Getting to the Ground(s) Truth: An Analysis of Coffee Shop Tipping Behavior

W241 - Final Project

4/20/2023

Authors: Marguerite Morgan, Zach Gutierrez, and Mike Varner

https://github.com/mike-varner/W241_Final_Project

Abstract

How does changing default gratuity settings impact pre-service tipping behavior at coffee shops? In the United States, tipping is a critical component of many workers' income in the food and beverage industry and default gratuity settings on point-of-sale (POS) systems have become increasingly widespread. Prior research has shown that increasing the range of default tipping recommendations leads to fewer tips but higher overall revenue from those that do tip. We investigated whether increasing some but not all default tip options leads to higher average tip percentages. We conducted a between-subjects, two group, post-test randomized experiment at two coffee shops: Mercury Cafe and Ryan Bros (located in San Francisco and San Diego repsectively). All patrons who visited one of these coffee shops, while the experiment was being conducted and who paid with a Debit or Credit card only, had their transaction randomized to the control group (N = 270) or treatment group (N = 279). We found that on average, individuals in the treatment group tipped 1.2% higher than individuals in the control group. However, this finding was not significant at a 5% level. Our results are encouraging, however, as relatively modest increases in default settings (3-5%) can lead to higher tip revenue. Moving forward, we recommend increasing the sample size and adjusting the treatment tip percentages to be even higher in order to achieve sufficient power and further investigate this relationship.

Introduction

Tipping has become an extremely important part of the American food and beverage industry. As of 2011, it was estimated that annual tips represent roughly $47B in the food industry alone in the United States. In many states, tipped employees have a lower minimum wage, making tipping a crucial piece of their income. Among those that work at counter service restaurants in the US, it's estimated that as high as 60% of their income may come from tips (Azar, 2020). Some historians believe that the practice of tipping originated in England in the 16th century where local pubs and coffee houses left out urns labeled "To Insure Promptitude." Patrons would add money to the urn ahead of ordering to ensure they received good service. Today, the practice of tipping in bars, restaurants and coffee shops has become even more prevalent with the introduction of tipping guidelines on POS systems. Given the importance of tipping to so many American workers and the widespread prevalence of tipping guidelines at many food and beverage establishments, our research focuses on the following:

How does changing default gratuity settings impact pre-service tipping behavior at coffee shops?

Existing research has demonstrated that both providing tipping guidelines as well as showing a higher range within those guidelines increases overall revenue driven from tips. A 2011 study found that tipping guidelines, in the form of calculation assistance, led to a 15% increase in tips (Seiter et al., 2011). Another study conducted in 2021, found that larger default tip options led to a lower percentage of individuals tipping but larger overall tipping revenues. Simultaneously, a survey conducted across individuals in the US and Israel in 2010 found that the most common reason for tipping was a feeling of wanting to abide by a social norm and avoid the guilt or embarrassment associated with not tipping (Azar, 2020). These three studies suggest that individuals are more likely to tip when guidelines are provided, more likely to tip larger amounts when the default tipping settings are higher, and that a customer's desire to tip is typically driven by wanting to follow social norms. Based on this foundation of research, we hypothesized that increasing the tipping guidelines for some but not all tipping options would lead to a similar number of people tipping and would also lead many to tip higher thereby increasing overall average tip percentages for workers.

Experimental Details

Experimental Design

The experiment involved adjusting the default gratuity settings on the POS system at two different coffee shops: Mercury Cafe in San Francisco and Ryan Bros in San Diego. The experimental design included a between-subjects, two-group, post-test only, randomized experiment as summarized in the ROXO diagram below:

Figure 1

R	X	O
R		O

Coffee shop patrons' transactions were randomized either into the control group where they experienced the coffee shop's default tip settings or the treatment group where they experienced a slight increase in some (but not all) of the coffee shop's default tip settings. The potential outcomes being observed were the percent tip per transaction. We compared the average percent tip within the treatment group to the average percent tip within the control group in order to learn about a potential causal effect.

Randomization

Coffee shop patrons were randomized based on when they visited each coffee shop. At Ryan Bros, randomization happened on a day level. There were two weekdays and two weekend days during which the control settings were used and there were two weekdays and two weekend days during which the treatment settings were used. Deploying the treatment and control during the week and on the weekends ensured that there were no inherent differences in weekend vs.

weekday customers that might impact the tipping behavior observed. At Mercury, the experiment was conducted over the course of one day and individuals were randomized every hour from 9am to 3pm. These randomization schemes were based on what each coffee shop felt was feasible for them.

<u>Treatment</u>

The treatment included increasing the lowest default tip option by 5% and the second lowest default tip option by 3%. The other default tip options remained the same.

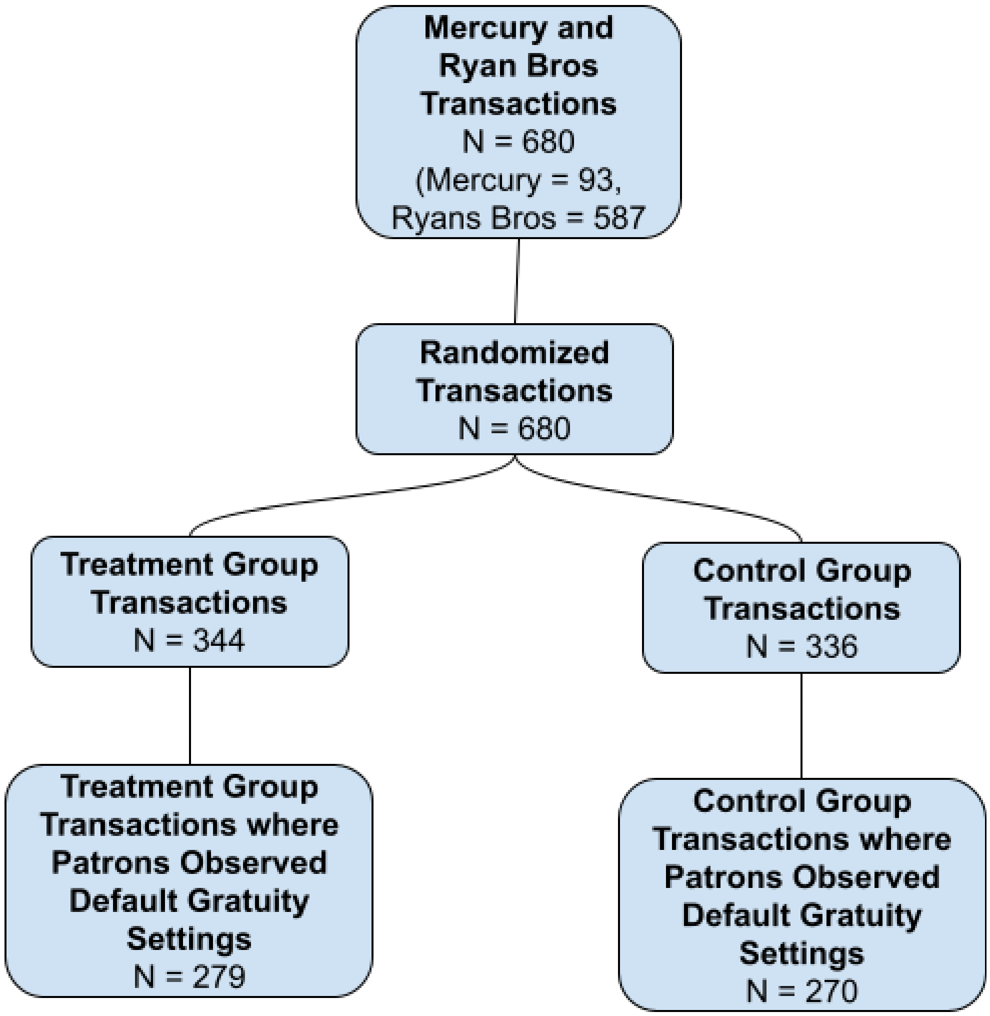
Figure 2 shows that Mercury Cafe and Ryan Bros had different baseline default tip options but the treatment applied to each was consistent. This treatment was selected in order to test the hypothesis that increasing some but not all of the default tip options leads to a similar percent of people tipping but more people tipping higher and thus more overall tip revenue. The specific percentage increase used was identified in partnership with the coffee shops based on what they were comfortable with.

Figure 2

Coffee Shop	**Control [Current default settings]**	**Treatment**
Ryan Bro’s	10%	15%
	15%	18%
	20%	20%
	30%	30%
	Custom Tip	Custom Tip
	No Tip	No Tip
Mercury Cafe	15%	20%
	20%	23%
	25%	25%
	Custom Tip	Custom Tip
	No Tip	No Tip

All Mercury and Ryan Bross patrons’ transactions that occurred on the dates of the experiment were considered for involvement in the experiment and randomized. There was some noncompliance with the treatment regimen as shown in Figure 3. Individuals who did not pay solely with a credit or debit card (i.e. with cash, a gift card, or a combination of payment methods) likely did not view the default gratuity settings for their full sale amount or potentially at all. These individuals' transactions were removed from both treatment and control groups given we were not confident that they fully received the treatment or control gratuity settings.

Figure 3



<u>Initial Power Analysis</u>

Prior to conducting the experiment, we expected to see an effect size of ~2.55% meaning that those in the treatment group would tip on average ~2.55% higher than those in the control group. We assumed the average tip in the control group would be 17.45% with a standard deviation of 8.26% and the average tip in the treatment group would be 20% with a standard deviation of 9%. These assumptions were based off of another study conducted on taxi ride tips in New York City between 2010 and 2018 (Donkor, 2021). This meant that a sample size of ~350 transactions would give our experiment roughly 80% power.

Analysis

<u>Exploratory Data Analysis</u>

Given each coffee shop had a different POS system, the types of covariates each had were different. We were curious to get as many covariates as we could in hopes to improve the precision of treatment effect and to help sanity check our results. For example, tender type (Figure 4), showed a relatively even split between treatment and control with very similar means, aside from debit card in treatment being slightly higher. For order type (Figure 5), the results looked mostly consistent for dine in, but in-store pick up was a bit lopsided, both in terms of mean tip percent and count. Lastly, we wondered how the particular cashier working during the transaction might affect the outcome (Figure 6). The only issue here was that one cashier did not have a shift on a treatment day. Aside from that, the splits were fairly even.

Figure 4

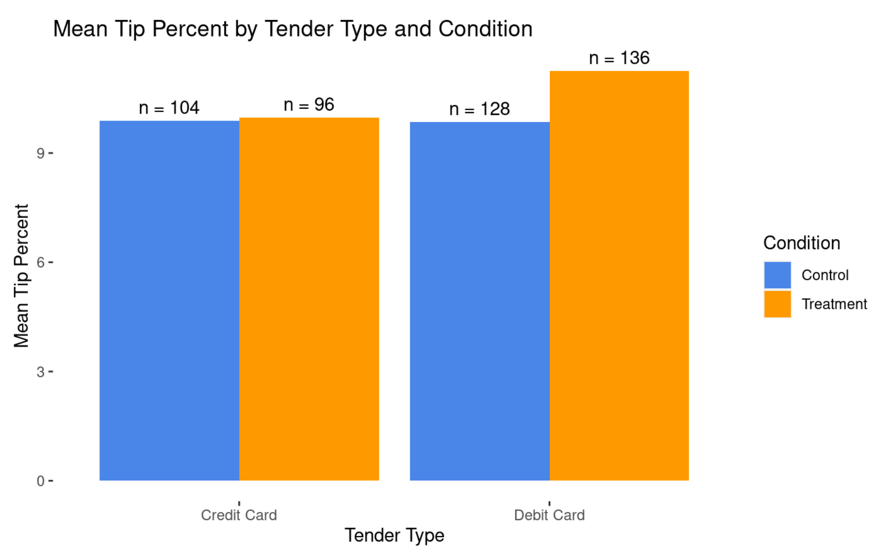


Figure 5

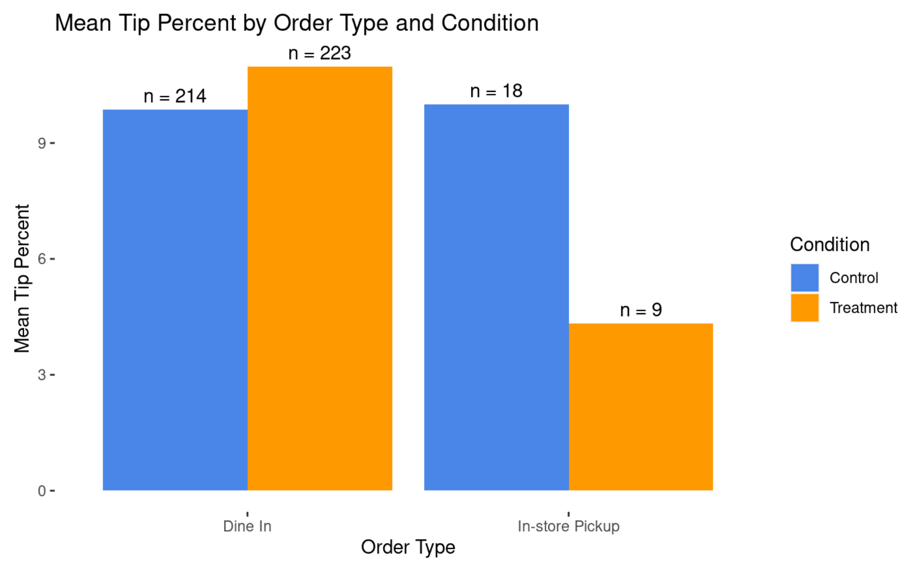
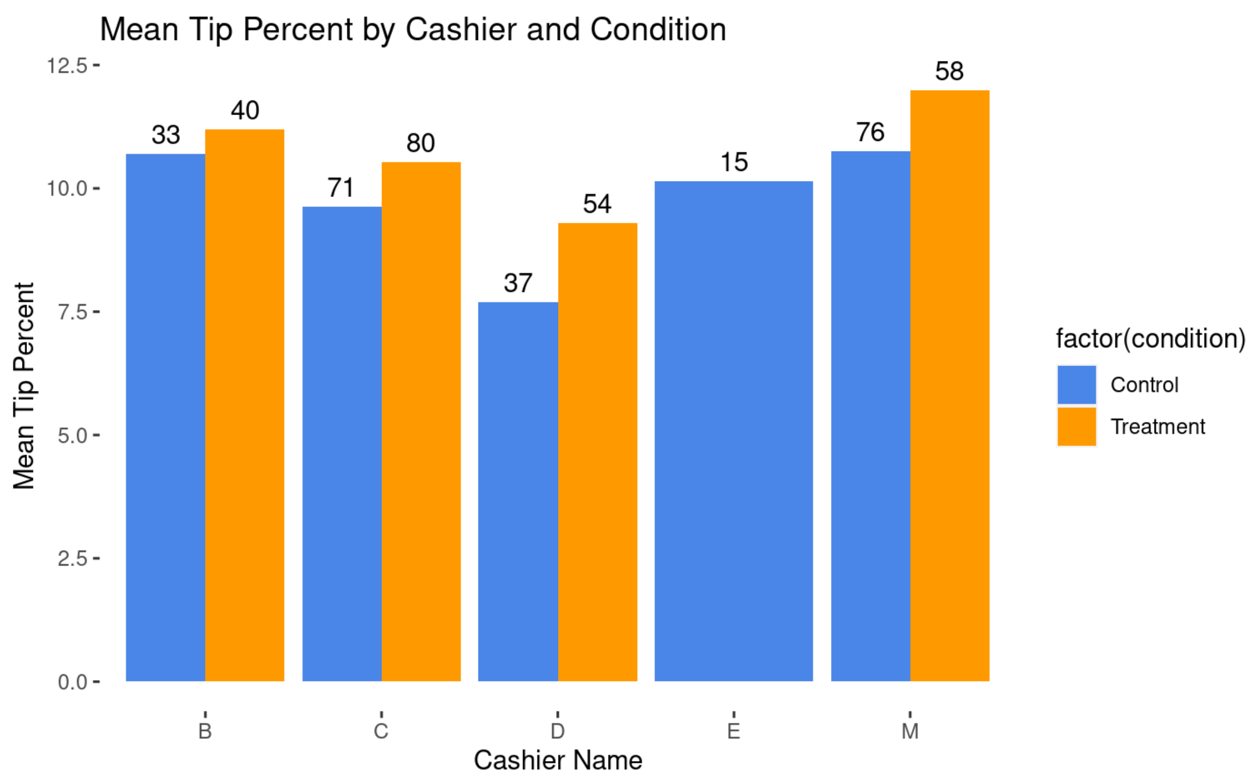


Figure 6



<u>Regression Results</u>

We first looked at the raw effect of our treatment on our outcome of interest, tip percent, and found that the treatment had an effect of ~1.4% (see base model results below in Figure 7). While this effect is only significant at the 10% level, we do think that this result is encouraging nonetheless. This 1.4% treatment effect is large in the context of our treatment which only moved some default options by 3-5%. Therefore, if in the future we increased the default options higher than this we may see a larger and potentially more significant effect.

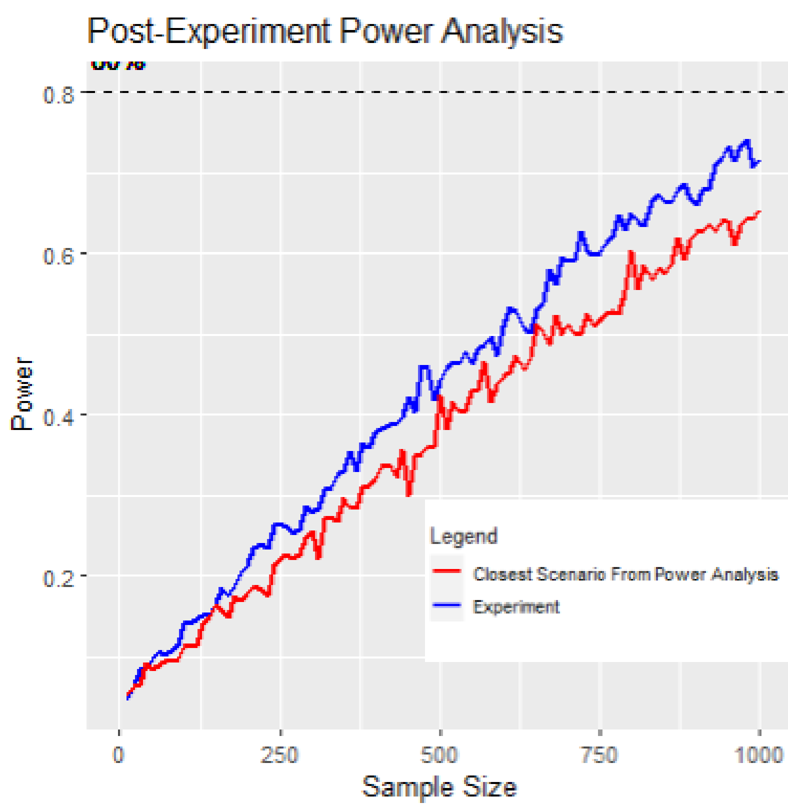
From the base model, we wanted to understand if location impacted our results given we ran this experiment at two different locations. The interaction model below did not show a significant location difference when interacted with treatment or when by itself. Based on this, we proceed to incorporate covariates and pool over location. However, given Mercury and Ryan Bros use different POS systems, they have completely different sets of covariates. To appropriately incorporate these we took a two-stage approach. First, we regressed tip percent for each store on its respective set of covariates (models Ryan Bros and Mercury in Figure 7). Secondly, we computed the residuals from these regressions and regressed these on our treatment and location indicator (final model in Figure 7). We found that the inclusion of covariates reduced our treatment effect slightly to 1.2% and stayed significant at the 10% level while also reducing the standard error of our treatment effect. We did see that the inclusion of covariates slightly improved our precision, although not as much as we would have hoped.

Figure 7

	Dependent variable:				
	tip_percent				residuals
	Base	Interaction	Ryan Bros	Mercury	Final
	(1)	(2)	(3)	(4)	(5)
treat	1.353*	3.829**			1.206*
	(0.713)	(1.665)			(0.708)
is_ryan		0.165			0.072
		(1.349)			(0.921)
treat:is_ryan		−2.972			
		(1.841)			
order_typeIn-store Pickup			−2.741		
			(1.843)		
order_typeTake Out			4.756***		
			(1.465)		
cashier_name_clean_2			−1.493		
			(1.339)		
cashier_name_clean_3			−0.629		
			(0.986)		
cashier_name_clean_4			−1.459		
			(1.490)		
tenderDebit Card			0.548		
			(0.806)		
is_reward			0.167		
			(1.260)		
day_of_weekSaturday			0.157		
			(1.236)		
day_of_weekSunday			−1.228		
			(1.362)		
day_of_weekThursday			−0.596		
			(1.235)		
card_entry_methodsSwiped				1.125	
				(1.926)	
card_entry_methodsTapped				−0.787	
				(2.087)	
card_brandDiscover				−0.521	
				(4.237)	
card_brandMasterCard				−2.935	
				(3.255)	
card_brandVisa				−3.033	
				(3.106)	
Constant	9.845***	9.703***	11.153***	14.908***	−0.675
	(0.509)	(1.229)	(1.198)	(3.399)	(0.925)
Covariates	No	No	Yes	Yes	Yes, but in first stage only
Observations	548	548	464	84	548
R^2	0.007	0.015	0.015	0.019	0.005
Adjusted R^2	0.005	0.009	−0.006	−0.044	0.002
Note:					*p<0.1; **p<0.05; ***p<0.01
					All coefficients are reported using robust standard errors.

After not finding a statistically significant result at the 5% level, we wanted to understand to what extent our study was underpowered. To do this, we reran our power analysis using our experiment's results and found that we had ~50% power for the number of observations we collected (Figure 8). While we did find that our experiment was more powered than the least powered scenario, from our original power analysis, we still weren't highly powered. Assuming there is a statistically significant difference in tip percentage between the treatment and control groups, there is only a 50% chance we would have been able to observe it given our experimental design. From our initial power analysis we found that power was highly sensitive to the ATE; our two other plausible power scenarios achieved ~80% power with our current sample size and only a modest increase in the assumed ATE. In hindsight our study was not sufficiently powered for the given ATE we observed.

Figure 8



Conclusions & Next Steps

We did not see a significant difference in average tip percentage at the 5% level when increasing some of the default tip settings on coffee shop POS systems. Our treatment did not lead to a statistically significant increase (via a t-test) in the percentage of non-tippers and we did see an increase in average tip percentage from control to treatment groups as hypothesized. However, the increase in average tip percentage among tippers was not large enough to be statistically meaningful at the 5% level. That said, given that the control tipping options were increased by 3-5%, the fact that those in the treatment group tipped on average ~1.2% higher is worth noting. For future research, we would recommend increasing the sample size in order to have a better powered experiment. In addition, we would recommend increasing the default tip recommendations in the treatment group to be slightly higher. For example, increasing the second lowest default option by 5% as opposed to 3% may still maintain the same number of non-tippers but increase the average tip percentage among tippers to reach significance at the 5% level. In addition, it would be helpful to identify coffee shops that have historical customer level tip percentage data. Many of our existing covariates were not extremely useful in increasing the precision of our estimated treatment effect. However, this additional covariate would likely help reduce the standard error of the treatment effect estimate. Lastly, it may be worthwhile to explore other treatments that could increase tip percentage such as displaying a sign saying "Tips are appreciated!" or adjusting barista behavior while leaving the tip options the same.

References:

Azar, Ofer H. 2020. "The Economics of Tipping." Journal of Economic Perspectives, 34 (2): 215-36. DOI: 10.1257/jep.34.2.215 https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.34.2.215

Alexander, Damon, Boone, Christopher, and Lynn, Michael (2021) The Effects of Tip Recommendations on Customer Tipping, Satisfaction, Repatronage, and Spending. Management Science 67(1):146-165. https://doi.org/10.1287/mnsc.2019.3541

Donkor, Kwabena Baah. "The Economic Value of Norm Conformity and Menu-Opt-out Costs." Stanford Graduate School of Business, 1 Dec. 2021, https://www.gsb.stanford.edu/faculty-research/working-papers/economic-value-norm-conformity-menu-opt-out-costs.

Seiter, J.S., Brownlee, G.M. and Sanders, M. (2011), Persuasion by Way of Example: Does Including Gratuity Guidelines on Customers' Checks Affect Restaurant Tipping Behavior?. Journal of Applied Social Psychology, 41: 150-159. https://doi-org.libproxy.berkeley.edu/10.1111/j.1559-1816.2010.00706.x

Square "Accept Tips with the Square App." Accept Tips with the Square App | Square Support Center - US, https://squareup.com/help/us/en/article/5069-accept-tips-with-the-square-app.f