A Comparison of Efficacies of Altruistic & Egocentric Incentives

Isaac Chau, Emma Russon, Jeremiah (JJ) Sahabu

I. Abstract:

Organizations are often looking to maximize their productivity through different incentive programs for their employees. Outside of companies, motivation is sometimes difficult to muster, such as when asking students to fill out course evaluations. Our experiment seeks to uncover whether there are statistical differences between different types of incentives, in this case egocentric and altruistic ones. From our trials, we did not find statistical significance in the difference between incentives, though our results are inconclusive and require further exploration.

II. Introduction:

Background

Behavioral economics is a field of great interest to many large organizations, including corporations, universities, and governments. The ability to convince individuals at scale to perform the tasks that help achieve an organization's goals is invaluable. Incentives are often used as a tool in this regard. For example, companies pay employees wages for their time and work, while some religions promise the favor of deities for followers who do good works. The Hunger Site, a website popular in the aughts, convinced visitors to spend time clicking their mouse in order to donate food to the needy. In a similar vein, our project seeks to explore the motivational effects of small incentives on people asked to perform small tasks. The results of our work have many potential applications including but not limited to increasing survey response rates, recruiting participants for online crowdsourcing tasks, and even increasing voter turnout.

Research Question

Our research centers on two types of extrinsic motivation: egocentric and altruistic motivation. The purpose of this experiment is to investigate whether there is a difference in motivation to perform a task based on the type of incentive offered to an individual.

Hypothesis

There exists a statistically significant difference in motivation between receiving an egocentric or altruistic incentive to perform a task.

III. Experimental Design:

Experimental Setup and Treatment Conditions

The experiment we designed to test our hypothesis was centered around measuring the rate at which study subjects completed a small task on their personal computer. Our experiment had two treatment conditions which we will refer to as the Egocentric treatment condition and the Altruistic treatment condition. For each treatment condition, we created a website through which users were offered a reward in exchange for pressing their keyboard's spacebar key 400 times.

Each website was a single page with a title at the top, “The Spacebar Challenge,” and in the body there was text explaining the task and reward as well as a widget in the center that showed how many spacebar taps had been accumulated. From the perspective of potential study subjects, each site was identical except for a bolded phrase within the body that specified the reward that was being offered. (See Figures A1 and A2 in the appendix for screenshots of the websites.) The Egocentric treatment website offered users a $3 direct payment while the Altruistic treatment website offered users $3 to donate to any charitable organization of their choice (Figure 1). Besides this sentence in the body text, there was no difference in the website appearance or functionality until the user keyed their spacebar 400 times.

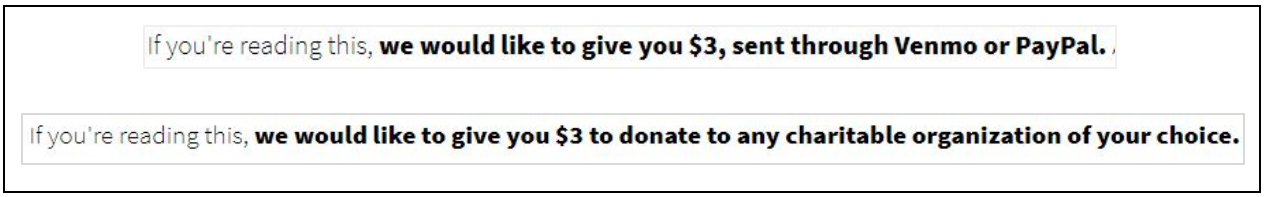


Figure 1. ***The unique message in the Egocentric treatment condition website (top) and the unique message in the Altruistic treatment condition website (bottom).***

Recruitment and Randomization

We used Facebook Ads to recruit participants for our experiment. We created two identical ads, each of which would link to one of the treatment websites. The ad placements we used were Desktop News Feed and Desktop Right Column. They appeared as shown in Figure 2.

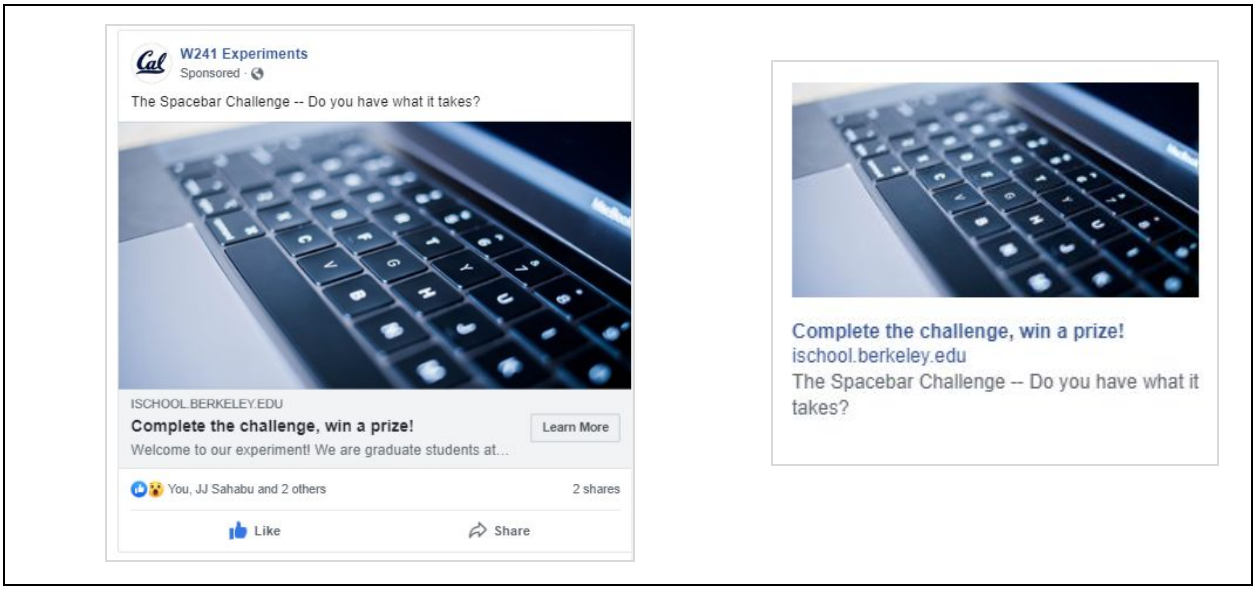


Figure 2. ***Appearances of the ad in the Desktop News Feed (left) and Desktop Right Column (right) placements. The ads for the links to the Egocentric and Altruistic treatment conditions were identical.***

To randomize the assignment of treatments to subjects, we used Facebook Ads’ built-in A/B testing feature which is normally used by advertisers to create experiments to test the efficacy of

different ad designs. The A/B testing tool created two ad campaigns, one for each of our ads, that were served to our target audience--Facebook users in the United States who were 18 years and older--according to Facebook's internal randomization procedure. Thus, each of our two treatment groups' ads was served to a randomized sample of our study population over the lifetime of the ad campaigns. The ads were run from 9 AM to 9 PM PST to allow for constant monitoring. By the end of the study, each ad had been served to more than 4500 users each and clicked on by more than 200 users each. See Figure 3 for more details on sampling and selection of study participants.

In order for a Facebook user to enter our study as a participant, they must have been served one of our ads on their desktop personal computer, clicked the ad, and loaded the linked webpage on their internet browser. Because the ad for each treatment group was identical in appearance to that of the other, we did not consider a potential study participant to be part of a treatment group until they entered our webpage through the ad they had been randomly assigned to receive.

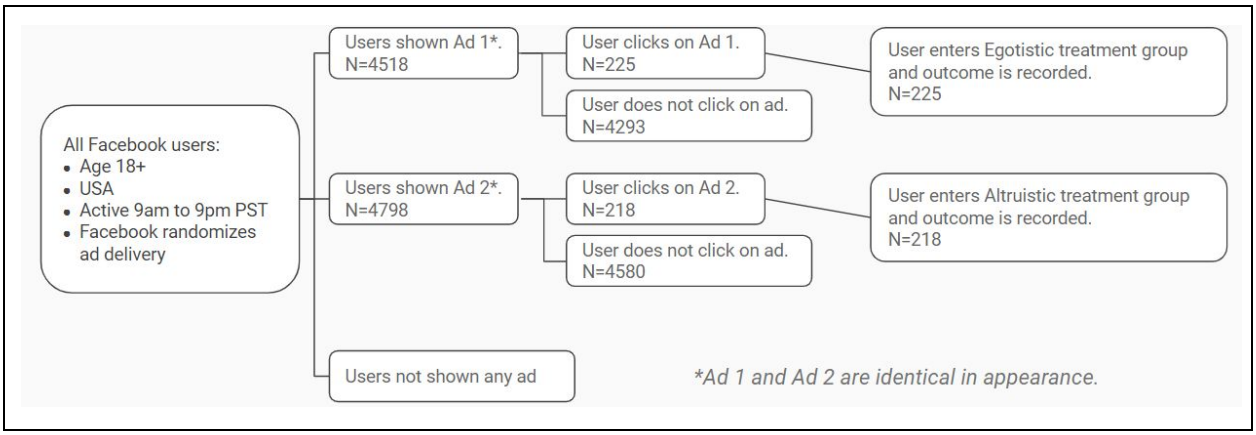


Figure 3. ***Flow diagram illustrating the sampling, treatment assignment, and outcome collection process. Possibility of noncompliance was practically eliminated by the experimental design. Possibility of attrition was practically eliminated by the method through which data was collected.***

Data Collection and Outcome Measures

Once a user loaded the webpage to which their assigned ad was linked, they were exposed to the treatment condition message (either Egocentric or Altruistic). At this point, the user became a study subject. The website they had navigated to would automatically begin recording information for several measures. It recorded the timestamp of the moment at which the user entered the page. It also recorded the length of time that elapsed between the user entering the page and either exiting the page or finishing the 400th tap of the spacebar. The number of times the spacebar key had been pressed was recorded and associated for each user whether they completed the task or not. If the user completed the 400 spacebar key presses, the webpage would show a form for the user to enter their email address and information for either the direct payment or charitable donation. Upon the user's action of closing the web page or, if the user completed the spacebar task, the submission of the reward information form, the web page

would send the recorded data to a database we maintained that was not associated with Facebook. Because Facebook users only became study participants when they loaded our web page and data collection for every subject was automated by the web page, attrition from the study was not possible.

Note that subject non-compliance (i.e. using the web page for the treatment group to which they were not assigned) was also practically impossible, as the users were not aware of the existence of another treatment group, and the only way for a user to access the other treatment website was by guessing its url.

The main outcome measure we were interested in was the proportion of users who completed the task of 400 spacebar taps for each treatment group. Other outcome measures we collected to examine treatment effects were the number of spacebar taps each user registered as well as the elapsed time between opening the webpage and either completing 400 spacebar taps or exiting the page by closing the window.

Power and Sample Size

Lack of previous experiments similar to our own made it difficult for us to estimate our ideal sample size, as we could only make an educated guess as to what effect size we would observe. If the task completion rates of the two treatment groups were to differ by 20%, we would require at least 162 total participants distributed between the two groups in order to reach a power level of 0.8 and a significant level of 0.95. For the same power and significance levels, a 10% difference in completion rates would require a sample size of at least 586 total participants. (See Appendix B for power calculation details). Due to this uncertainty, we maximized our sample size by allowing our experiment's two ad campaigns to run continuously under constant supervision until the combined expenditure on advertising and participant compensation reached our budget of $500.

IV. Obstacles:

During our experiment, we came across a few obstacles in which we believe could have biased our results. Please see below for the different ways we tackled these obstacles.

Ad Blockers

We recognized that a good portion of internet users use ad blockers. As such, there was a portion of internet users who were not exposed to our experiment and as a result, never would have been selected. Even though the sample selection for both treatment groups were equally affected by users who have ad blockers that prevent participation in our experiment, we saw this as a challenge to the external validity of our findings. As supplemented by the demographic distribution of our participants in Figure B1 in Appendix 2, we can see that our sample contains a relatively smaller proportion of individuals between 20 and 40 years old. We believe this is due to the use of Ad Blockers by users in this age group, which tends to be more technologically savvy than children and seniors. Knowing this, we hypothesize that it is unlikely that individuals

who use ad blockers experience highly different motivational responses to the same incentives as non-users.

Ad Consistency

Our experiment counted any individual who clicked into our site as a participant in our experiment. As a result, we wanted to ensure that the ads' appearances were consistently identical so that the only effects observed were attributable to the treatment variable. During our experiment, there were individuals who "shared" our ad, commented, and "reacted" to the ad. Our team was diligent in removing all comments, shares, and reactions from the ads in a timely manner, so we are confident that the appearance of the ads leading to the two treatment conditions were identical throughout the experiment. Nevertheless, we address the potential impact of shares on the non-interference assumption in the "Analysis" section.

V. Results & Analysis:

Data for Analysis

Since we wanted to test if there was a difference between the completion rates of a simple task when given an Altruistic or Egocentric incentive, our main outcome of interest was the number of subjects who completed the challenge out of all subjects assigned to a specific treatment group. We measured this by recording the number of spacebar taps each subject performed. Each treatment website was connected to a database on the backend that collected data from a subject when they closed the website or submitted the compensation information. For the subjects who exited the website without attempting the challenge, their observed number of taps was stored as 0 in the database. In turn, we had observations for every subject who clicked into our treatment websites.

The number of taps from every observation contributed to our analysis in two ways. To measure the primary outcome of interest, we created a binary variable that indicated if a subject reached 400 taps and calculated the average of this variable to get the completion rate. Furthermore, we decided that it would be interesting to see if the Egocentric and Altruistic treatment incentives had a different effect on the effort each subject put forth in the challenge. For this secondary outcome variable, we used the raw number of taps a subject performed as a proxy for effort, following the logic that more taps reflected more effort. The distribution of the number of spacebar taps from subjects and the respective completion rates for each treatment group is displayed in Figure 4.

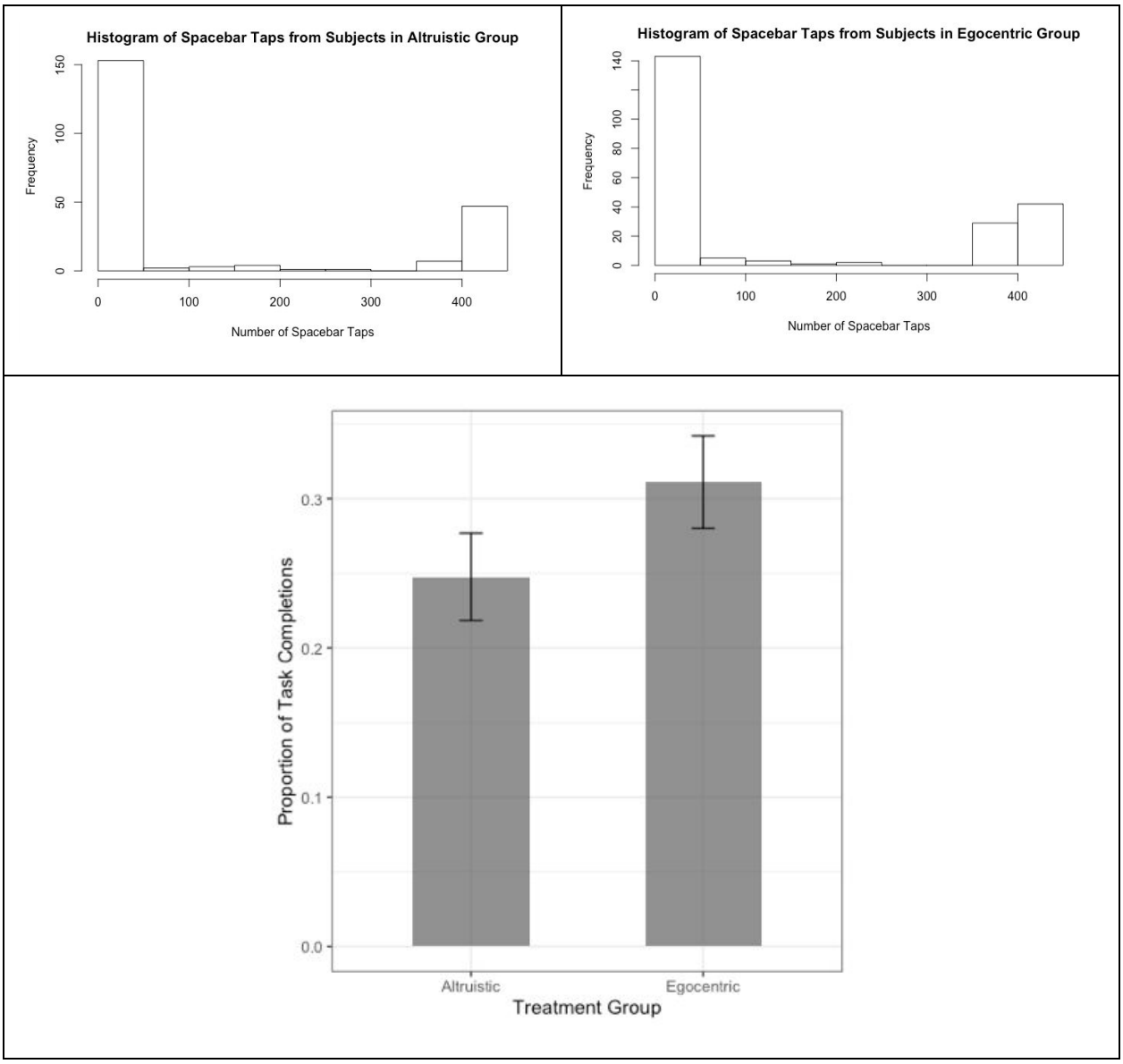


Figure 4. ***Distribution of spacebar tap counts from subjects in the Altruistic treatment group (left) and the Egocentric treatment group (right). Distribution of completion rate for each treatment group (bottom).***

Out of the 225 subjects assigned to the Egocentric treatment, 70 of them completed the spacebar challenge. For the Altruistic treatment group, 54 out of 218 subjects completed the challenge. This resulted in completion rates of 31.1% for the Egocentric treatment group and 24.8% for the Altruistic treatment group. In a simple difference in means test, the difference was not statistically significant, with a p-value of 0.138 (Figure 4). In other words, with the Altruistic incentive as the baseline, the effect size of 0.063 that the Egocentric incentive had on the completion rate is not statistically different from 0.

To incorporate covariates and additional outcomes in our analysis, we recorded the timestamp at which a subject clicked into the treatment website and the length of time a subject spent on the treatment website. The discrete hours from the timestamp were used as pretreatment covariates in our models to control for variability in our outcomes due to the time of day the subject was assigned to the treatment. For example, a subject who clicked into the treatment website at a later time during the day might be more likely to complete the challenge on the Altruistic treatment website after watching the evening news and seeing all the medical professionals risking their lives to combat the COVID-19 pandemic. However, after conducting a regression analysis, we found that including these pretreatment covariates did not significantly change our estimated treatment effect. (See “Regression Results.” The distribution of completions in every observed hour for each treatment group is displayed in Figure 5.)

We took the length of time users spent on the page as our third outcome measure. This was the time from when the subject clicked into the page to when the subject closed the page. It encompassed the length of time a subject took to read the website content, decided to attempt the challenge, and potentially completed the challenge. We decided this would be an interesting post-treatment variable to consider, since the two treatment incentives might have a different effect on how long a subject took to consider whether or not they wanted to attempt to complete the challenge. However, our analysis of this outcome variable also yielded a statistically insignificant difference between the two treatment groups. (See “Regression Results.” The distribution of the time in seconds subjects spent on the treatment website for each treatment group is displayed in Figure 6.)

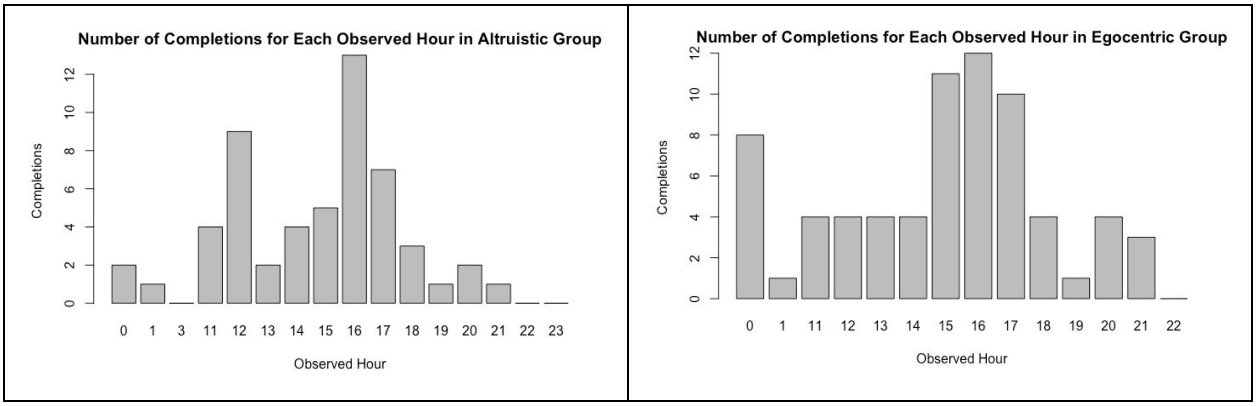


Figure 5. ***Distribution of the number of completions for each observed hour in the Altruistic treatment group (left) and the Egocentric treatment group (right).***

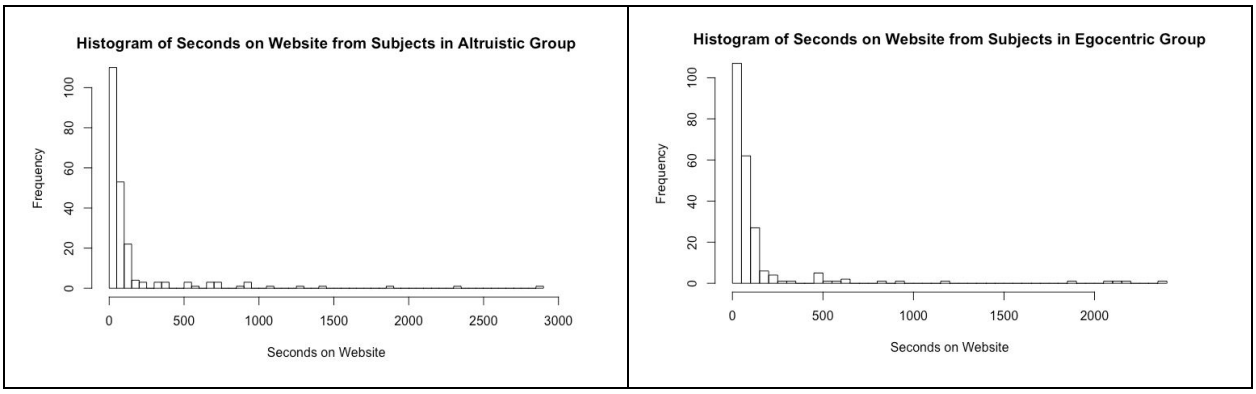


Figure 6. ***Distribution of the amount of time in seconds spent on the treatment website from subjects in the Altruistic treatment group (left) and the Egocentric treatment group (right).***

Models

For the main outcome of interest, our base model to estimate the simple "treatment-control" contrast on task completion rate was operationalized as shown below, with the Altruistic incentive acting as the control and the Egocentric incentive acting as the treatment.

$$Completion\ Rate = \beta_0 + \beta_1 Treatment + \varepsilon \quad (1)$$

To improve the precision of the estimates, we included the covariate that indicated in which hour subjects entered the treatment website.

$$Completion\ Rate = \beta_0 + \beta_1 Treatment + \beta_2 Hour + \varepsilon \quad (2)$$

The models for our secondary outcomes of interest were operationalized as follows.

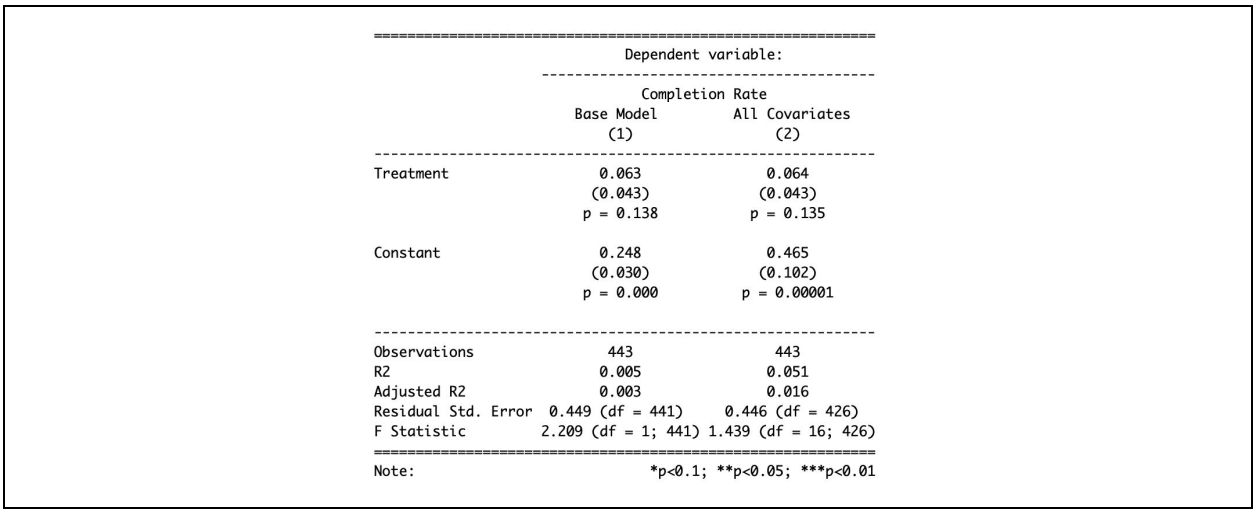
$$Effort = \beta_0 + \beta_1 Treatment + \varepsilon \quad (3)$$
$$Effort = \beta_0 + \beta_1 Treatment + \beta_2 Hour + \varepsilon \quad (4)$$

$$log(Time\ Elapsed) = \beta_0 + \beta_1 Treatment + \varepsilon \quad (5)$$
$$log(Time\ Elapsed) = \beta_0 + \beta_1 Treatment + \beta_2 Hour + \varepsilon \quad (6)$$

Since the distribution of the time in seconds spent on the treatment website had a strong right skew for both treatment groups, a log transform was applied to the Time Elapsed variable.

Regression Results

The regression results for the models of task completion rates are displayed below. We performed Levene's test to determine whether or not we needed heteroskedastic robust standard errors, and concluded that the variance was not significantly different along the hour covariate; thus robust standard errors were not used (Appendix E).

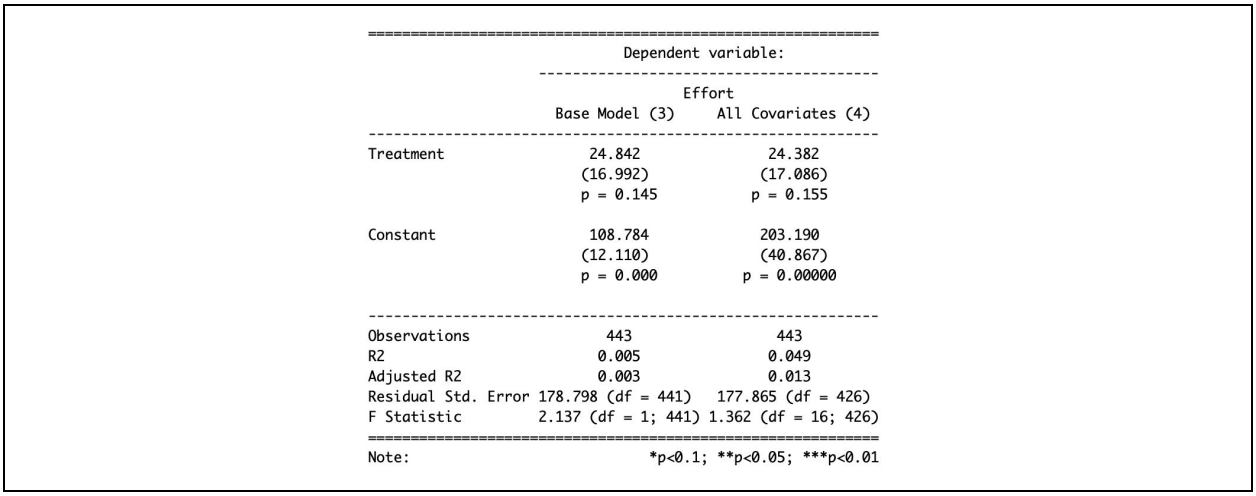


	Dependent variable:	
	Completion Rate	
	Base Model (1)	All Covariates (2)
Treatment	0.063	0.064
	(0.043)	(0.043)
	p = 0.138	p = 0.135
Constant	0.248	0.465
	(0.030)	(0.102)
	p = 0.000	p = 0.00001
Observations	443	443
R2	0.005	0.051
Adjusted R2	0.003	0.016
Residual Std. Error	0.449 (df = 441)	0.446 (df = 426)
F Statistic	2.209 (df = 1; 441)	1.439 (df = 16; 426)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 7. ***Regression results for model with just treatment variable (left) and for model with both treatment and hour covariate (right) on completion rates, with the Altruistic treatment as the baseline. Note: Hour covariate results were hidden. Full regression results can be found in Appendix D.***

Looking at the base model, the estimated difference in completion rates for the Egocentric group was 0.063 (0.043), which was not statistically significant with a p-value of 0.138. Including the covariate for the hour at which subjects entered the treatment website did not explain much of the variability and improve the model estimates; the difference in task completion rates between the Egocentric and Altruistic groups did not change to become statistically significant.

Next, the regression results for the models of a subject's effort on attempting the challenge are displayed below. We performed another Levene's test on this outcome variable and concluded that the variance was not significantly different along the hour covariate; thus robust standard errors were not used (Appendix E).

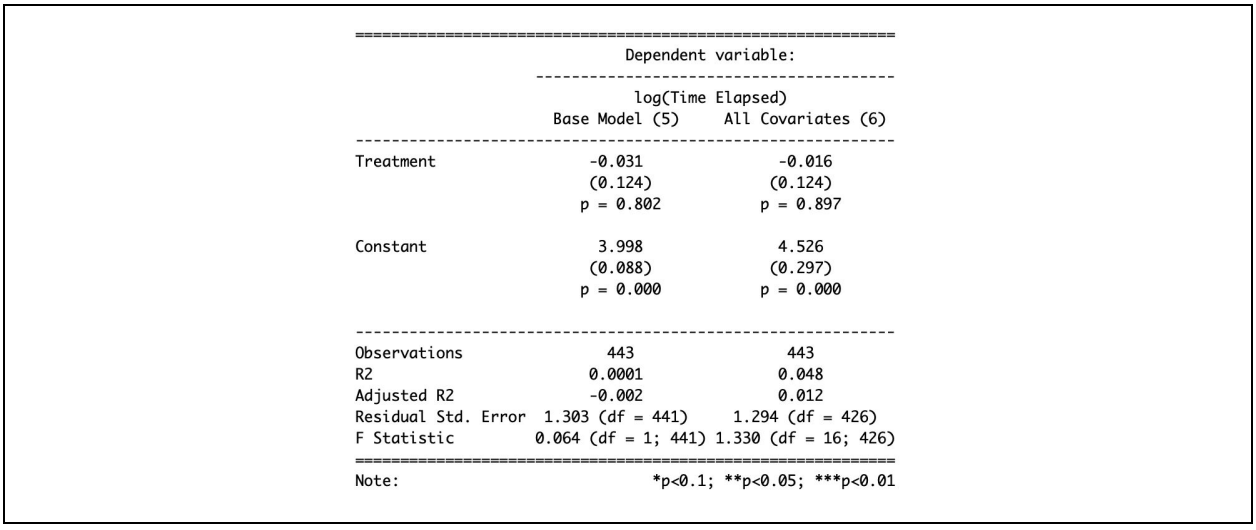


	Dependent variable:	
	Effort	
	Base Model (3)	All Covariates (4)
Treatment	24.842	24.382
	(16.992)	(17.086)
	p = 0.145	p = 0.155
Constant	108.784	203.190
	(12.110)	(40.867)
	p = 0.000	p = 0.00000
Observations	443	443
R2	0.005	0.049
Adjusted R2	0.003	0.013
Residual Std. Error	178.798 (df = 441)	177.865 (df = 426)
F Statistic	2.137 (df = 1; 441)	1.362 (df = 16; 426)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 8. ***Regression results for model with just treatment variable (left) and for model with both treatment and hour covariate (right) on effort, with the Altruistic treatment as the baseline. Note: Hour covariate results were hidden. Full regression results can be found in Appendix D.***

Considering the base model, the estimated difference in effort exerted from subjects in the Egocentric group was 24.84 (16.99) taps, which was not statistically significant with a p-value of 0.145. Again, including the covariate for the hour at which subjects entered the treatment website did not improve the model estimates; the difference in effort between the Egocentric and Altruistic groups did not change to become statistically significant.

Finally, the regression results for the models of the time subjects spent on the treatment website are displayed below. We performed one more Levene's test on this outcome variable and concluded that the variance was not significantly different along the hour covariate; thus robust standard errors were not used (Appendix E).



	Dependent variable:	
	log(Time Elapsed)	
	Base Model (5)	All Covariates (6)
Treatment	-0.031	-0.016
	(0.124)	(0.124)
	p = 0.802	p = 0.897
Constant	3.998	4.526
	(0.088)	(0.297)
	p = 0.000	p = 0.000
Observations	443	443
R2	0.0001	0.048
Adjusted R2	-0.002	0.012
Residual Std. Error	1.303 (df = 441)	1.294 (df = 426)
F Statistic	0.064 (df = 1; 441)	1.330 (df = 16; 426)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 9. ***Regression results for model with just treatment variable (left) and for model with both treatment and hour covariate (right) on log of time elapsed, with the Altruistic treatment as the baseline. Note: Hour covariate results were hidden. Full regression results can be found in Appendix D.***

Looking at the base model, the time a subject in the Egocentric group spent on the treatment website is approximately 3% (0.124) less than the time subjects on the Altruistic group spent on the website. However, this estimate was not statistically significant with a very large p-value of 0.802. Again, including the covariate for the hour at which subjects entered the treatment website did not improve the model estimates; the difference in effort between the Egocentric and Altruistic groups did not change to become statistically significant.

Sample Size and Achieving Statistical Significance

Our study's sample size was limited by financial constraints. Given our $500 budget and commitment to reward each subject who completed the spacebar task with $3, we knew our experiment would be limited by the number of participants who earned the compensation. Our sample's respective task completion rates of 31.1% and 24.8% for the Egocentric and Altruistic treatment groups did not achieve a statistically significant difference. As a thought experiment, let us imagine that our point estimate in the difference in completion rates for the two treatments

reflected the true value, and our ads had recruited study participants according to the same proportion observed in our experiment's data ($N_{Egocentric}$=225, $N_{Altruistic}$=218). Then we would have required the assignment of 798 subjects to the Egocentric condition and 773 subjects to the Altruistic condition for a total of 1571 subjects in order to reach statistical significance at α=0.05 and β=0.20. (See Appendix B for calculation details). The costs of this hypothetical experiment would be nearly $1320 for participant compensation and about $230 for Facebook advertising for a total of $1550.

Check on Interference from "Shares"

One issue that arose during the experiment was caused by two users "sharing" the News Feed placement ad for the Altruistic treatment website. This had the potential to bias our estimates because subjects not randomly assigned the treatment by Facebook could have entered our study through the shared posts. However, the number of participants in the Altruistic treatment group did not seem to reflect illegitimate subject recruitment since it had fewer subjects assigned to it, despite the sharing. Additionally, we propose that any illegitimate participants in the Altruistic treatment would have been more likely to have completed the spacebar task, when encouraged by the subject who had likely completed the task themselves and shared the page. Any additional task completions in the Altruistic group would have reduced our measurement of the main treatment effect, making our estimate more conservative.

To address the possible violation of the non-interference assumption, we conducted a difference in means test to see if there was a statistically significant difference in the rate of task completions before and after the first share of the Altruistic treatment website. With a p-value of 0.2173, there was no evidence to show that the difference in completion rates before and after the share was statistically significant. This is good evidence that our estimates are fairly unbiased, and non-interference was satisfied.

VI. Conclusion:

Overall, we are pleased with the results of our experiment. Although the differences in the effects of our two treatment incentives were not statistically significant, there nevertheless was an observable treatment effect that would be practically significant if future research reveals similar results. If we were to iterate on this experiment, a larger budget, changes in parameters such as the effort or skill to complete the task, and changes in the magnitude of compensation could help increase the likelihood of significant results by enlarging the treatment effect and/or allowing for a larger sample size. Future experiments conducted by researchers with better web development skills could integrate Facebook Ads' tools to collect demographic information about individual subjects for subgroup analysis. Given our project timeline and novice web development skills, this was not a possibility for our experiment. In spite of this, our current analysis among multiple

VII. Appendix

Appendix A: Website Appearances

Figure A1. ***Screenshot of the web page for the Egocentric treatment condition.***

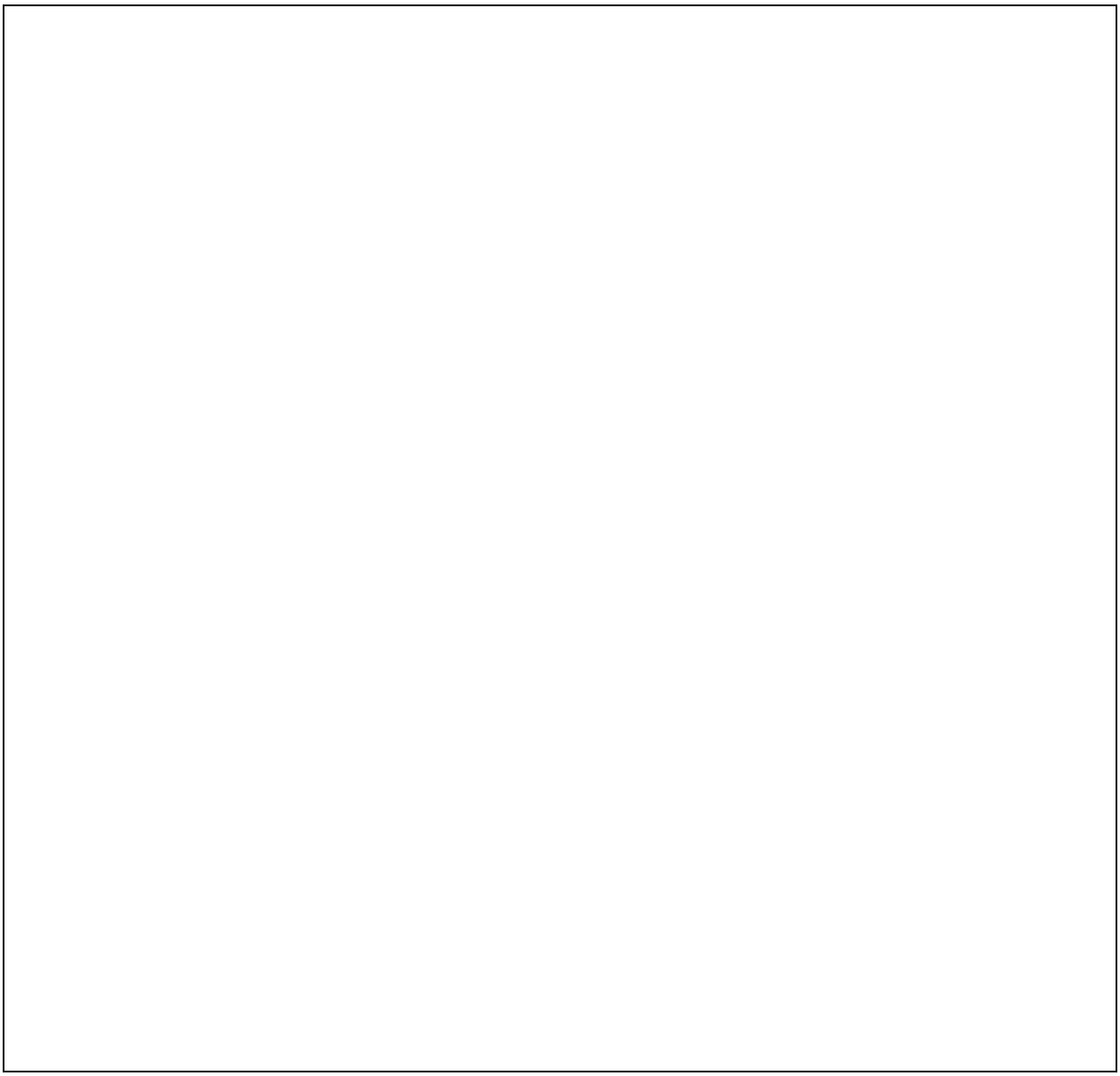


Figure A2. ***Screenshot of the web page for the Altruistic treatment condition.***

Appendix B: Power Calculations

$$N_1 = \left\{ z_{1-\alpha/2} * \sqrt{\bar{p} * \bar{q} * (1 + \frac{1}{k})} + z_{1-\beta} * \sqrt{p_1 * q_1 + (\frac{p_2 * q_2}{k})} \right\}^2 / \Delta^2$$

$$q_1 = 1 - p_1$$

$$q_2 = 1 - p_2$$

$$\bar{p} = \frac{p_1 + k p_2}{1 + K}$$

$$\bar{q} = 1 - \bar{p}$$

$$N_1 = \left\{ 1.96 * \sqrt{0.3 * 0.7 * (1 + \frac{1}{1})} + 0.84 * \sqrt{0.2 * 0.8 + (\frac{0.4 * 0.6}{1})} \right\}^2 / 0.2^2$$

$$N_1 = 81$$

$$N_2 = K * N_1 = 81$$

p_1, p_2 = proportion (incidence) of groups #1 and #2
$\Delta = |p_2 - p_1|$ = absolute difference between two proportions
n_1 = sample size for group #1
n_2 = sample size for group #2
α = probability of type I error (usually 0.05)
β = probability of type II error (usually 0.2)
z = critical Z value for a given α or β
K = ratio of sample size for group #2 to group #1

Figure B1. ***Power calculation given = 0.05 and = 0.20 and prospective task completion rates of Treatment$_i$ = 0.40 and Treatment$_j$ = 0.20 for a difference of 0.20.***

Kane SP. Sample Size Calculator. ClinCalc: https://clincalc.com/stats/samplesize.aspx. Updated July 24, 2019. Accessed April 19, 2020.

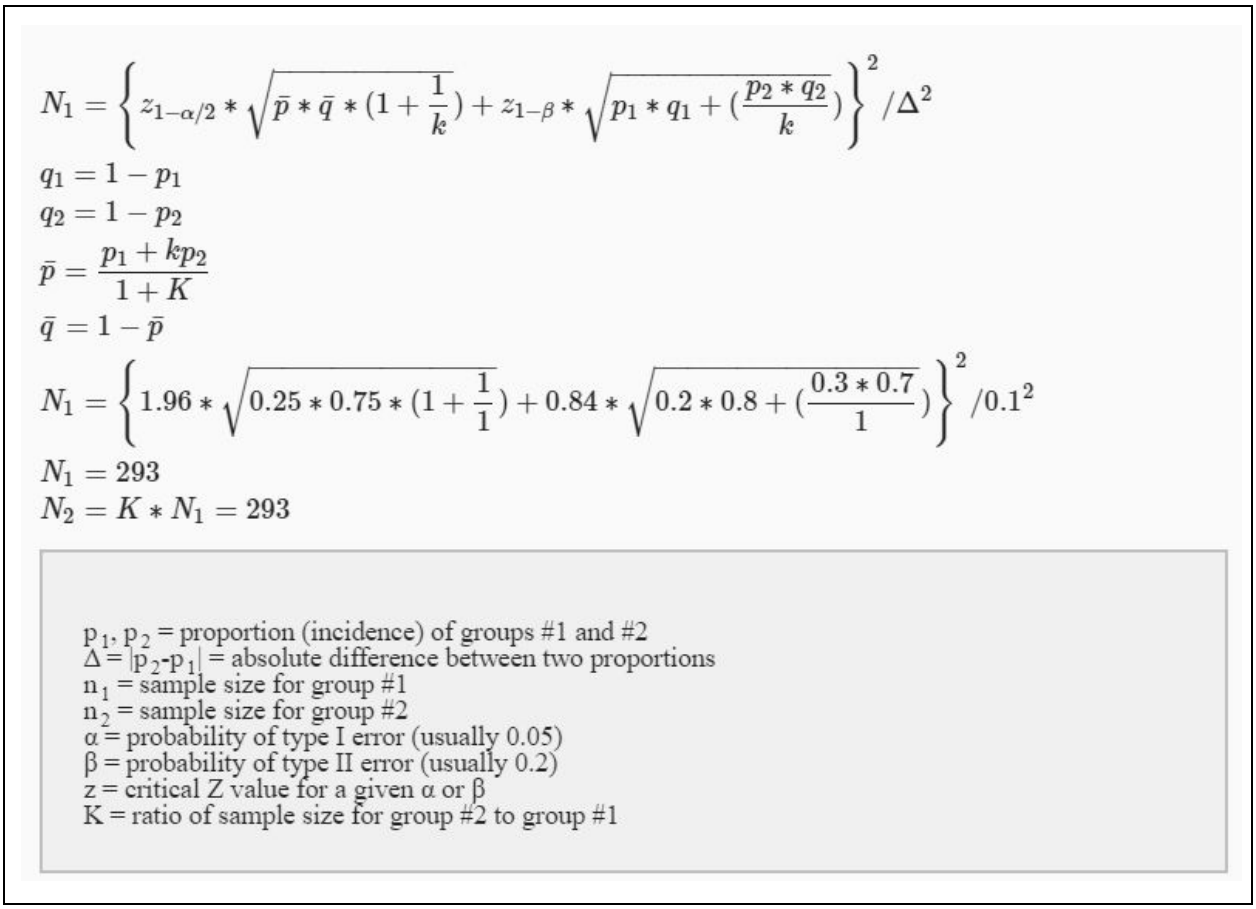


Figure B2. ***Power calculation given = 0.05 and = 0.20 and prospective task completion rates of treatment$_i$ = 0.30 and treatment$_j$ = 0.20 for a difference of 0.10.***

Kane SP. Sample Size Calculator. ClinCalc: https://clincalc.com/stats/samplesize.aspx. Updated July 24, 2019. Accessed April 19, 2020.

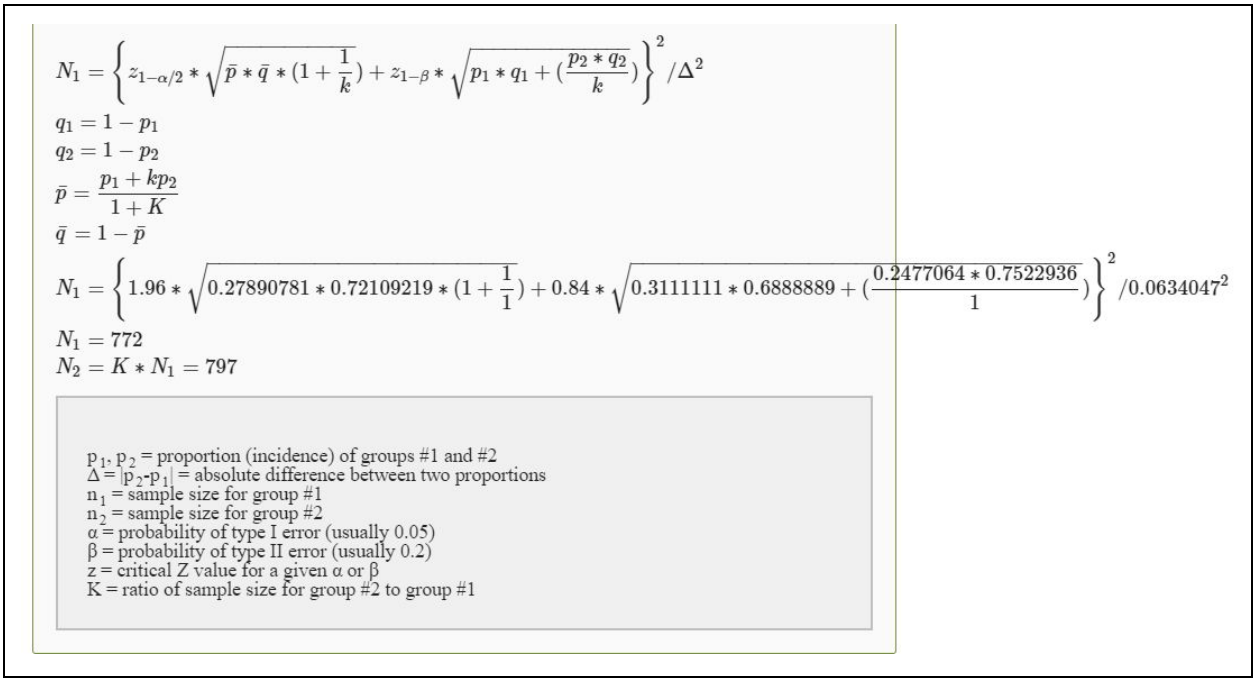


Figure B3. ***Sample size calculation given = 0.05 and = 0.20 and observed task completion rates of treatment$_{Egocentric}$ = 0.311 and treatment$_{Altruistic}$ = 0.248 for a difference of 0.0634.***

Kane SP. Sample Size Calculator. ClinCalc: https://clincalc.com/stats/samplesize.aspx. Updated July 24, 2019. Accessed April 19, 2020.

Appendix C: Participant Demographic Distribution

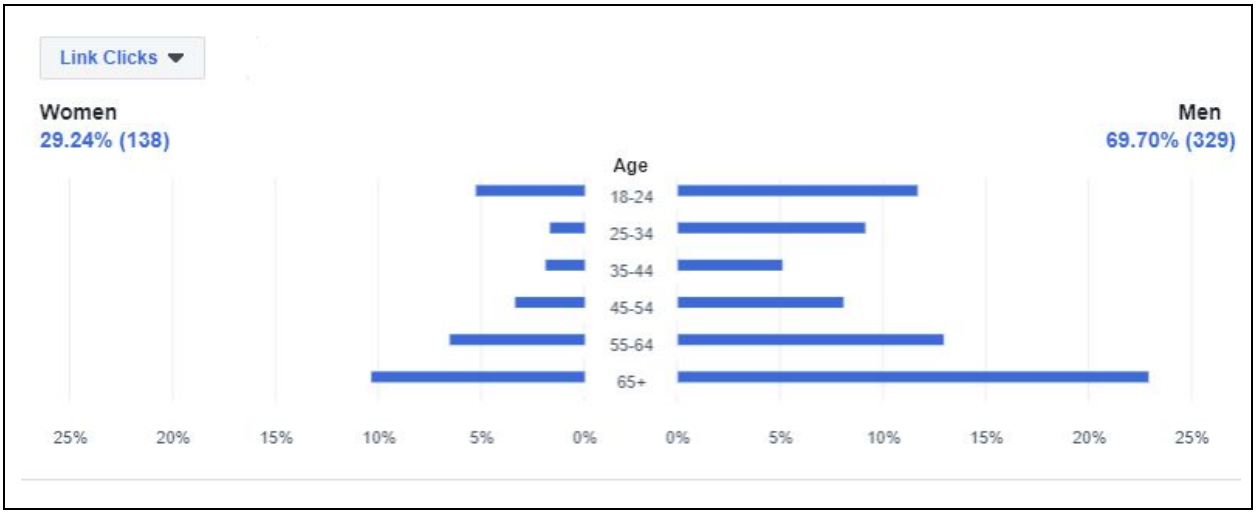
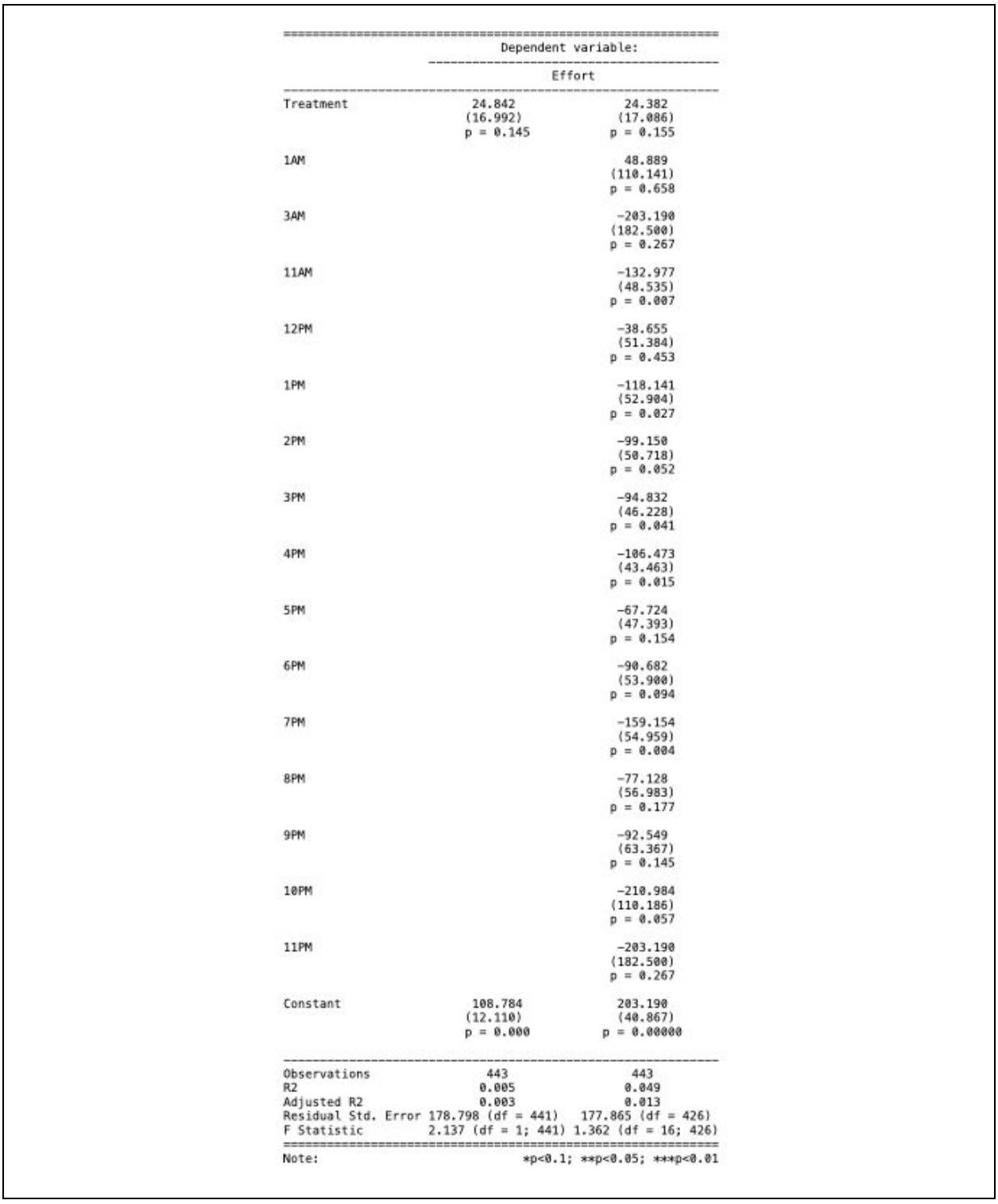


Figure C1. ***Screenshot of the distribution of ages of men and women for the experiment.***

Appendix D: Regression Tables

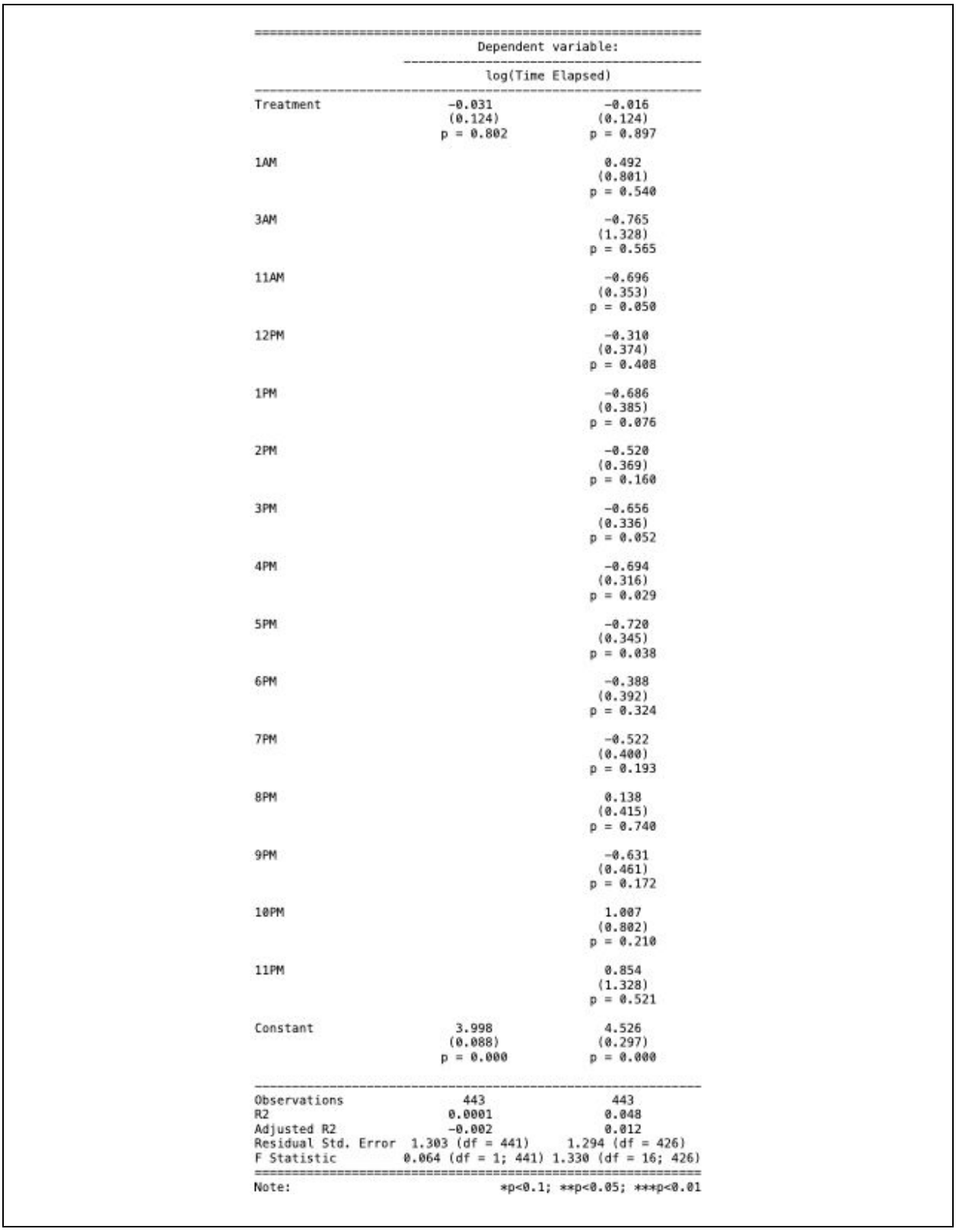
	Dependent variable:	
	Completion Rate	
	(1)	(2)
Treatment	0.063	0.064
	(0.043)	(0.043)
	p = 0.138	p = 0.135
1AM		0.159
		(0.276)
		p = 0.565
3AM		−0.465
		(0.458)
		p = 0.311
11AM		−0.299
		(0.122)
		p = 0.015
12PM		−0.059
		(0.129)
		p = 0.647
1PM		−0.271
		(0.133)
		p = 0.042
2PM		−0.245
		(0.127)
		p = 0.055
3PM		−0.221
		(0.116)
		p = 0.058
4PM		−0.258
		(0.109)
		p = 0.019
5PM		−0.137
		(0.119)
		p = 0.249
6PM		−0.200
		(0.135)
		p = 0.141
7PM		−0.406
		(0.138)
		p = 0.004
8PM		−0.183
		(0.143)
		p = 0.202
9PM		−0.192
		(0.159)
		p = 0.229
10PM		−0.486
		(0.276)
		p = 0.080
11PM		−0.465
		(0.458)
		p = 0.311
Constant	0.248	0.465
	(0.030)	(0.102)
	p = 0.000	p = 0.00001
Observations	443	443
R2	0.005	0.051
Adjusted R2	0.003	0.016
Residual Std. Error	0.449 (df = 441)	0.446 (df = 426)
F Statistic	2.209 (df = 1; 441)	1.439 (df = 16; 426)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure D1. ***Regression results for model with just treatment variable (left) and for model with both treatment and hour covariate (right) on completion rates, with the Altruistic treatment as the baseline.***



	Dependent variable:	
	Effort	
Treatment	24.842	24.382
	(16.992)	(17.086)
	p = 0.145	p = 0.155
1AM		48.889
		(110.141)
		p = 0.658
3AM		−203.190
		(182.500)
		p = 0.267
11AM		−132.977
		(48.535)
		p = 0.007
12PM		−38.655
		(51.384)
		p = 0.453
1PM		−118.141
		(52.904)
		p = 0.027
2PM		−99.150
		(50.718)
		p = 0.052
3PM		−94.832
		(46.228)
		p = 0.041
4PM		−106.473
		(43.463)
		p = 0.015
5PM		−67.724
		(47.393)
		p = 0.154
6PM		−90.682
		(53.900)
		p = 0.094
7PM		−159.154
		(54.959)
		p = 0.004
8PM		−77.128
		(56.983)
		p = 0.177
9PM		−92.549
		(63.367)
		p = 0.145
10PM		−210.984
		(110.186)
		p = 0.057
11PM		−203.190
		(182.500)
		p = 0.267
Constant	108.784	203.190
	(12.110)	(40.867)
	p = 0.000	p = 0.00000
Observations	443	443
R2	0.005	0.049
Adjusted R2	0.003	0.013
Residual Std. Error	178.798 (df = 441)	177.865 (df = 426)
F Statistic	2.137 (df = 1; 441)	1.362 (df = 16; 426)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure D2. ***Regression results for model with just treatment variable (left) and for model with both treatment and hour covariate (right) on effort, with the Altruistic treatment as the baseline.***



	Dependent variable:	
	log(Time Elapsed)	
Treatment	−0.031	−0.016
	(0.124)	(0.124)
	p = 0.802	p = 0.897
1AM		0.492
		(0.801)
		p = 0.540
3AM		−0.765
		(1.328)
		p = 0.565
11AM		−0.696
		(0.353)
		p = 0.050
12PM		−0.310
		(0.374)
		p = 0.408
1PM		−0.686
		(0.385)
		p = 0.076
2PM		−0.520
		(0.369)
		p = 0.160
3PM		−0.656
		(0.336)
		p = 0.052
4PM		−0.694
		(0.316)
		p = 0.029
5PM		−0.720
		(0.345)
		p = 0.038
6PM		−0.388
		(0.392)
		p = 0.324
7PM		−0.522
		(0.400)
		p = 0.193
8PM		0.138
		(0.415)
		p = 0.740
9PM		−0.631
		(0.461)
		p = 0.172
10PM		1.007
		(0.802)
		p = 0.210
11PM		0.854
		(1.328)
		p = 0.521
Constant	3.998	4.526
	(0.088)	(0.297)
	p = 0.000	p = 0.000
Observations	443	443
R2	0.0001	0.048
Adjusted R2	−0.002	0.012
Residual Std. Error	1.303 (df = 441)	1.294 (df = 426)
F Statistic	0.064 (df = 1; 441)	1.330 (df = 16; 426)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure D3. ***Regression results for model with just treatment variable (left) and for model with both treatment and hour covariate (right) on log of time elapsed, with the Altruistic treatment as the baseline.***

Appendix E: Levene's Test for Heteroskedasticity

```r
# every hour as a categorical variable for completion
model1 <- lm(data$complete ~ factor(data$treat))
model2 <- lm(data$complete ~ factor(data$treat) * factor(data$hour) )
```

```r
leveneTest(model1)
leveneTest(model2)
```

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1  2.2094 0.1379
      441
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group  29  0.8388  0.709
      413
```

Figure E1. ***Levene's test results for completion rate models.***

```r
# every hour as a categorical variable for completion
model3 <- lm(data$taps ~ factor(data$treat))
model4 <- lm(data$taps ~ factor(data$treat) * factor(data$hour))
```

```r
leveneTest(model3)
leveneTest(model4)
```

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1  2.1374 0.1445
      441
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group  29  0.7275 0.8502
      413
```

Figure E2. ***Levene's test results for effort models.***

```r
# every hour as a categorical variable for completion
model5 <- lm(log(data$seconds) ~ factor(data$treat))
model6 <- lm(log(data$seconds) ~ factor(data$treat) * factor(data$hour))
```

```r
leveneTest(model5)
leveneTest(model6)
```

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1  0.0442 0.8335
      441
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group  29  1.0613 0.3824
      413
```

Figure E3. ***Levene's test results for log time elapsed models.***