Does Spending on Others Make you Happier than Spending on Yourself?

Julia Buffinton, Morris Burkhardt, I-Wae Niu, Jayashree Raman

Introduction

As global incomes continue to rise and millions transition into the middle class with disposable incomes, companies are spending more than ever on advertising and marketing new consumer and household products to this new generation of consumers. With advertising slogans such as “Because you’re worth it!” (L’Oreal), “Don’t dream it. Drive it!!” (Jaguar), “The sweet smell of success” (Macallan), consumers are encouraged to spend more on themselves if they want to be happier or more successful in life.

So can money buy happiness? As early as 1972, the Himalayan Kingdom of Bhutan coined the term “Gross National Happiness” (GNH) declaring it more important than “Gross Domestic Product” (GDP). In 2012, the UN issued the first World Happiness Report and encouraged the spread of Bhutan’s GNH philosophy. Despite having the strongest economy in the world and rising incomes, the US has dropped four spots in the latest UN World Happiness Report ranking to #18. There seems to be a disconnect between wealth and happiness.

Our study aims to answer the question “Does spending on others make you happier than spending on yourself?” We believe that a better understanding of the science behind spending and happiness will ultimately help people make better decisions to live happier lives, and even potentially help countries translate greater wealth (GDP) into greater happiness (GNH) by encouraging more prosocial giving among their citizens.

Background

There is much research on the effect of income on happiness that concludes that money can indeed buy happiness but only to an extent, and less than one might think. Of particular interest to us is an older paper by Dunn, Aknin & Norton (2008) which concluded that spending one’s income on others was a greater predictor of happiness than spending on themselves. The researchers reached their conclusions through two observational studies testing initial correlation and one field experiment to demonstrate causal impact of prosocial spending.

The first observational study was conducted among a nationally representative sample of 632 Americans, asking them to rate their happiness levels, report their annual income, and to estimate how much they spend on various categories which were combined to provide an index

of personal vs. prosocial spending. A regression predicting happiness levels indicated that personal spending had little impact on happiness (B=-0.02, NS) while prosocial spending was associated with greater happiness levels (B=0.11, p <0.01). The second observational study examined 16 employees' happiness levels before and after they received profit-sharing bonuses, recorded their income and categories they spent their bonuses on which were combined to provide an index of personal vs. prosocial spending. Once again, a regression predicting Time 2 happiness levels revealed that prosocial spending was the only significant predictor of happiness (B=0.81, p<0.02) even when controlled for income (B=0.96, p<0.02).

The researchers sought to demonstrate the causal impact of prosocial spending by conducting a field experiment (n=46) at one university campus by randomly assigning participants to receive either $5 or $20 in the morning with instructions to either spend it on themselves or on someone else by 5pm that same day (2x2 design). Participants were asked to rate their happiness levels before receiving the "treatment" and afterwards through a phone call survey in the evening. The analysis revealed prosocial spenders reporting greater post happiness levels compared to personal spenders, irrespective of windfall size.

It is this causal impact of spending on others vs. spending on oneself on happiness that we would like to measure through another field experiment. We hope to strengthen and add to research in this area of study by attempting to replicate some aspects of the Dunn, Aknin & Norton (2008) field experiment while improving on some methodological aspects i.e. using a more diverse and larger sample and utilizing real-time surveys for responses (vs. waiting 6-8 hours before calling participants back for a response).

Methods/Experimental Design

To answer the question of "Does spending on others make you happier than spending on yourself?" we conducted a randomized control experiment over the course of 5 weekends in July and August. The subjects were drawn by a random selection of Starbucks customers who visited the store on weekend mornings and Friday late afternoon. The two treatment groups received $5 gift cards, the first of which was instructed to spend on themselves, and the second group was told to buy a beverage for someone else. Our experiment measured the happiness levels pre- and post-treatment for all the subjects in the treatment and control groups. The primary outcome we measured was the change in happiness ratings for each participant. We hypothesized to see positive changes in happiness ratings for both treatment groups after the spending activity but more importantly, we anticipate this increase in happiness to be greater in "treat someone else" (treatment 2) vs. "treat yourself" (treatment 1) groups.

Research Design

Our research design can be summarized using the "ROXO grammar" seen in Figure 1:

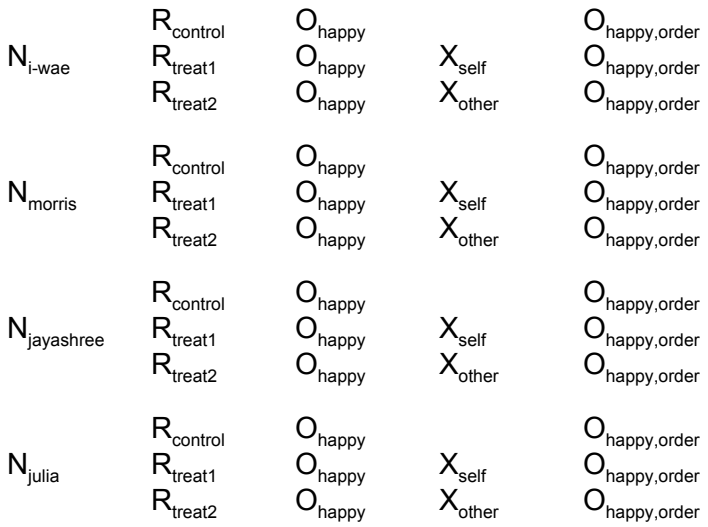

Figure 1: Experimental design using ROXO grammar, as described by Trochim and Donnelly (2016).

We collected pre- and post-treatment happiness ratings in order to calculate a “happiness difference” as a result of treatment. The baseline measurement was key for minimizing variation in the happiness measurement and allowing us to better identify an effect. Due to our pretest-posttest design, we did not to rely much on covariates for power. The only covariate we observed is gender, for potential heterogeneous treatment effects.

Each researcher recruited participants at a local Starbucks store. Participants were assigned to treatment groups by alternating assignment (Control, Treatment 1, Treatment 2, Control), to ensure that approximately the same number of participants in each group were collected each day, to remove any confound of the collection date with treatment. Participants for each treatment group were split evenly across researchers, thus, randomization was blocked by researcher/city.

Researchers approached all Starbucks patrons and asked them if they were willing to participate in a brief survey for a school project. If they agreed, they were told that this was a three-question survey; we would ask them one question before they entered the store and two questions after they returned, including what they ordered. The question about their order was included to mask the actual objectives of the survey. At this point, we did not alert participants to which treatment they would receive.

Each participant’s happiness level was measured before entering the store, prior to treatment instructions (Figure 1). After recording their baseline happiness, if the participant was in one of the treatments groups, they were instructed that there was one more part to the survey that had not yet been mentioned. At this point, they received a $5 Starbucks gift card with instructions to either spend it on themself during this visit (Treatment 1) or give it to another patron from them

to use (Treatment 2). Control participants did not receive a gift card. Researchers attempted to minimize interference of hearing/seeing participants in other treatment groups by waiting several minutes after a participant entered the store before recruiting the next, so as few participants as possible were inside the store at any point. For this experiment, it did not matter whether they knew the person who received the gift card.

After returning from Starbucks, participants recorded what they had ordered and their happiness level. To monitor non-compliance without the researcher entering Starbucks to directly observe it, participants were asked whether they were able to complete the task. Some participants in Treatment 2 shared anecdotes about approaching strangers with the gift card. For the purposes of this experiment, treatment began when the participant opened the envelope to read the instructions for what to do inside Starbucks. The inclusion of a pure control group is key for measuring non-compliance under the intent to treat effect (Control vs. Treatment 1, and Control vs. Treatment 2). Researchers also tracked and recorded attrition rates, although attrition was expected to be low as the researchers were directly outside Starbucks and could make sure they were able follow up with participants.

Pilot Study

To test our research design, procedures and materials, a pilot study was conducted with 20 participants. Initial power analysis calculations indicated our planned sample sizes (Control n=104; Treatment 1 n=52; Treatment 2 n=52) were sufficiently powered to detect an intent to treat effect (ITT) between Treatment 1 vs. Control and Treatment 2 vs. Control, however, was underpowered to detect an ITT effect between Treatment 1 vs. Treatment 2 (see Figure 2). We would have required almost four times as many observations (380 vs. 104) to achieve 80% power. Due to time and budget constraints, our team decided to implement our experiment with the original planned sample sizes, despite its lack of power.

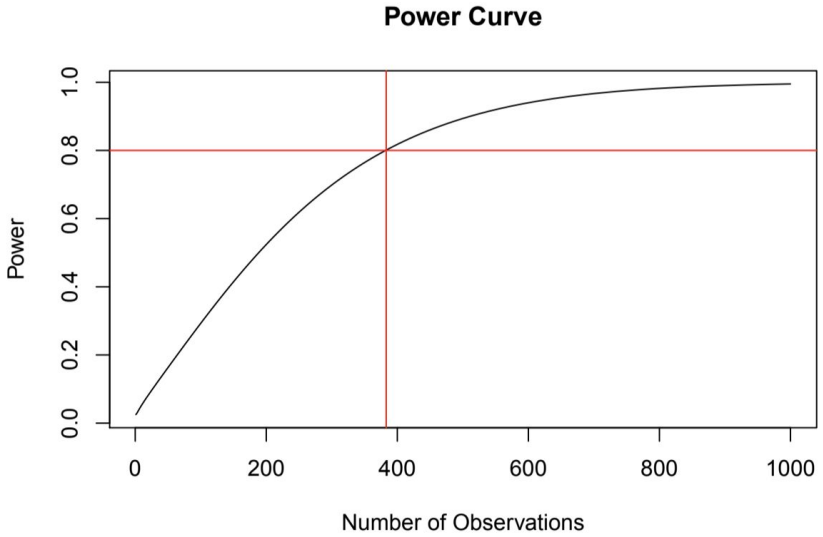

Figure 2: Power curve for Treatment 1 vs. Treatment 2

Participants

A total of 208 participants were recruited by all four researchers outside of Starbucks stores in four cities in the United States and Canada (Vancouver, BC; Culver City, CA; Sunnyvale, CA; Arlington, VA). Each researcher recruited participants from one or two Starbucks locations in their respective city on Saturday and Sunday mornings between 9:00 a.m. and 12:00 p.m and on Friday late afternoon between 3:00 and 5:00 p.m. These participants opted in to the study; researchers asked Starbucks patrons as they approached the store if they'd be willing to complete a short survey for a school project. Approximately half (n=102) of the participants were in the control group; and the remaining participants were divided evenly between treatment group 1 (n=53) and treatment group 2 (n=53). See Figure 3 for a flow diagram of experiment observations. The distribution of participants in control vs. treatment is not exactly even due to experimenter error in allocating treatment. A discussion of attrition and non-compliance rates is in our results section below.

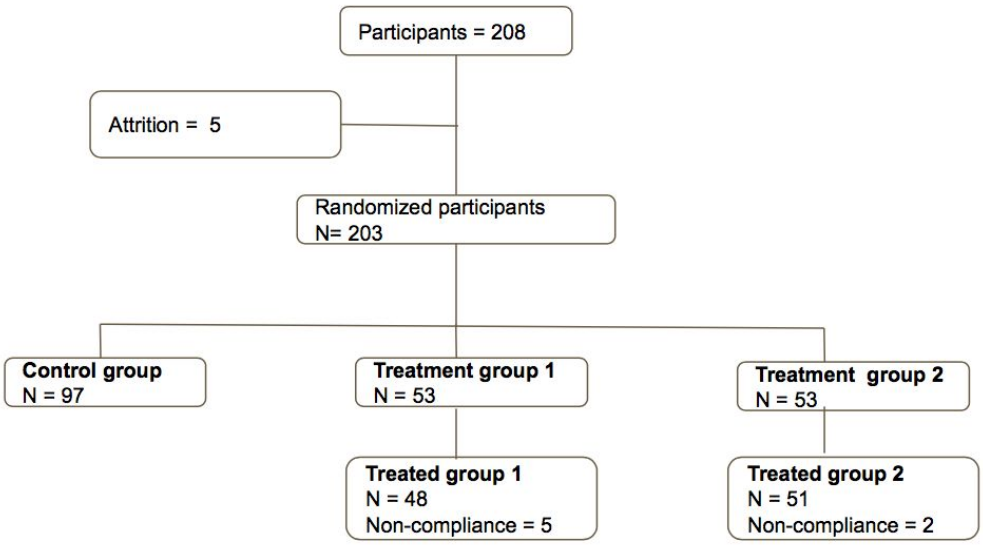

Figure 3: Flow diagram of experiment observations.

Materials & Measurement

We measured happiness levels pre- and post-treatment using a 10-point rating scale (Figure 4).

Figure 4: Happiness survey question

Participants were instructed to rate their current happiness from unhappiest they've ever been (1) to the happiest they've ever been (10). The anchors for the rating scale were selected to minimize the potential of top- or bottom-coding, which would maximize the chance that we are able to observe an effect (if there is one). A scale from 1 to 10 was used because it is the most natural rating scale for most individuals. Additionally, after treatment, all participants were asked to record what they ordered at Starbucks. This was included as a mask for the true objective of the experiment and to increase plausibility of the excludability assumption by ensuring uniform handling of treatment and control groups (Figure 5).

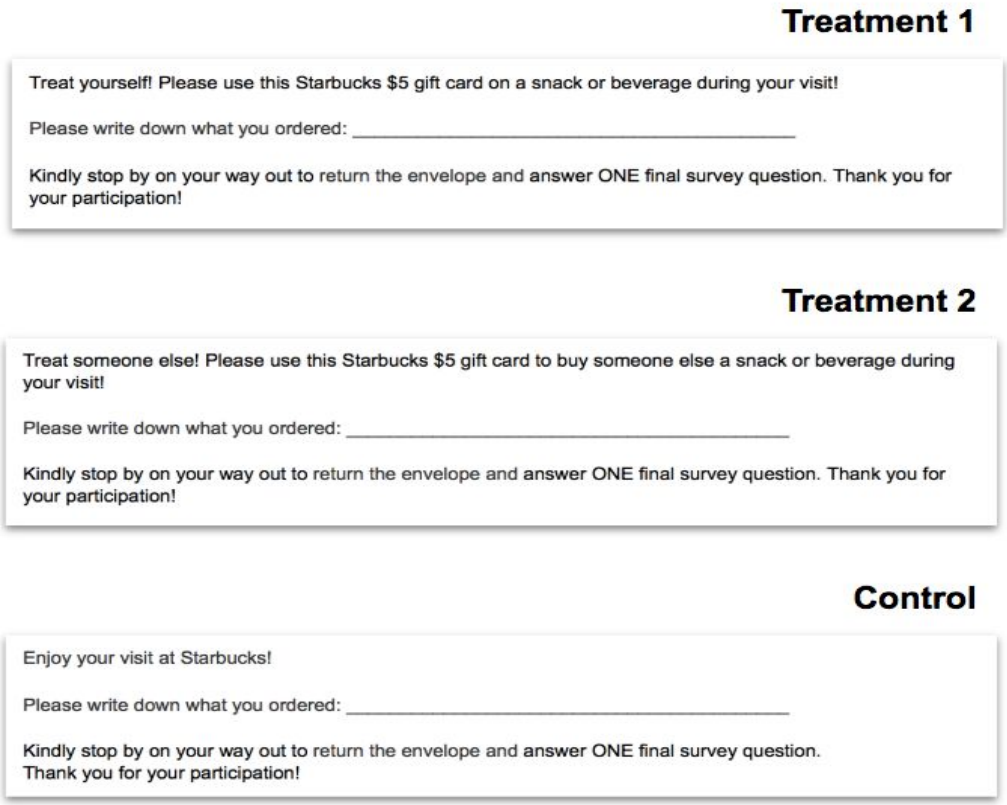

Figure 5: Instruction cards

Procedures

We collected pre- and post-treatment happiness ratings in order to calculate a “happiness difference” as a result of treatment. The baseline measurement was key for minimizing variation in the happiness measurement and allowing us to better identify an effect. Due to our pretest-posttest design, we did not to rely much on covariates for power. The only covariate we observed is gender, for potential heterogeneous treatment effects.

After individuals agreed to participate and before they entered the store, they answered the happiness question (see Figure 1). Upon completion, they were instructed to continue into the store for their purchase and record what they ordered. Additionally, depending on treatment, participants received further instructions for their order. Control participants (n=102) received no further instructions beyond reporting their order. Participants in treatment group 1 (n=53) received instructions to “Treat yourself! Please use this Starbucks $5 gift card on a snack or beverage during your visit.” Participants in treatment group 2 (n=53) received instructions to spend their gift card on another patron in the store: “Treat someone else! Please use this Starbucks $5 gift card to buy someone else a snack or beverage during your visit.” For this experiment, it did not matter whether they knew the person who received the gift card. Lastly, all participants were asked to answer the happiness question after returning from the store.

Analytic Plan

Once we completed the data collection, we pursued the following analytic plan:

1. **Descriptive Statistics**
    a. Researcher, Gender and Treatment Groups
    b. Attrition and Non-compliance
    c. Happiness rating (pre-, post-, difference)

2. **Attrition & Non-Compliance**
    A pilot study of the first 20 participants suggested that the attrition and non-compliance rate should be well below 5% and independent of both treatment group assignment and potential outcomes. Should this hold for the complete experiment, all data points for participants who show attrition or non-compliance will be ignored. If this however does not hold, we will place bounds on the intent to treat (ITT) effect. We will do so by filling in the minimum or maximum observed outcome for each of our data points that exhibits attrition or non-compliance. This will enable us to calculate the minimum and maximum intent to treat effect.

3. **Visualizations**
    To conduct exploratory data analysis, we will examine the distribution of participants by gender, researcher/location, baseline happiness, and post-treatment happiness using

boxplots and histograms. We will do this for the complete dataset and by each treatment group and other covariates.

We will also display the distribution (again using boxplots and histograms) of the 'difference in happiness' variable (post-test happiness rating minus pre-test happiness rating) overall and by each covariate. We will also binarize the difference in happiness variable, assigning 0 for a decrease or no change in happiness and 1 for a positive change in happiness.

4. **Regression Models for the Intent To Treat Effect**
   We propose the following regression models to examine the intent to treat effect:

Dependent variable	***Treatment Group***	***Gender (covariate)***	***Researcher (covariate)***	***Happiness Baseline (covariate)***
Difference in Happiness	X			
	X	X, I		
	X		X, I	
	X			X
	X	X	X	X
Binarized Difference in Happiness	X	X	X	X

Table 1: Regression models

X: variable is included as independent variable in model
I: interaction term of variable with 'treatment group' variable is included in model

Whenever 'Difference in Happiness' is specified as the dependent variable, we will be fitting a linear regression model and when 'Binarized Difference in Happiness' is specified as dependent variable, we will be fitting a logistic regression model. We decided to include the logistic regression model since we suspect that people are not very good at quantifying differences in their happiness level. A simple 0/1 indicator to differentiate between 'no increase in happiness' and 'increase in happiness' might be able to filter out some of the noise.

We will be reporting heteroskedastic robust standard errors for all of our linear regression models. Aside from the ‘treatment group’ variable, we are only including covariates (gender, researcher, and baseline happiness) in our regression models from which we are not intending to draw inference. Hence, we will not be running into the multiple comparison problem and do not need to adjust the significance level. We will therefore be working with a 5% significance level for every regression model.

To check for data points that have high influence on our regression, we will create a plot of Cook’s distance for our baseline linear model (without covariates).

Descriptive Statistics (EDA)

Each researcher collected 52 participants. Our plan was to collect double the number of control participants (26) as participants in each treatment group (13). All researchers were able to do so, except for one, who collected data from 24 control participants and 28 treatment participants (14 in each treatment group), due to mis-counting in allocating treatment.

Gender

We are interested in gender as a potential source of heterogeneous treatment effect. However, we did not actively attempt to control the number of participants from each gender in each treatment group. This is only an issue if there is a drastic difference between the number of participants in each gender. We only had slightly more female participants than male participants (116 vs. 92). This could have been from variation due to the small sample size or simply the distribution of gender in Starbucks patrons at those locations. This is not necessarily problematic, but it does become more interesting as it breaks down by treatment group.

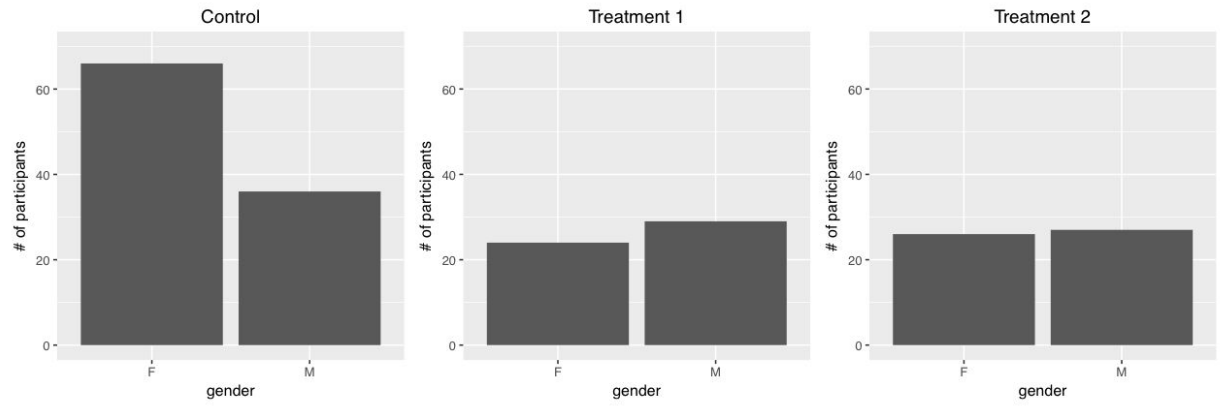

Figure 6: Distribution of gender by treatment group.

Gender was approximately balanced across both treatment groups, with a few more men in each treatment group. Because there were more female participants overall, this meant that there were many more female participants (66) than male participants (36) in the control group.

Attrition and Non-compliance

Both non-compliance (3.4%) and attrition rates (2.4%) were low, relative to typical survey experiments (see table 2). This is because participants completed pre- and post-treatment surveys immediately before and after the survey, in person, in the same location. Given the design, attrition should be zero or very close to it, but in some cases, the participant remained in the store longer than the researcher was present, or they somehow left the store without the researcher noticing.

	Non-compliance	% Non-compliance	Attrition	% Attrition
No gift card	0	-	3	2.9%
Treat yourself (1)	5	9.4%	1	1.9%
Treat someone else (2)	2	3.8%	1	1.9%
Total	7	3.4%	5	2.4%

Table 2: Attrition and non-compliance rates by treatment group, and in total.

Approximately 2.4% of participants attrited. All of them were female and they were all recruited by Julia and Morris. They had a range of pre-treatment happiness ratings. Attrition rates were pretty consistent across treatment groups (2.9% vs. 1.9%).

The non-compliance rate of 3.4% is higher and different across treatment groups. Of the seven participants that did not comply, six of them (86%) were men, and most were recruited by I-Wae and Julia. Two of these participants reported a baseline happiness rating of 1. Typically, non-compliance was due to participants' actions regarding the gift card, so we did not observe non-compliance in the control group. Several participants did not want to spend the gift card on themselves and returned it after treatment. One person could not find a stranger who was willing to accept the gift card, and another elected to give their gift card to someone more needy than a Starbucks patron. While admirable, this non-compliance was much higher in treatment group 1 (9.4%) than treatment group 2 (3.8%), so in order to examine the effect of treatment 1 vs. treatment 2, we placed bounds on the ITT.

Baseline Happiness & Difference in Happiness

Just over half (108/208) of the participants reported a baseline happiness of 7 or 8 (see Figure 7). About three quarters (154/208) of the participants reported baseline happiness 7 or above; however, a few people reported happiness levels below 3.

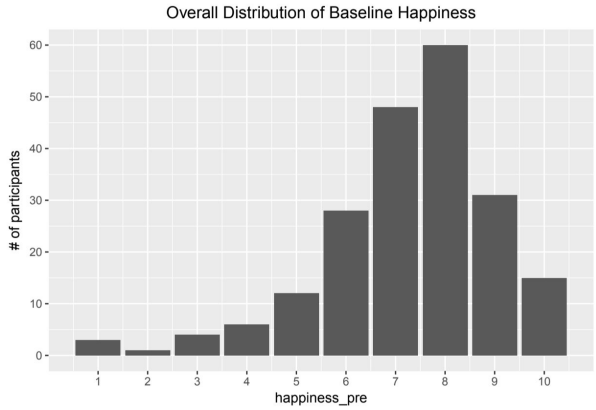

Figure 7: Distribution of baseline happiness.

It is noteworthy that a person's potential change in happiness is dependent on their baseline (pre-test) happiness level. Figure 8 displays boxplots of the difference in happiness dependent on the baseline happiness. In general, people who started happier, saw smaller differences in happiness than those who startet less happy. Fifteen individuals reported "10" for their baseline happiness. This meant that even if there was an increase in their happiness levels, we would not be able to measure it. Luckily, those individuals made up roughly the same proportion of each treatment group, and they all reported a 10 post-treatment, which means that at worst we will be under-estimating the effect.

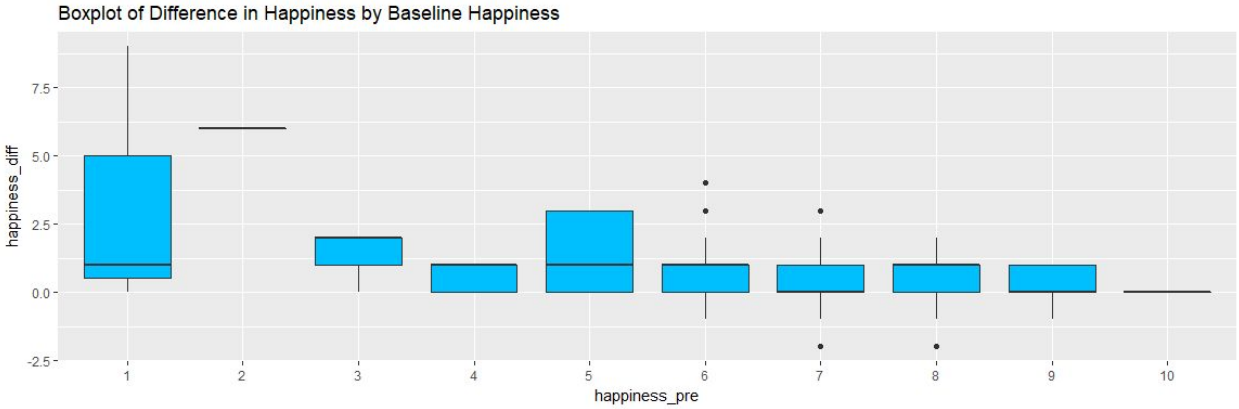

Figure 8: Difference in Happiness by Baseline Happiness

Over 75% of the subjects indicated a difference in happiness of either 0 or 1, while only nine participants (~4%) reported a decrease in happiness (Figure 9). All of these participants were in the control group. When asked, almost all of them cited long lines as the reason they were unhappy. Others were dissatisfied with the music inside the store or the accuracy of their order. On the other hand, a few people saw great change in their happiness level, with changes of 6 and 9 between pre- and post-treatment surveys. The participant who reported a change of 9, although a bit of an outlier compared to the rest of the dataset, is not a concern for our analysis because the Cook's distance (in our baseline regression model) for this value is only .35, so its influence is not too large (Figure 10).

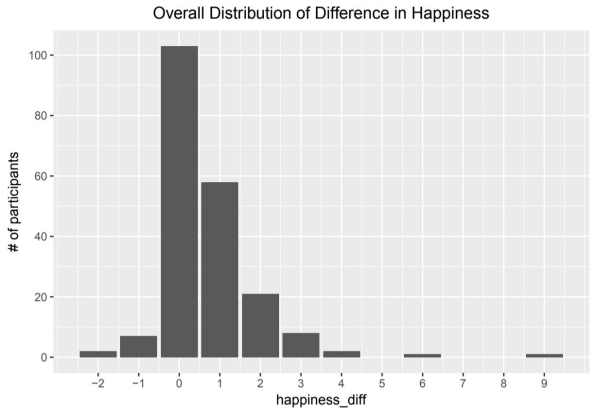

Figure 9: Distribution of difference in happiness between pre- and post-treatment surveys.

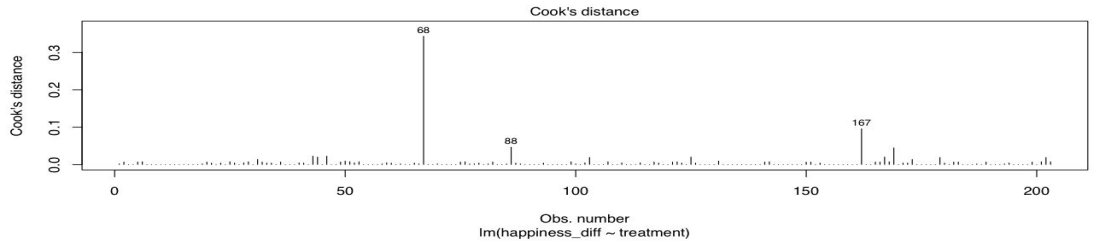

Figure 10: Plot of Cook's distance.

We are interested in ultimately adding baseline happiness as a covariate in the regression, because baseline happiness places bounds on the potential happiness change each participant could achieve. For example, a participant with a baseline happiness rating of 7 can only increase by 3 to 10, the top of our scale, and decrease by 6. This is very different from a participant who initially rates themself as a 3 on the happiness scale, as they have the potential to increase by 7 but only decrease by 2. We also examined baseline happiness and difference in happiness levels for each researcher, but there were no noticeable differences in these distributions across researchers.

Regression Results & Model Selection

Table 1 shows the regression models we ran, in accordance with the analytics plan outlined above, with every column in the table representing one model. We fit five linear models with our dependent variable, happiness difference', and one binary response model for which we used logistic regression. For the 'Binarized Difference in Happiness' we coded 0 to mean 'negative or no change in happiness' and 1 to mean 'positive change in happiness'. Table 3 outlines the regression results from all six models.

Focusing on the linear models without interaction terms - models (1), (4) and (5) - the ITT on happiness difference for Treatment 1 ("treat yourself") ranged between around 0.86 and 0.9, while the ITT for Treatment 2 ("treat someone else") was slightly higher in all of these models. For models (1), (4) and (5), the difference between the ITT for Treatment 2 and Treatment 1, which is the ITT we care about, ranged between +0.06 and +0.125. This means that "treat someone else" tended to have a directionally higher happiness difference than "treat yourself." The difference is not statistically significant in any of the models. This did not come as a surprise, given the conclusions from our pilot studies' power analysis indicating an underpowered comparison between Treatment 2 vs. Treatment 1 unless we increased our combined treatment group sample size from 104 to about 380. Unfortunately, due to time and budget constraints we were unable to increase our treatment group sample size beyond 104.

The covariates 'gender' (model 2) and 'researcher' (model 3) together with their respective 'treatment' interaction terms were overall not statistically significant according to F-Tests, as can be seen in Table 3. For the model with 'gender' and gender interactions (model 2), the ITT for Treatment 2 vs Treatment 1 is +0.22 (= 1.039 - 0.083 - 0.212 - [0.898 - 0.083 + 0.149]) for men and -0.141 (= 0.898 - 1.039) for women, which would indicate that men were more altruistic. This difference however is not statistically significant. For the model with 'researcher' and researcher interactions (model 3), the ITT for Treatment 2 vs. Treatment 1 depends on the researcher and ranges from around -0.538 for researcher Julia to around +0.923 for researcher Jayashree.

The covariate 'happiness_pre', an individual's happiness baseline, is statistically and practically significant in both models in which it is included. This makes sense, since a person's potential change in happiness is very much dependent on their baseline happiness, as we saw in the EDA section in Figure 8.

When we back-transform the coefficients of the logistic regression model, we receive an ITT of about 0.922 for Treatment 1 and 0.958 for Treatment 2. The difference between these ITTs is 0.036, which is smaller than any of the ITTs we receive for linear regression models. The binarization of the happiness difference variable seems to diminish the differences in the ITTs rather than 'clean out' noise, as we had hypothesized. The logistic regression model will therefore not be further considered as a candidate for the final model.

	Dependent variable:					
	happiness_diff					hapiness_diff_bin
	OLS					*logistic*
	(1)	(2)	(3)	(4)	(5)	(6)
treatment1	0.904***	1.039***	0.846*	0.864***	0.882***	2.467***
	(0.227)	(0.277)	(0.332)	(0.207)	(0.200)	(0.465)
treatment2	0.962***	0.898***	0.923***	0.989***	1.005***	3.128***
	(0.162)	(0.208)	(0.277)	(0.145)	(0.145)	(0.493)
genderM		−0.083			−0.112	−0.261
		(0.181)			(0.161)	(0.395)
treatment1:genderM		−0.212				
		(0.441)				
treatment2:genderM		0.149				
		(0.321)				
researcherJayashree			0.077		0.321	0.058
			(0.309)		(0.223)	(0.514)
researcherJulia			0.583*		0.484*	1.612**
			(0.242)		(0.231)	(0.524)
researcherMorris			0.130		0.154	0.416
			(0.206)		(0.194)	(0.525)
treatment1:researcherJayashree			−0.308			
			(0.490)			
treatment2:researcherJayashree			0.538			
			(0.527)			
treatment1:researcherJulia			0.032			
			(0.779)			
treatment2:researcherJulia			−0.583			
			(0.375)			
treatment1:researcherMorris			0.485			
			(0.479)			
treatment2:researcherMorris			0.177			
			(0.403)			
happiness_pre				−0.268**	−0.280**	−0.484***
				(0.094)	(0.099)	(0.127)
Constant	0.192*	0.222	0.000	2.157**	2.047**	1.387
	(0.095)	(0.134)	(0.170)	(0.724)	(0.739)	(0.944)
F-Test p-value against model (1)	n/a	7.80e-01	1.61e-01	4.80e-10	1.15e-08	n/a
Observations	203	203	203	203	203	203
Akaike Inf. Crit.	620.300	625.182	624.687	582.668	583.282	207.801
Note:						*p<0.05; **p<0.01; ***p<0.001

Table 3: Regression results with Control as omitted variable.

Finally, we used F-Tests and AIC to pick the 'best' model among the five linear models. As can be seen in Table 3, there are two candidates with low, almost equivalent AIC, models (4) and (5). Table 3 further shows that the same two models are also the only two models that showed statistical significance in F-Tests against the baseline model. We ended up choosing the more parsimonious model, model (4) as our final model - that is the model with only 'Baseline Happiness' as a covariate. For the final model, the ITT of Treatment 2 vs. Treatment 1 is 0.125, but - as mentioned above - it is not statistically significant. This can be seen from the standard errors respectively the resulting confidence interval.

To get proper test statistics, without examining the confidence intervals of the coefficients, we re-fit our final model using Treatment 1 as the reference level for the treatment variable (see Table 4). The 'treatment_alt0' level represents the control group and the 'treatment_alt2' level represents the treatment 2 group. From this model, we can directly read off the test statistics for the ITT of Treatment 2 vs. Treatment 1. The regression results confirm our previous assessment that the ITT between Treatment 2 vs. Treatment 1 is 0.125. The standard error for this ITT is roughly 0.212 and the p-value is around 0.557.

```
t test of coefficients:

               Estimate Std. Error t value  Pr(>|t|)
(Intercept)     3.020739   0.818586  3.6902 0.0002892 ***
treatment_alt0 -0.863947   0.206579 -4.1822 4.327e-05 ***
treatment_alt2  0.124769   0.212319  0.5876 0.5574343
happiness_pre  -0.268307   0.094421 -2.8416 0.0049560 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 4: Regression Results for Final Model

Bounds for Attrition & Non-compliance

Due to differential attrition and non-compliance between treatment groups, we calculated bounds for the measured ITT. Since our primary comparison of interest is Treatment 2 vs. Treatment 1, we decided to run two regressions with the same specification as our final model. For the first regression, we assigned all data points that exhibited attrition or non-compliance in Treatment 1 to the maximum observed outcome for Treatment 1, and all data points that exhibited attrition or non-compliance in Treatment 2 to the minimum observed outcome for Treatment 2. For the second regression we did this vice versa, with Treatment 1 data points as the minimum observed outcome and Treatment 2 data points as the maximum observed outcome. For our minimum and maximum value, we were originally planning on inserting the smallest and largest observation for the 'Difference in Happiness' in our data, but eventually decided to exclude one extreme outlier, who indicated an improvement of 9 points on the happiness scale. Our final minimum/maximum adjustments were -2 and +6, respectively.

The results from our bounds calculation are as follows:

- Bounds for ITT_treatment1: [0.45, 1.27]
- Bounds for ITT_treatment2: [0.84, 1.13]

The bounds are wider for Treatment 1 than for Treatment 2, since we had 6 cases of non-compliance/attrition for Treatment 1 and only 3 cases of non-compliance/attrition for Treatment 2.

From the bounds, we conclude that the minimum ITT for Treatment 2 vs. Treatment 1 is -0.44 (= 0.84 - 1.27), and the maximum ITT is +0.68 (= 1.13 - 0.45).

Discussion and Conclusions

In conclusion, we see that people become happier when they receive $5. This is not particularly surprising, as it seems reasonable that most people would enjoy receiving a windfall of any amount. However, they also become just as much happier when they give away $5, suggesting that they derive the same amount of happiness from prosocial spending. Both results are noteworthy, because $5 is a relatively small amount of money for most people (or at least for the typical Starbucks patron), and it seems to make a noticeable difference. For both treatment groups, we observed an average increase of 1 on the 10-point happiness rating scale. It's unclear what that actually means in terms of a person's state of mind, but it does show that the difference is large enough to the individual to report, which is a promising start for future studies.

It's very likely that we observed a demand effect in this study. Many participants figured out the significance of being asked twice about their happiness in this study. Some even said, "I should be happier, right?" However, this should have affected participants equally across treatment groups, and even in control (especially because they were asked about their order, some assumed that we were studying the effects of caffeine), so it was not a concern for this study.

Our study set out to answer the question is "Does spending on others make you happier than spending on yourself?" Although we observed directionally higher happiness differences in "treat someone else" compared to "treat yourself", it is important to note that we were unable to draw conclusions about the causal effect of giving a gift card compared to receiving a gift card, as the slight difference in a person's change in happiness between those two groups is not significant. As mentioned in our results section, this effect could have been better detected through larger treatment group sample sizes which we were unfortunately unable to implement given time and budget constraints.

However, it was still an interesting finding that the act of giving the gift card increased happiness compared to the baseline. It seems to be the case that money can in fact buy happiness, but so can doing good deeds such as giving money away.

It seems reasonable to conclude that these results could generalize to other Starbucks locations in other cities (especially because they were collected in various locations across the country). However, Starbucks patrons are a distinct subset of the population, so at this time it seems inappropriate to generalize to all Americans and Canadians. Future studies could explore whether this effect holds for different sums of money (previous work had explored $5 and $20 amounts) - perhaps the effects of giving and receiving money are only similar for small amounts, but they diverge as the amounts grow.

In future studies, we would like to be able to actually make the comparison we want, i.e. Treatment 1 vs. Treatment 2. A first step in doing this would be to increase the number of participants, especially in treatment groups, to increase the power of our study. Additionally, it could be worth exploring a few other potential sources that could affect whether we saw an effect. For example, what is the sustainability of the effect? We asked participants immediately after treatment about their happiness level, but the original study surveyed participants hours afterwards. It may be the case that the effect on people's happiness from giving a gift card is longer sustained, or even grows over time as participants get a chance to reflect on their prosocial behavior.

Further developments for this study include adding several other covariates to our analysis. We noticed that there were other environmental covariates that seemed to have a big effect (i.e., caused a noticeable change in a person's rating on the happiness scale). Almost all of the people who exhibited a decrease in happiness cited a long line or an inaccurate order as the reason. Additionally, consuming the beverage/food in the store before responding may have caused people to be happier than if they had not consumed it yet. We also noticed that people seemed to derive a greater change in happiness when they treated someone they didn't know and that person was very grateful, however, we did not consider these for our original study.

We hope that as the body of knowledge and science behind spending and happiness grows, people will start making better decisions to live happier lives by encouraging more prosocial behaviors. This could ultimately translate into parenting decisions, school programs and government policies that encourage and reward prosocial giving among their citizens.

References

Wikipedia contributors. (2018, May 14). Gross National Happiness. In Wikipedia, The Free Encyclopedia. Retrieved 12:55, May 31, 2018, from https://en.wikipedia.org/w/index.php?title=Gross_National_Happiness&oldid=841225085

Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending Money on Others Promotes Happiness. *Science, 319*(5870), 1687-1688. doi:10.1126/science.1150952

Dunn, E. W., Gilbert, D. T., & Wilson, T. D. (2011). Research Dialogue: If money doesn't make you happy, then you probably aren't spending it right. Journal Of Consumer Psychology, 21115-125. doi:10.1016/j.jcps.2011.02.002

Trochim, W. M., Donnelly, J. P., & Arora, K. (2016). *Research methods: The essential knowledge base*. Boston, MA: Cengage learning.