cold coffee
<b>Sweet or not sweet</b>	If participants prefers sweet or not sweet coffee
<b>Good &amp; Gather Awareness</b>	Indicating whether participants are aware of the Good & Gather brand and their thoughts.
<b>Chameleon Awareness</b>	Indicating whether participants are aware of the Chameleon brand and their thoughts.

## Data Completeness

From the power statistic analysis, we would need at least 80 participants for this experiment to have a meaningful treatment effect. Initially, the plan was to have 40 participants in both control and treatment. For those of us who were conducting the experiment in-office, we collected data on the control group for one day. Then on another day the data from the treatment group was collected. Unfortunately, some participants from the treatment could not make it, which resulted in a control group with a sample size of 53 and treatment group of 39.

Through this exploration, we found that two participants from the control group didn't provide an actual integer for their age. For example, for one instance, a participant put in "old" for age. We removed these two cases from the control group which resulted in a sample size of 51. In total, we had a sample size of 89 which is above what our goal for the experiment was.



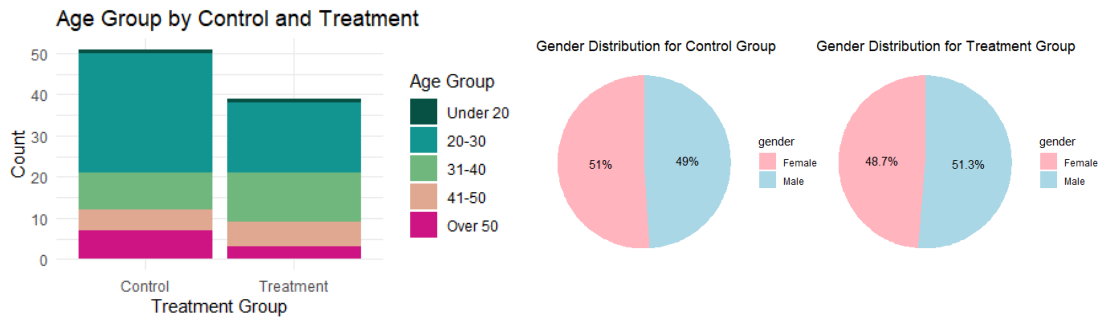
## Exploratory Data Analysis

### Demographics

After finalizing the data, we did exploratory analysis to get a better understanding of the demographics of our participants. When it comes to age, we bucked the ages into 5 groups (Under 20, 20-30, 31-40, Over 50) to help with summarizing the data and doing a comparison between the control and treatment group. From the graph titled, “Age Group by Control and Treatment”, we have more participants in the age range of 20 - 30 compared to any other age bucket. This is consistent in both groups, however we do notice the buckets are not completely equal between the groups. When it comes to gender, the control group has 51% female and 49% male. For the treatment group it is 48.7% female and 51.3% male.

Because the participants were randomly assigned to the control and treatment. The difference in the demographic distributions are likely due to chance rather than bias. Randomization helps with confounding variables, even for those variables that are not explicitly measured. Considering this, we do not expect the difference in demographics to have an impact on the results. We have also included a covariate imbalance analysis further in the report.

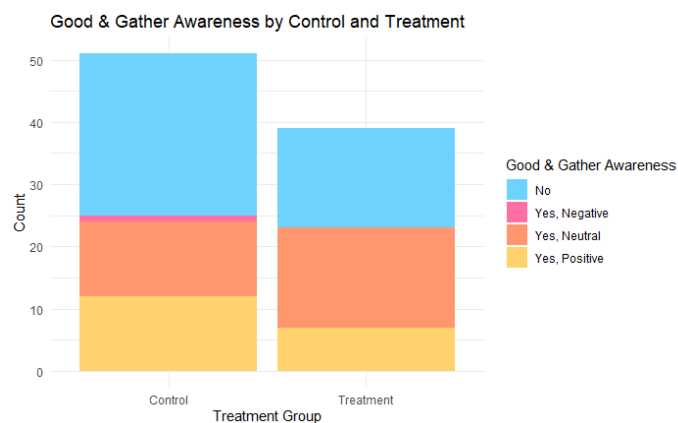




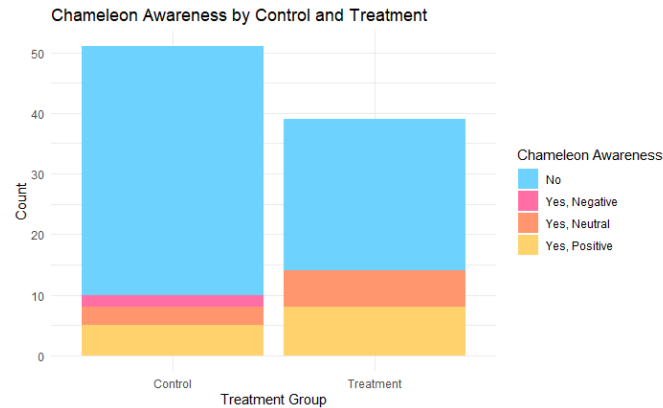
## Coffee Brand Awareness

When it comes to rating the taste of the coffee, we were interested in knowing if participants were aware of the Good & Gather or Chameleon brand in general and what their views are. Participants were asked to share if they were familiar with each brand and their overall view. They could choose from four options: “No”, “Yes, Negative”, “Yes, Neutral”, or “Yes, Positive.”

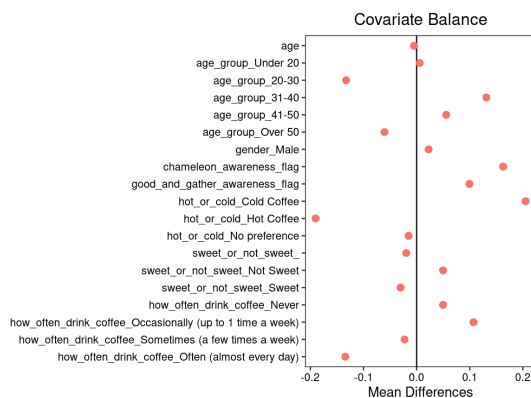
The bar chart shows that participants had some familiarity with the Good & Gather brand. This aligns with our expectations, since the generic Target brand is more recognizable. Both control and treatment groups did express they have more of a neutral perception of the brand even though they have some awareness.



For the Chameleon brand, most participants were not familiar with the brand. If they were, it was mostly with a Neutral or Positive view. Since we see the Chameleon brand to be more of a specialty coffee brand, our expectations were that most participants will not be familiar, but if they are, they would have more of a positive view.



## Covariate Balance Analysis



To confirm the success of our randomization, we conducted a covariate balance analysis using Cobalt, with results shown on the left. We found no severe imbalances, but do have a few moderate imbalances that exceed -0.1 or 0.1 to be aware of as we interpret our results:

- Cold vs Hot Coffee Preference
- Chameleon Awareness
- Age Group 20-30 and 31-40
- Often Coffee Drinkers

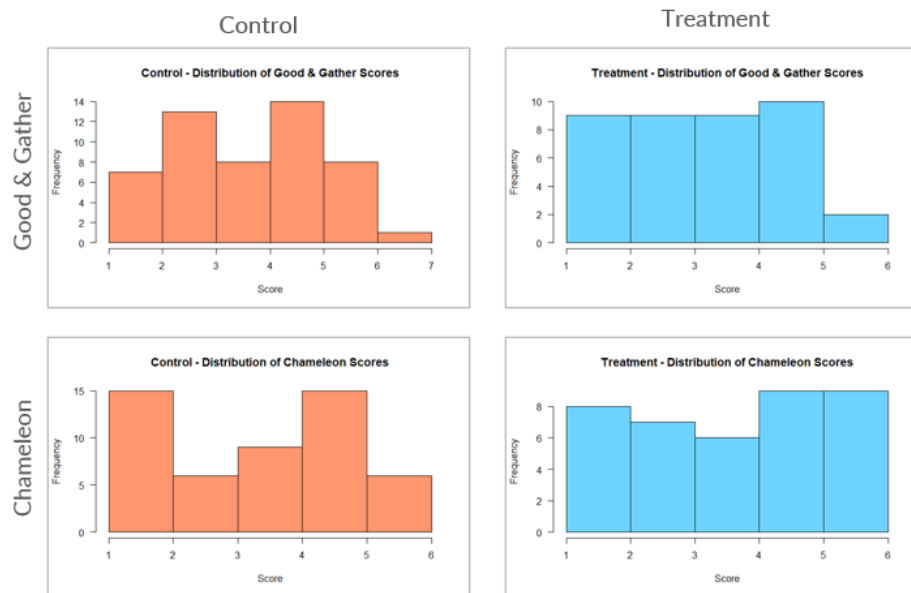
Overall, we do not expect most of these covariates to be very strongly related to the taste score of the coffees, aside from the often coffee drinkers. In general, people who don't drink coffee often may not enjoy this taste, especially since we are not including any sweetener or using milk as a mixer. This is one factor that we will keep in mind throughout the analysis.

## Results

For this experiment, we wanted to observe the difference in treatment effect separately for the two coffee brands. This is because we are interested in seeing if the rating of the coffee changes based on the brand. These two brands have different reputations, and it would be interesting to see how that impacts rating for the participants.

## Score Overview

Below are the distributions of taste scores for each brand for treatment and control. For control, where brands are not available, Good & Gather has an almost normal distribution, with a mean of 4.06 and standard deviation of 1.49. Chameleon has more of a bimodal distribution, with a mean of 3.76 and standard deviation of 1.53. After branding is exposed for the Treatment group, we see that Good & Gather has a lower mean taste score of 3.62 with a standard deviation of 1.33, while Chameleon has a higher mean taste score of 4.05 with a standard deviation of 1.57.



## Model Exploration

Since we collected various covariates, we wanted to identify which ones had a statistically significant impact on the model's output. We began with two baseline linear regression models, one for each coffee brand, and tested different covariates to evaluate how they impacted the treatment effect.

### Good & Gather

When it comes to the Good & Gather, we are starting off with examining the relationship between treatment and the good\_and\_gather\_score. The good\_and\_gather\_score is the rating the participants shared for the taste of the coffee that is between 1- 7.

#### Baseline Model:

From the results in Table 1, we can see that the intercept is 4.059, this is the predicted score for the control group. The overall treatment effect estimate is -0.443. This tells us there is a decrease in the rating when the brand is displayed. However, the treatment is not statistically significant because it has a p-value of 0.147, which is higher than our threshold of 0.05. We also have an R-square value of 0.024, which tells us the model explains very little variance for good\_and\_gather\_score.

Table 1: Baseline Model for Good and Gather Score

	Dependent variable:
	good_and_gather_score
Treatment	-0.443 p = 0.147
Constant	4.059**** p = 0.000
Observations	90
R <sup>2</sup>	0.024
Adjusted R <sup>2</sup>	0.013
Residual Std. Error	1.422 (df = 88)
F Statistic	2.148 (df = 1; 88)
Note: *p<0.1; **p<0.05; ***p<0.01	

### Model Exploration:

Given the results, we still wanted to continue to explore covariates contributions to the model. We started off with age and gender. Tested different covariates and the one that was more promising was the chameleon\_awareness. To test if other covariates have any impact in improving the model, we performed an ANOVA test. From the results, we found that the p-value is 0.447, which has strong evidence that adding this good\_and\_gather\_awareness to the model has no meaningful effect on explaining the variance in good\_and\_gather\_score.

### Good & Gather - ANOVA Test

#### Analysis of Variance Table

```
Model 1: good_and_gather_score ~ treatment + log(age) + gender + chameleon_awareness
Model 2: good_and_gather_score ~ treatment + gender + log(age) + chameleon_awareness +
good_and_gather_awareness
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	83	140.57				
2	80	136.00	3	4.5703	0.8961	0.447

### Fine Tune Model:

Finding that chameleon\_awareness was more significant was interesting because how someone views one brand might sway their rating for the other coffee brand in question.

From the linear regression summary (model 2) in Table 2, we can see when the participant is aware of Chameleon and has a Negative view, the p-value is 0.05 which is statistically significant. When the participant has a Neutral view, the p-value is 0.053, which is marginally significant. After applying these covariates, we see that the treatment decreases even further to -0.568 with a p-value of 0.051 which is statistically significant. The model has a R-square value of 0.229, which indicates a moderate explanatory power with the overall model being statistically significant based on F-statistic.

Table 2: Regression Results for Good and Gather Score

	<i>Dependent variable:</i>	
	good_and_gather_score	
	(1)	(2)
Treatment	-0.443 p = 0.147	-0.568* p = 0.051
log(Age)		1.303*** p = 0.004
Gender		0.385 p = 0.168
Chameleon Awareness (Negative)		-1.887** p = 0.050
Chameleon Awareness (Neutral)		0.932* p = 0.053
Chameleon Awareness (Positive)		-0.617 p = 0.128
Constant	4.059**** p = 0.000	-0.524 p = 0.731
Observations	90	90
R <sup>2</sup>	0.024	0.229
Adjusted R <sup>2</sup>	0.013	0.174
Residual Std. Error	1.422 (df = 88)	1.301 (df = 83)
F Statistic	2.148 (df = 1; 88)	4.116*** (df = 6; 83)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

## Chameleon

Following the same process like we did for Good & Gather, we are starting off with examining the relationship between treatment and chameleon\_score.

### Baseline Model:

From the results, we have an intercept of 3.765, which is the predicted score for the control group when rating the coffee. The treatment effect estimate is 0.287, this tells us there is an increase in rating for the Chameleon coffee when the brand is displayed. However, the treatment effect is not statistically significant because the p-value is 0.387 which is higher than our 0.05 threshold. We also have an R-square value of 0.009 which tells us the model explains very little variance for chameleon\_score. The F-statistic also tells us the model lacks explanatory power because the p-value is insignificant.

Table 3: Baseline Model for Chameleon Score

	<i>Dependent variable:</i>
	chameleon_score
Treatment	0.287 p = 0.387
Constant	3.765**** p = 0.000
Observations	90
R <sup>2</sup>	0.009
Adjusted R <sup>2</sup>	-0.003
Residual Std. Error	1.549 (df = 88)
F Statistic	0.757 (df = 1; 88)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

### Model Exploration:

Given the results, we still tested different covariates to see if any of them had statistically significant improvement to the model. We again started off with adding in age and gender to test different covariates. What was surprising was that chameleon\_awareness was statistically significant in improving the model. We again wanted to test if adding in good\_and\_gather\_awareness would have a statistically significant improvement to the model by using the ANOVA test. From the results, we can see the p-value is 0.267, which has strong evidence that adding this good\_and\_gather\_awareness to the model has no meaningful effect on explaining the variance in chameleon\_score.

### *Chameleon - ANOVA Test*

#### Analysis of Variance Table

```
Model 1: chameleon_score ~ treatment + log(age) + gender + chameleon_awareness
Model 2: chameleon_score ~ treatment + log(age) + gender + chameleon_awareness +
good_and_gather_awareness
Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      83 164.74
2      80 156.86  3    7.8875 1.3409 0.267
```

### Fine Tune Model:

For our final Chameleon model, when the participant had a Neutral or Positive view of the Chameleon brand, the p-value are both under 0.05 indicating there is statistical significance. After applying these covariates, the treatment effect decreases to 0.018, but the p-value is 0.953 indicating the treatment does not have a meaningful impact on chameleon\_score. The model has a R-square of 0.226 with a F-statistic that is significant, indicating the model as a whole is statistically significant. Overall, even though treatment does not have a significant effect on the score, chameleon brand awareness does.

Table 4: Regression Results for Chameleon Score

	<i>Dependent variable:</i>	
	chameleon_score	
	(1)	(2)
Treatment	0.287 p = 0.387	0.018 p = 0.953
log(Age)		1.196** p = 0.012
Gender		0.276 p = 0.360
Chameleon Awareness (Negative)		1.250 p = 0.226
Chameleon Awareness (Neutral)		1.192** p = 0.023
Chameleon Awareness (Positive)		1.675**** p = 0.0003
Constant	3.765**** p = 0.000	-0.760 p = 0.644
Observations	90	90
R <sup>2</sup>	0.009	0.226
Adjusted R <sup>2</sup>	-0.003	0.170
Residual Std. Error	1.549 (df = 88)	1.409 (df = 83)
F Statistic	0.757 (df = 1; 88)	4.043*** (df = 6; 83)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

### Combined Data:

We also created a combined model where we stack the scores for both Good & Gather and Chameleon to see what the treatment effect is overall. When we do this, we can see from Model 1 in table 5 that the average treatment effect is -0.077. This means that after the member is aware of the brand, they are rating the coffee lower. However, the p-value is 0.735 tells us the treatment is insignificant overall and that seeing the brand does not have a significant impact on the taste score provided.

In order to account for whether or not the treatment effect depends on the coffee brand, we decided to add in the interaction term (Treatment x Brand) as well as an interaction term for brand and Chameleon awareness as well as Good & Gather awareness. The treatment effect for Chameleon is -0.020 with a p-value of 0.950 indicating the treatment had no significant impact on the score for Chameleon. The Treatment effect with the interaction term for brand indicates that compared to Chameleon, Good & Gather's scores are 1.082 higher with a p-value of 0.003, which suggests that the treatment effect does differ for the two brands. Looking at awareness for both brands and which relationships are significant, those with a neutral view of Good & Gather rate Good & Gather -1.099 points lower than Chameleon. Overall, though the effect of branding was not significant for either coffee brand, our model points out that the awareness of one brand has the ability to influence the score of another brand.

Table 5: Regression Results		
	<i>Dependent variable:</i>	
	score	
	(1)	(2)
Treatment	−0.077 p = 0.735	−0.020 p = 0.950
Brand (Good and Gather)		1.082*** p = 0.003
Treatment x Brand		1.164**** p = 0.0004
log(Age)		1.467 p = 0.142
Chameleon Awareness: Negative		1.249** p = 0.021
Chameleon Awareness: Neutral		1.477**** p = 0.001
Chameleon Awareness: Positive		1.085 p = 0.434
Brand x Chameleon Awareness: Negative		0.488 p = 0.164
Brand x Chameleon Awareness: Neutral		0.543 p = 0.183
Brand x Chameleon Awareness: Positive		−0.482 p = 0.270
Good and Gather Awareness: Negative		−3.582** p = 0.012
Good and Gather Awareness: Neutral		−0.281 p = 0.709
Good and Gather Awareness: Positive		−1.994*** p = 0.002
Brand x Good and Gather Awareness: Negative		−2.082 p = 0.287
Brand x Good and Gather Awareness: Neutral		−1.099** p = 0.026
Brand x Good and Gather Awareness: Positive		−0.738 p = 0.201
Constant	3.910**** p = 0.000	−0.735 p = 0.524
Observations	178	178
R <sup>2</sup>	0.001	0.248
Adjusted R <sup>2</sup>	−0.005	0.173
Residual Std. Error	32 1.494 (df = 176)	1.355 (df = 161)
F Statistic	0.115 (df = 1; 176)	3.320**** (df = 16; 161)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

## Model Comparison and Outcomes

To summarize the results of the regression models for the two coffee brands, we see a notable difference in how the treatment affects the participants' scores for taste. When it comes to Good & Gather, the treatment group seems to rate the coffee slightly lower compared to the control group. This could suggest



that when the treatment is provided, it negatively impacts the rating of the coffee's taste. Additionally, the views of the participants that are aware of the Chameleon brand has an impact on the Good & Gather scores. If there is a Negative view, the scores decrease. If there is a Neutral view, the scores increase. This indicates a spillover effect, because the views for one brand has some influence on ratings for the other.

When it comes to the Chameleon brand, the treatment group gave slightly higher scores than those in the control group. However, the difference isn't large or statistically significant. Again, the views the participants had for the Chameleon brand did have a statistically significant effect on how participants rated the coffee. If there were Neutral or Positive views, that increased the rates participants gave for Chameleon.

Once we look at the treatment effect with an interaction term for the brand, the effect of the branding did result in statistically significant results. It does point out that there may be some relationship between awareness of one brand and how that can affect the scores of another brand based on the fact that those with a positive awareness of Chameleon scored Good & Gather lower. This could be due to some biases relating to brand comparison rather than actually being based on the taste of the two coffees independently.

Overall, Good & Gather ratings were lower after treatment, which had some influence from the Chameleon brand views. This shows how other brands can influence how someone views the quality of a competing brand. When it comes to the Chameleon ratings, we see a positive increase in rating from its own brand views. This demonstrates how brand views shape a customer's rating on quality.

## Conclusion

This experiment aimed to investigate the impact of branding on taste satisfaction for two different coffee brands, Good & Gather and Chameleon by comparing the satisfaction scores between the treatment group (brand exposure) and the control group (blind taste test). Our findings revealed that branding negatively affected the perceived taste of Good & Gather, with participants in the treatment group rating it lower than those in the control group. Further analysis revealed that participants' views on Chameleon, particularly the negative and neutral perceptions, significantly influenced their ratings for Good & Gather, highlighting a spillover effect from one brand's reputation onto another. Branding had a marginally positive effect on the perceived taste of Chameleon, with treatment group participants rating it slightly higher than the control group. Even though our findings are not statistically significant, there are still some suggestions that branding can both positively and negatively influence taste perceptions, depending on pre-existing brand reputations. The experiment as a whole highlights the pivotal role branding plays in consumer perceptions.

## Limitations and Future Enhancements

Looking back at how our experiment was conducted, one aspect of our experiment that could potentially impact the generalizability of our results is how our sample was conducted. Since we were collecting our data in person, we conducted a convenience sample based on people we interact with in our daily life. Our sample was made up of colleagues, friends, and family members. Because of this, we are likely not

getting a sample that is representative of the entire population. Rather, our family members, friends, and colleagues may have similarities in their coffee drinking tendencies or preferences or demographics which could lead to biased results. For example, colleagues who are used to drinking the coffee machine in the office may develop a preference for it. Another limitation of conducting an in person experiment is sufficient sample size. If we were able to increase our sample size, this could have potentially allowed us to get statistically significant results.

In general, we aimed to keep the conditions in which we conducted the experiment consistent across all our samples. Coffee was always made by mixing in water, subjects were given the same amounts of coffee, and coffee was always served cold. One factor that varied amongst subjects was the time of day in which the samples were conducted; some were conducted in the afternoon at work and some in the evening when our team was with family and friends. If we were to conduct this experiment again, we would prioritize collecting our samples at a similar time of day to limit any variance in scores. For example, coffee may be more desirable in the morning. Another potential aspect of our experiment design that could have led to some bias in our results is the fact that each participant tried both coffees. We did this for the sake of maximizing the data we collected but it may have had some negative impacts. For example, we had our subjects try coffee 1 and then coffee 2, but there is the possibility that coffee 1 left a residual taste impacting how the subjects rated coffee 2. Additionally, based on our analysis, awareness of one brand can impact how someone scores another brand which brings in biases around brand comparison rather than focusing on the coffee taste and branding of the two bottles separately. In the future one way that we could isolate this effect is by having two separate controls and treatments for each coffee brand. This would be ideal if we had access to a much larger sample size.

Lastly, one factor that we noticed when conducting our experiment was that some subjects found the taste of the coffee to be very harsh. Since we opted to use an unsweetened cold brew concentrate, this choice may have been strong for the average coffee consumer. In the future, if we were to run a similar experiment, it may be wise to opt for a cold brew rather than a concentrate so we could still isolate the effects of an unsweetened coffee without having a taste that may be too strong for our average coffee consumer.

# Appendix

## Resources

Petersen, J. M. (1971). Computer decision making for well child care. *JAMA Pediatrics*, 122(5), 375.  
<https://doi.org/10.1001/archpedi.1971.02110160061006>

Lowengart, O. (2012). The effect of branding on consumer choice through blind and non-blind taste tests. *Innovative Marketing*, 8(4).

Shukla, A., Mishra, A.& Dwivedi, Y.K. (2023) *Expectation Confirmation Theory: A review*. In S. Papagiannidis (Ed), TheoryHub Book. Available at <https://open.ncl.ac.uk> / ISBN: 9781739604400

# Control Group Survey

## Coffee Survey Group I

Thank you for participating!

simran.gill@berkeley.edu [Switch account](#)

🔒 Not shared

\* Indicates required question

What is your name? \*

Your answer \_\_\_\_\_

How would you score the taste for Coffee I? (7 being the best, 1 being the worst) \*

1 2 3 4 5 6 7

☐ ☐ ☐ ☐ ☐ ☐ ☐

How would you score the taste for Coffee II? (7 being the best, 1 being the worst) \*

1 2 3 4 5 6 7

☐ ☐ ☐ ☐ ☐ ☐ ☐

What is your age? \*

Your answer \_\_\_\_\_

What is your gender? \*

☐ M

☐ F

☐ Other: \_\_\_\_\_

How often do you drink coffee?

☐ Never

☐ Occasionally (up to 1 time a week)

☐ Sometimes (a few times a week)

☐ Often (almost every day) every day

If you drink coffee, do you usually take it hot or cold? \*

☐ Hot Coffee

☐ Cold Coffee

☐ No preference

If you drink coffee, do you usually prefer it sweet or not sweet?

☐ Sweet

☐ Not Sweet

Do you have any medical conditions that may impact your sense of taste? \*

☐ Yes

☐ No

Are you aware of the brand Good & Gather? If yes, do you view them positively, negatively, or neutral? \*

☐ Yes, Positive

☐ Yes, Neutral

☐ Yes, Negative

☐ No

Are you aware of the brand Chameleon? If yes, do you view them positively, negatively, or neutral?

☐ Yes, Positive

☐ Yes, Neutral

☐ Yes, Negative

☐ No

Submit Clear form

# Treatment Group Survey

## Coffee Survey Group 2

Thank you for participating!

simran.gill@berkeley.edu [Switch account](#)

Not shared

\* Indicates required question

What is your name? \*

Your answer

How would you score the taste for the Good & Gather coffee? (7 being the best, 1 being the worst) \*

1 2 3 4 5 6 7

☐ ☐ ☐ ☐ ☐ ☐ ☐

How would you score the taste for the Chameleon coffee? (7 being the best, 1 being the worst) \*

1 2 3 4 5 6 7

☐ ☐ ☐ ☐ ☐ ☐ ☐

What is your age? \*

Your answer

What is your gender? \*

☐ M

☐ F

☐ Other:

How often do you drink coffee?

☐ Never

☐ Occasionally (up to 1 time a week)

☐ Sometimes (a few times a week)

☐ Often (almost every day)

If you drink coffee, do you usually take it hot or cold? \*

☐ Hot Coffee

☐ Cold Coffee

☐ No preference

If you drink coffee, do you usually prefer it sweet or not sweet?

☐ Sweet

☐ Not Sweet

Do you have any medical conditions that may impact your sense of taste? \*

☐ Yes

☐ No

Are you aware of the brand Good & Gather? If yes, do you view them positively, negatively, or neutral? \*

☐ Yes, Positive

☐ Yes, Neutral

☐ Yes, Negative

☐ No

Are you aware of the brand Chameleon? If yes, do you view them positively, negatively, or neutral?

☐ Yes, Positive

☐ Yes, Neutral

☐ Yes, Negative

☐ No

Submit

Clear form