Michael Yazdani
Julie Lai
Daniel Lampert
W241 Final

Effect of Grey Scale on Screen Time

Abstract

During the coronavirus pandemic, our lives have moved onto our phones, and in many cases in an unhealthy way. A high amount of screen time and phone use can have a serious effect on mental health and physical well being. To tackle this issue, our project explores the use of removing colors from mobile phones to reduce overall screen time. We gathered around 50 participants for a 3 day experiment to test if reducing the color stimulus causes individuals to use their phones less compared to their prior averages. This project's results show that even with a specific treatment to make phones unappealing to use, there is not a statistically significant change in phone usage after changing the screen filter to grayscale

Background

Over the years people are spending an increasing amount of time on cell phones. In the United States it is estimated that the average individual spends more than four hours on their cell phone a day. Consequences of long-term digital exposure exist in forms such as eye strains, migraines, low attention spans, and quality of sleep. There have already been efforts to reduce screen time, such as setting time limits on apps, wearing blue light glasses, and turning night-shift mode on. Our team aims to explore the hypothesis that long digital activity can be caused by visual (color) stimulation. In particular, we planned to investigate the effect of turning screen settings to grayscale on the length of an individual's screen time activity.

Hypothesis

The question we attempt to answer in this experiment is: Does changing a cell phone screen to black and white reduce cell phone usage?

We are optimistic that changing the screen to black and white will reduce screen time because it effectively makes the phone less interesting and therefore people will be less likely to spend a lot of time on it.

Experiment Overview

To test our hypothesis, we planned to run an experiment for three days. We chose three days because we felt that was long enough to capture an effect, but not too long where attrition might occur. Our process began with recruitment to assess participants' interest and to collect contact information. After recruitment, our team randomly assigned half the participants to treatment and control, where the treatment group was asked to turn their phone screen to greyscale for the next three days. The control group was not asked to change anything of their phone settings. We checked back in four days later to collect screen time data from the last three days.

The experiment itself was supported through Qualtrics surveys: an initial survey and a final survey. Our initial survey asked participants to screenshot their screen time report from last week, given that they already had the screen time report feature already turned on. If not, we asked them to turn it on for the duration of our experiment. Those in the treatment group were additionally asked and given directions to turn the greyscale color filter on for the next three days. We also collected several other covariates including age group, phone type, and country. Our final survey asked participants to screenshot their daily screen time report from the last three days.

The reality of our project timeline was slightly different than initially planned. When we checked back in with our participants, only roughly 15 people had participated in the experiment at all. Because we felt like it was crucial to collect more data, we decided to run the experiment again, asking those who have not participated to now participate in the second round of the experiment. We included a covariate to indicate the study group the individuals were in.

Project Timeline

Initial Plan

Start Recruitment	**Start Experiment**	**Participants gather data for 3 days**	**End Experiment**
March 11	March 21	March 22-24	March 25

Actual Plan

Start Recruitment	**Start Experiment**	**Participants gather data for 3 days**	**End Experiment & Start Experiment (second group)**	**Participants (second group) gather data for 3 days**	**End Experiment (second group)**
March 11	March 21	March 22-24	March 25	March 26-28	March 29

Enrollment & Recruitment Process

Because our experiment was planned to be run over time, we wanted to collect contact information so that we could follow up and keep track with our participants. We sent out an interest form through Google Forms, asking people to provide their email address if they were interested in participating. Our recruitment pool consisted of our personal network, including friends, family, and MIDS. As an incentive, we noted that we would randomly select 10 participants to win $50 Amazon gift cards. In total, 108 people filled out the interest form.

Randomization

After gathering contact information of people who would be interested in participating, we randomly shuffled the list and assigned one half of the randomly-shuffled list to the treatment group and the other half to the control group.

Tools - Qualtrics

We used Qualtrics, a survey platform, to send out initial and final surveys to both the treatment and control groups. We chose Qualtrics as it allowed us to add logic to questions. In other words, it allowed us to show participants different questions given a specific answer. For example, given that a user had either an IPhone or an Android, the survey would be tailored to give them instructions to turn on screen time report, to screenshot, and to turn on the grayscale filter specific to their phone.

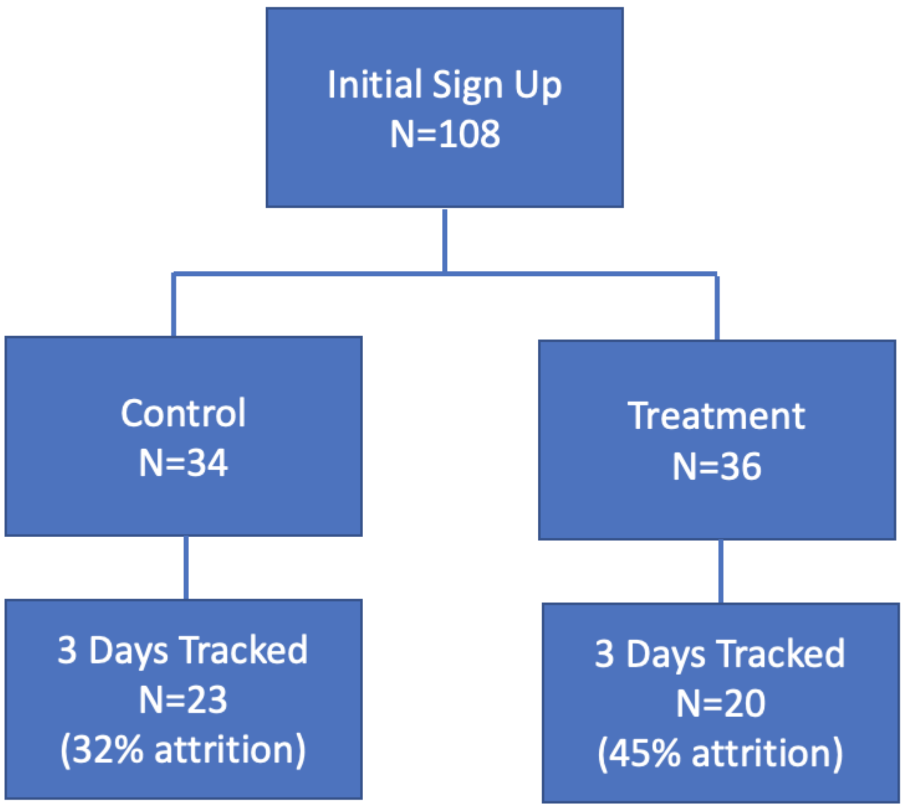
Data Completeness

As seen in the flow diagram below, we started the experiment by recruiting 108 people using a variety of methods, including a Facebook post, messages to Ischool students, and Twitter. We define non-compliance as signing up for the study, being placed in the treatment group, and not changing one's phone to grayscale mode or going back and forth during the treatment time. There were likely several instances of non-compliance but unfortunately we did not have an effective way to measure it. To minimize its effect, we asked participants to indicate if they were able to change their device to black and white. 100% of participants who completed the initial survey and the final survey indicated that they were able to change their settings to black and white.

After the initial recruitment effort we assigned 50% of the group to treatment and another 50% to the control group. However, we lost a sizable percentage of people who initially indicated they were interested in participating. Of people assigned to the control group, 37% percent did not respond to the survey invitation. In the treatment group, 33% percent did not respond to the survey invitation.

The largest issue we had with missing data was with attrition. As seen in the flow chart, in the control group 32% of participants who participated in the initial control group survey did not fill out the final survey. In the treatment group, the problem of attrition was even greater. 45% of individuals who participated in the initial treatment group survey did not respond in the final survey.

Treatment vs Control Flow Chart



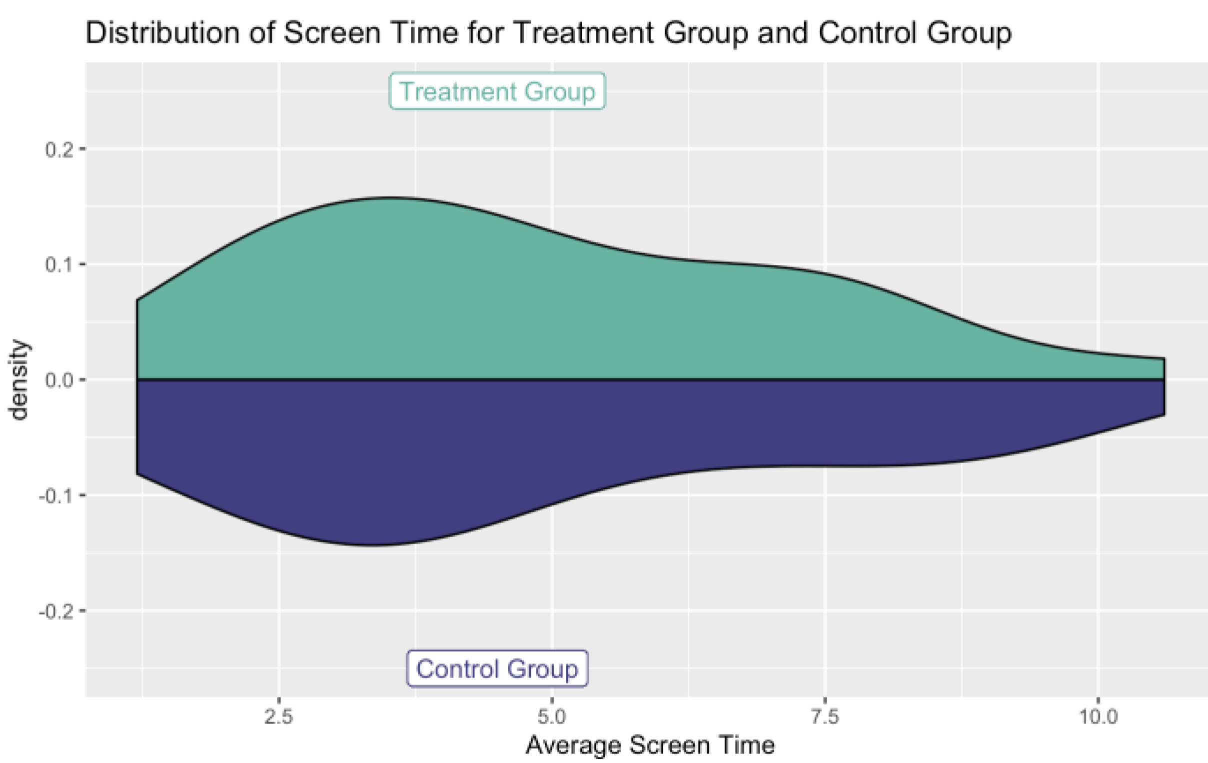
Observations and Outcome Measures

Our primary outcome measures of interest were the estimate for the average treatment effect (ATE), the Intent to Treat Effect (ITT), and the Complier Average Causal Effect (CACE). A quick review: The Average Treatment Effect (ATE) measures the effect of treatment on a randomly selected person. Intent-to-Treat (ITT) measures the effect of being made eligible for treatment, regardless of the fraction of the treatment group that's actually treated. The Complier Average

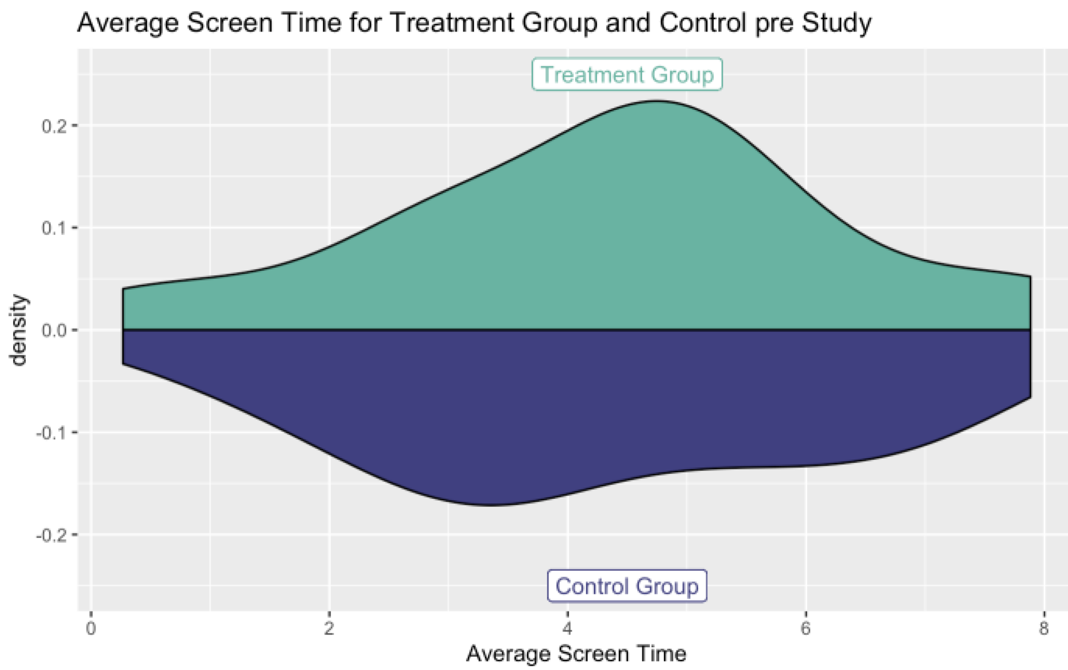
Causal Effect (CACE) is the average treatment effect of compliers. It can also be defined as $\frac{ITT}{ITT_d}$, where ITT_d is the difference between the proportion of subjects who are treated in the event that they are assigned to the treatment group and the proportion who would be treated even if they had been assigned to the control group.

Results

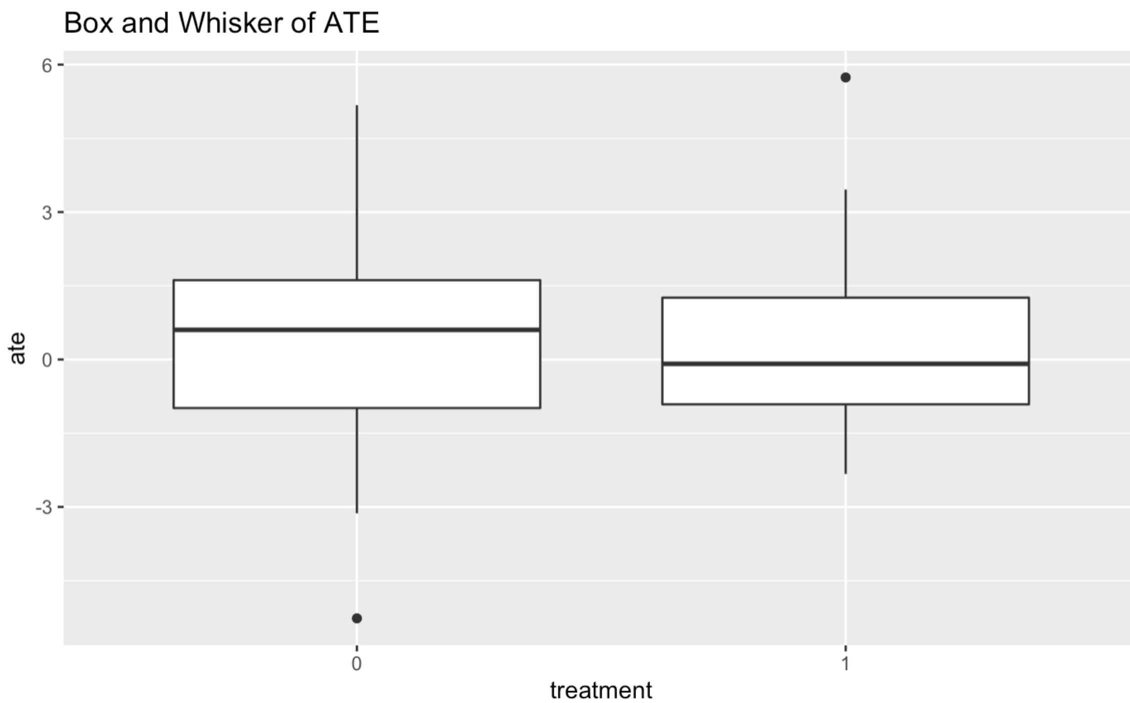
In our analyses, we wanted to see if changing one's screen color to grayscale mode would have a causal effect on screen time. To see the impact of changing a device to grayscale, we had individuals upload their screen time report for the three days they participated in the study. More explicitly, individuals in the control group uploaded three photos of their screen time without making any changes to their screen whereas the treatment group uploaded three photos of their screen time having adjusted their screen to greyscale. For our outcome variable, we decided to use the average screen time over the three days individuals were in the experiment. For the treatment group, the average screen time over the three days was 4 hours and 51 minutes with a standard deviation of 1 hour and 40 minutes. For the control group, the average screen time was 4 hours and 41 minutes with a standard deviation of 1 hour and 51 minutes. We also collected the prior weeks average screen time which was 4 hours and 21 minutes for the treatment group and 4 hours and 19 minutes for the control group. Below is a graphic showing the distribution of the treatment group and the control group.



The distribution of screen time in the experiment has substantial variability and a right skewed distribution. The following distribution shows the distribution of average screen time for the week before the experiment. Not surprisingly, the distribution for the treatment group shifted more than the control group. This likely can be attributed to the treatment effect even though we did not see any significant effects.



In this Box and Whisker plot below, we look at the average treatment effect for the control group (0) and treatment group (1) separately to get a better comparison of the difference in effect of the two groups.



To test to see if the differences in the mean between the treatment and control group are significant we decided to use a paired T-test to see if changing one's screen color to black and white has an impact on mean screentime. The difference in means proved to be highly insignificant with a confidence interval that includes zero and p-value of nearly one.

```
        Welch Two Sample t-test

data:  treatment_data[, screen_time_avg_final] and control_data[, screen_time_avg_final]
t = 0.051059, df = 29.598, p-value = 0.9596
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.675090  1.760945
sample estimates:
mean of x mean of y
 4.852594  4.809667
```

One major issue with this simplistic model is that it fails to account for covariates that likely have an impact on screen time, in particular, the average screen time for the week before treatment.

Models

In the initial survey we decided to ask for several covariates that could have an important impact on screen time. The first variable we decided to collect was the average screen time for the week before participation in our study. We believed this variable would be important because prior usage should be predictive of future usage. Some other covariates we collected were cellphone type (Android vs IPhone), nationality, age group, and a variable denoting if they participated in the first study cohort or the second study cohort. To analyze the relative importance of these covariates we decided to use an Anova F Test to see which model fit the data the best.

Ignoring Attrition

ATE

In our first models, we estimate the ATE, ITT, and CACE ignoring attrition.

$ATE = E[W_i | S_i = 1] - E[W_i | S_i = 0] \simeq Y_i = \beta_0 + \beta_1 (Binary\ Treatment)$

```
t test of coefficients:

                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            0.6393306  1.2643976  0.5056  0.61647
treatment             -0.0060024  0.6658367 -0.0090  0.99286
screen_time_avg_initial  0.5764728  0.2716128  2.1224  0.04140 *
country_binary         2.1090114  0.8574302  2.4597  0.01931 *
iphone                 1.1442797  0.8983256  1.2738  0.21164
first_group           -0.1124066  1.0406607 -0.1080  0.91464
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

This analysis regresses final average screen time (`screen_time_avg_final`) on treatment assignment (`treatment`), while controlling for initial average screen time (`screen_time_avg_initial`), country, and phone. Results suggest that the average treatment effect is -0.00033505 with a 95% confidence interval of +/- 1.272597 and that there is not a statistically significant treatment effect on final average screen time.

ITT

Intent to Treat Analysis

$ITT_i = \beta_0 + \beta_1 (Binary\ Treatment)$

```
t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 4.809667   0.700562  6.8654 4.286e-08 ***
treatment   0.042928   0.865303  0.0496    0.9607
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In this analysis, the final average screen time (`screen_time_avg_final`) is regressed on treatment assignment (`treatment`). Screen time is a continuous variable in units of hours and treatment assignment (`treatment`) is scored 1 if the subject was assigned to the treatment group and 0 otherwise. The ITT is 0.042928 with a 95% confidence interval of 1.730606. This result is not statistically significant.

CACE

Complier Average Causal Effect

$CACE_i = \frac{ITT_i}{ITT_d}$ Where $ITT_d = \beta_0 + \beta_1(Assignment)$

```
t test of coefficients:

                Estimate Std. Error t value  Pr(>|t|)
(Intercept)     4.809667   0.700562  6.8654 4.286e-08 ***
greyscale_binary 0.047016   0.947490  0.0496    0.9607
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In this instrumental variables regression model, the final average screen time (`screen_time_avg_final`) is regressed on actual treatment (`greyscale_binary`) using treatment assignment(`treatment`) as an instrument. The results suggest that turning on grayscale increased screen time usage among Compliers by 4.7% with a 95% confidence interval of 1.89498. However, the results suggest that this effect is not statistically insignificant.

ANOVA

In this section, we aim to compare models to understand if adding covariates improves the performance of the causal model.

```
Analysis of Variance Table

Model 1: screen_time_avg_final ~ treatment
Model 2: screen_time_avg_final ~ treatment + screen_time_avg_initial +
    country_binary + iphone + first_group
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     37 235.28
2     33 132.71  4    102.57 6.3764 0.0006459 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Analysis of Variance Table

Model 1: screen_time_avg_final ~ treatment + screen_time_avg_initial +
    country_binary + iphone + first_group
Model 2: screen_time_avg_final ~ treatment + screen_time_avg_initial +
    country_binary + first_group
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     33 132.71
2     34 140.32 -1   -7.6145 1.8934 0.1781
```

The increased information from the added covariates does improve the performance of the causal model. The F-test indicates that the inclusion of these covariates changes the standard errors of the estimates a statistically significant amount. Without the block fixed effects, the estimate has greater variance. The anova test has a p-value of 0.0001938 which suggests that there is a statistically significant difference in the different treatment effects for the added covariates. However, the addition of the study date for the two study groups does not appear to have any effect. We thought that the study date, in particular days of the week, would be more indicative of screen time.

Extreme Value Bounds

Due to the high levels of attrition, it is necessary to use extreme value bounds to see the possible range of treatment effects. To see the potential variability, we decided to impute the screen time for those who attrited with the minimum screen time value and the maximum screen time value. These two models were the most conservative approaches. We also fitted a model ignoring attrition, one where we imputed with the 75% for screen time, one where we imputed with the 25% percentile for screen time, and one where we imputed with the mean. Unlike the models where we imputed with the minimum and the maximum, the other models require a rather large assumption.

Extreme Value Bounds Analysis

	Dependent variable:					
	screen_time_avg_final	low_screen_time	high_screen_time	mean_screen_time	top_screen_time	bottom_screen_time
	(1)	(2)	(3)	(4)	(5)	(6)
treatment	-0.087	0.993*	-1.653**	-0.029	-0.524	0.477
		(0.574)	(0.674)	(1.046)	(0.436)	(0.541)
screen_time_avg_initial	0.598	0.468*	0.154	0.347	0.288**	0.407**
		(0.279)	(0.206)	(0.310)	(0.143)	(0.170)
as.factor(country)United States	-1.792	-1.815*	-1.165	-1.563	-1.442**	-1.688**
		(0.986)	(0.870)	(1.080)	(0.611)	(0.658)
iphone_binary	1.168	0.916	1.120	0.995	1.033	0.956
		(0.905)	(0.922)	(1.329)	(0.648)	(0.752)
Constant	2.398***	1.300	7.348***	3.638**	4.769***	2.481***
	(0.043)	(1.153)	(0.779)	(1.734)	(0.494)	(0.808)
Observations	28	49	49	49	49	49
R2	0.529	0.333	0.102	0.379	0.258	0.396
Adjusted R2	0.447	0.272	0.020	0.322	0.190	0.341
Note:						*p<0.1; **p<0.05; ***p<0.01

These results suggest that we are unable to precisely measure the ATE which shows substantial variability across the extreme value bounds. In order to properly assess the ATE, a study with methods to reduce attrition needs to be conducted.

Power

For our power calculation, we got 0.05 which is the probability that the test we set up will correctly reject the null hypothesis. We utilized the randomization inference process in order to repeatedly sample from our population and to product p-values. We will ultimately need a larger sample size in order to increase our probability of correctly rejecting the null hypothesis.

Conclusions

Even given the relative uncertainty due to attrition, there appears to be no recognizable causal effect of changing a phone screen to a grayscale on screen time. We thought that since a greyscale screen is less interesting than people would use their phone less. This may suggest that more interventions are needed to reduce screen time.

Limitations and Future Enhancements

In order to have a more concrete conclusion we would like to re-run the study in both a more controlled environment and with more participants. One potential issue that might have occurred but we could not measure was non-compliance. In a future study, it would be best to create an application that could track participants' compliance and also automatically collect the screentime information to reduce the issue of attrition. Another change that we would implement would be to use a much larger group of participants to reduce sampling variability.

Appendix

Image 1: Facebook Post Recruiting Participants

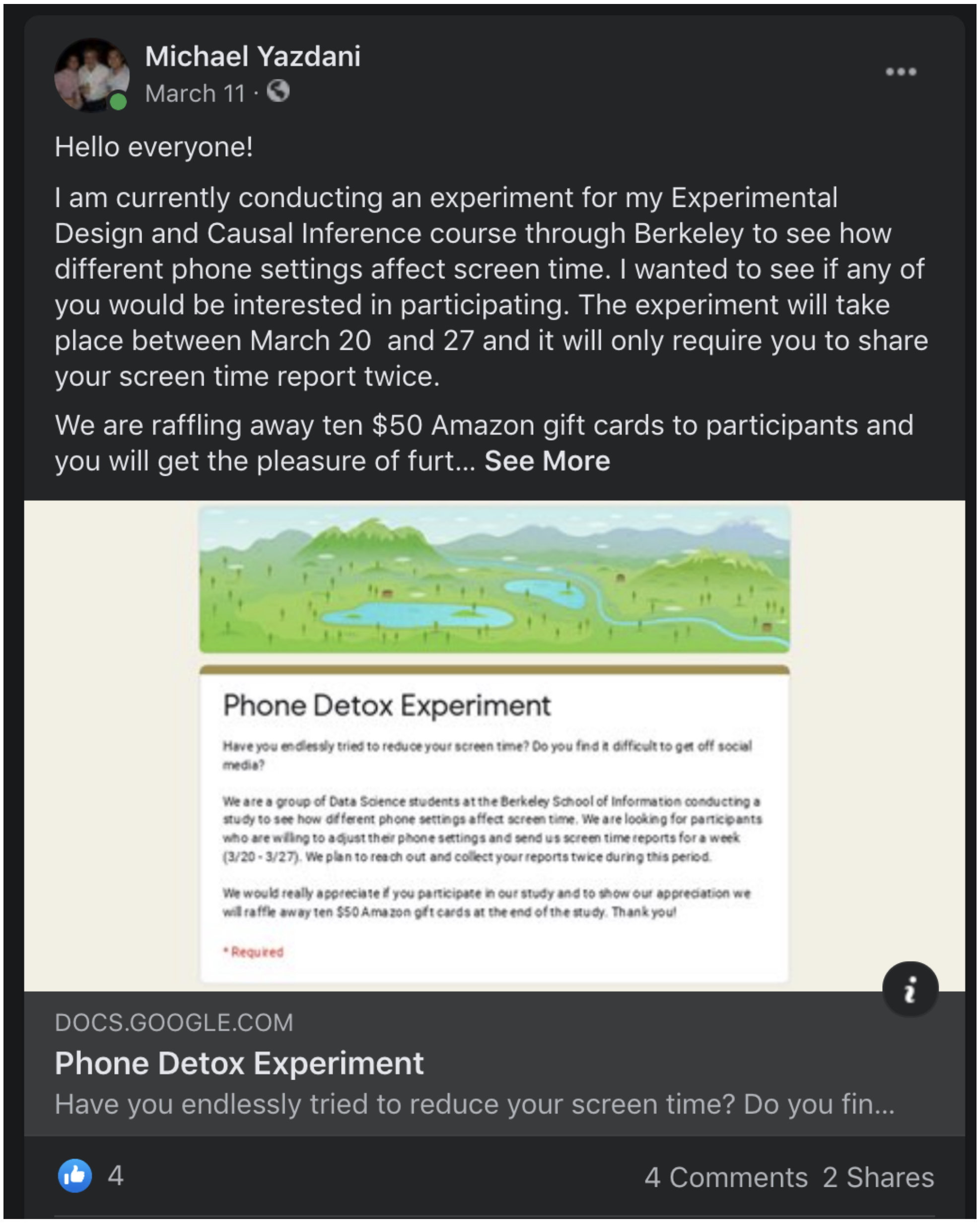


Image 2: Google Form Collecting Potential Participant's Email

Phone Detox Experiment

Have you endlessly tried to reduce your screen time? Do you find it difficult to get off social media?

We are a group of Data Science students at the Berkeley School of Information conducting a study to see how different phone settings affect screen time. We are looking for participants who are willing to adjust their phone settings and send us screen time reports for a week (3/20 - 3/27). We plan to reach out and collect your reports twice during this period.

We would really appreciate if you participate in our study and to show our appreciation we will raffle away ten $50 Amazon gift cards at the end of the study. Thank you!

First Name *

Short answer text

Last Name *

Short answer text

Email *

Short answer text

Image 3: Initial Survey

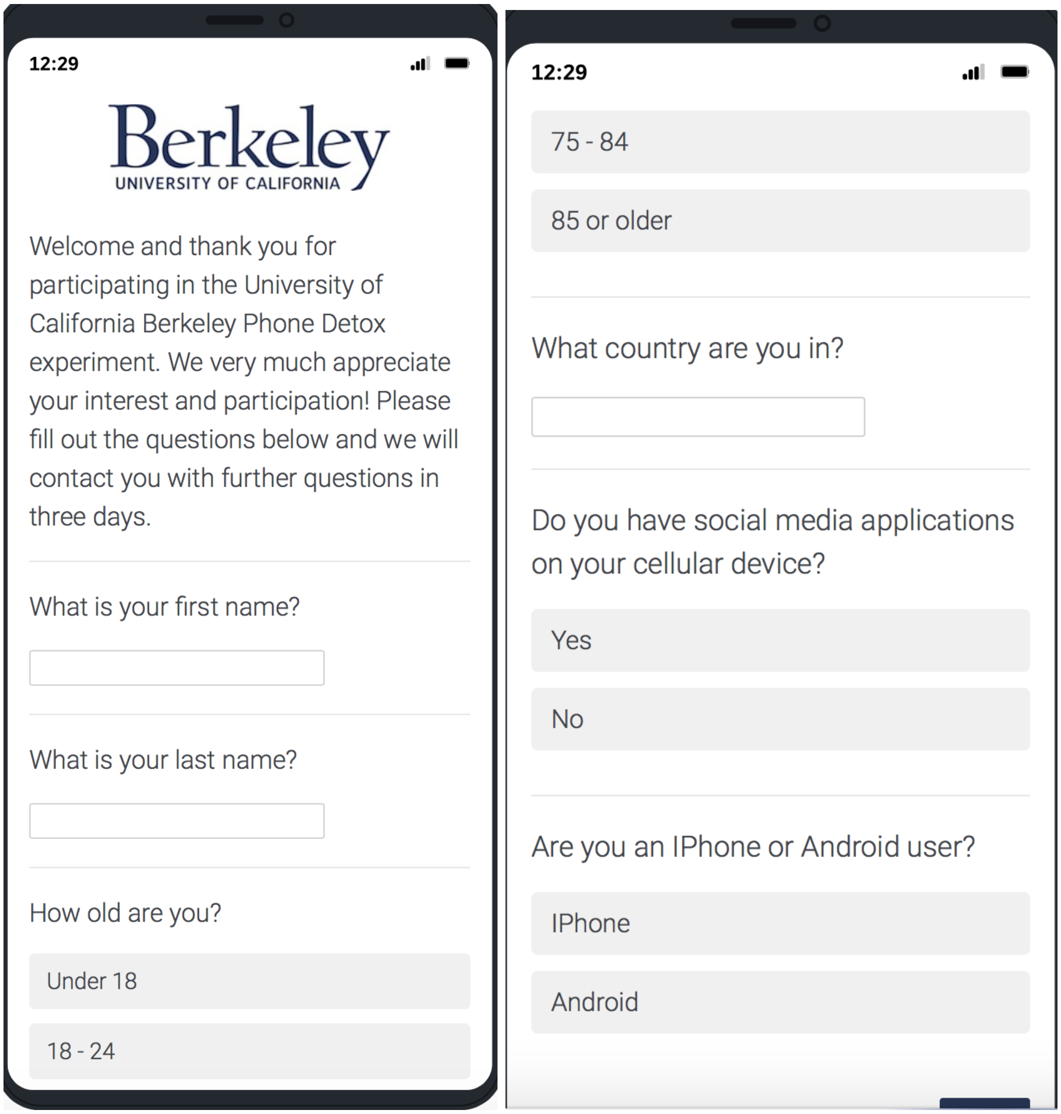


Image 4: Initial Survey

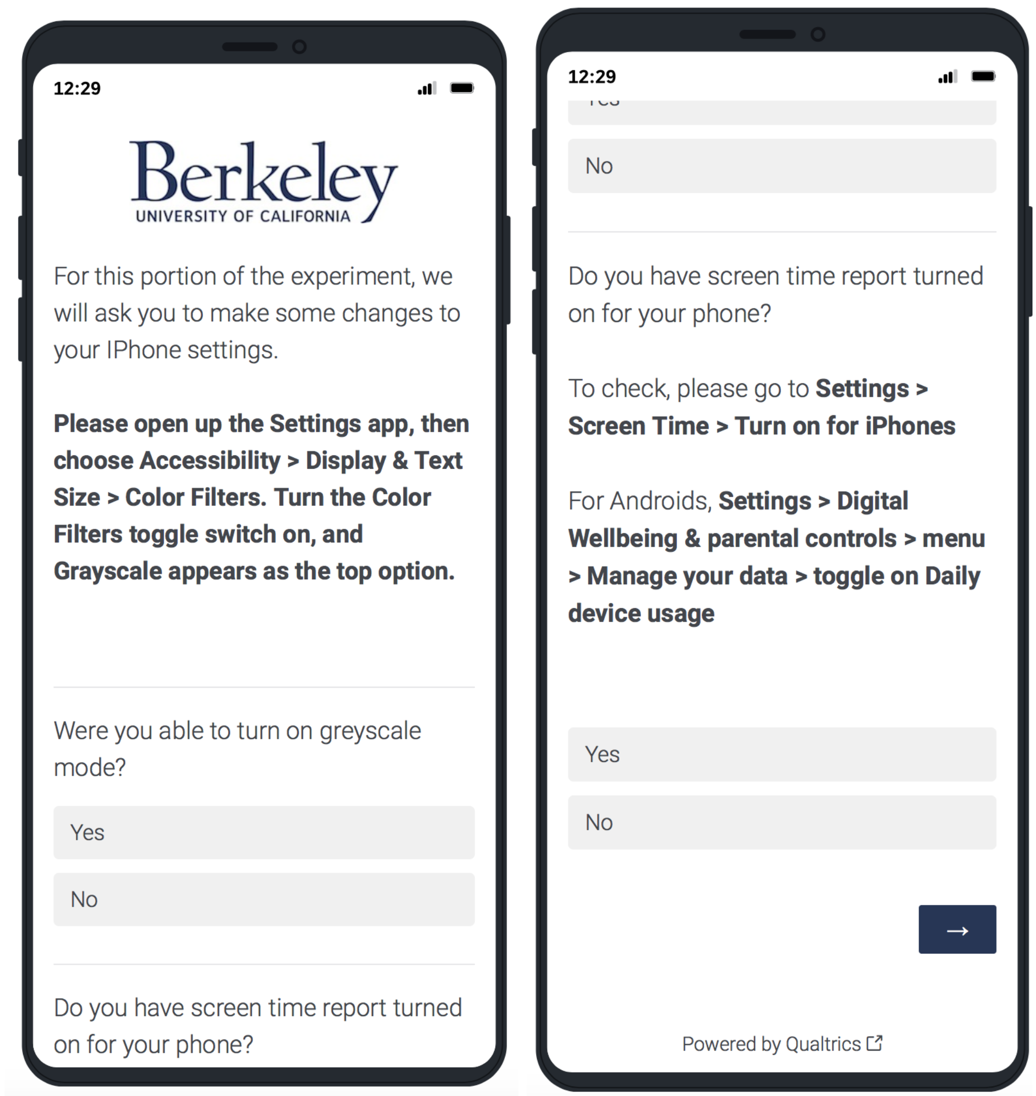


Image 5: Initial Survey

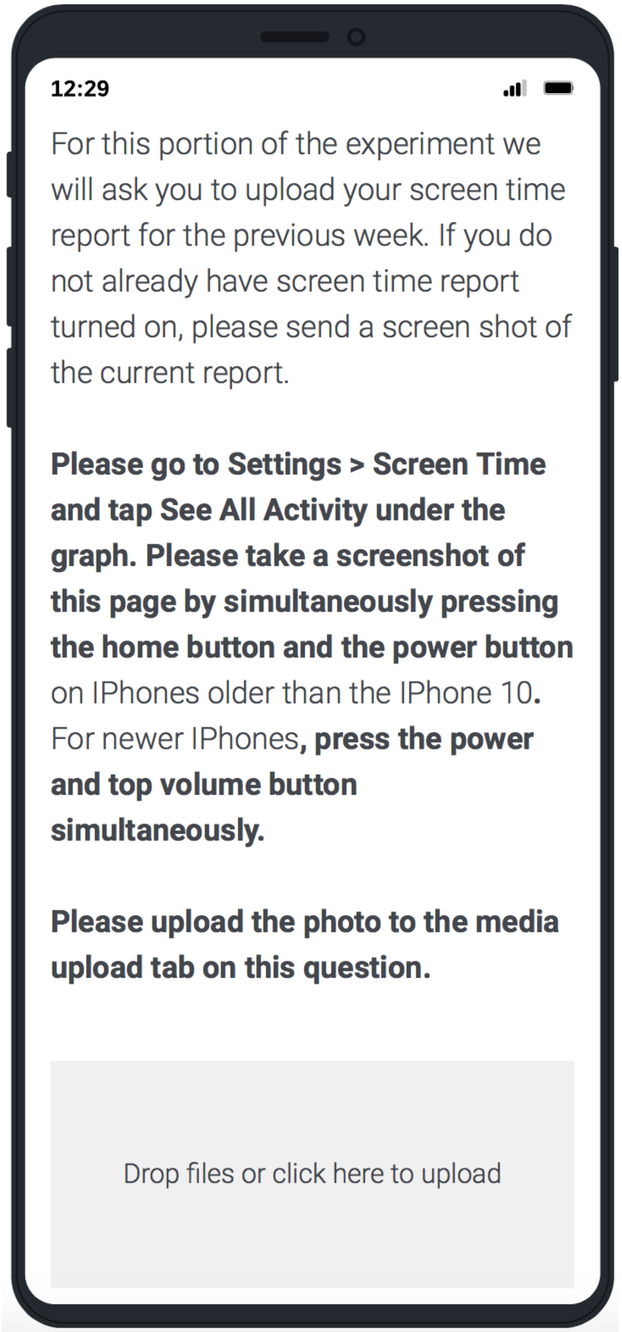
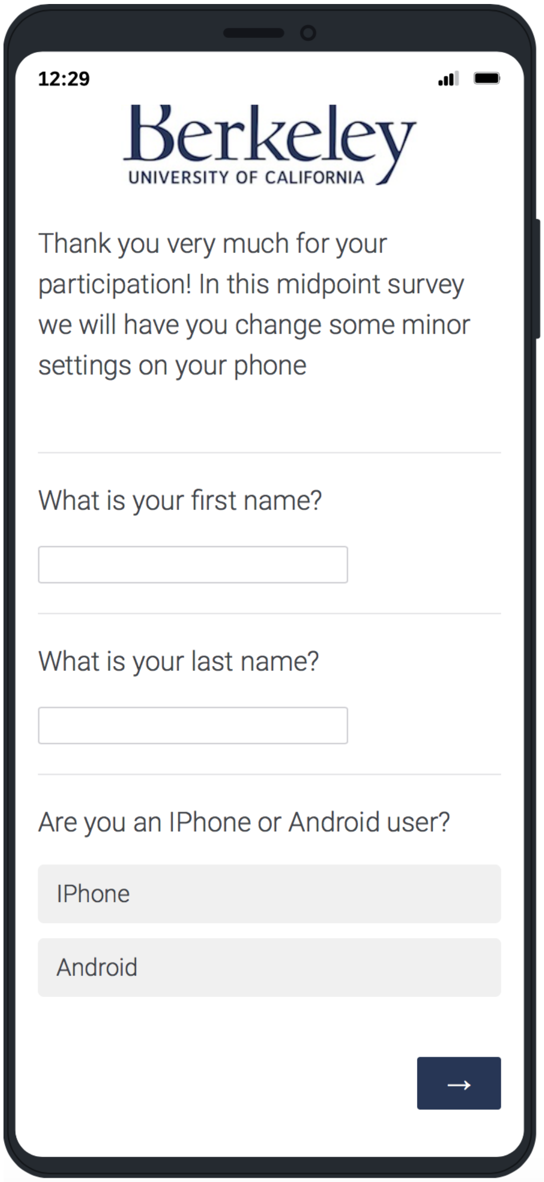
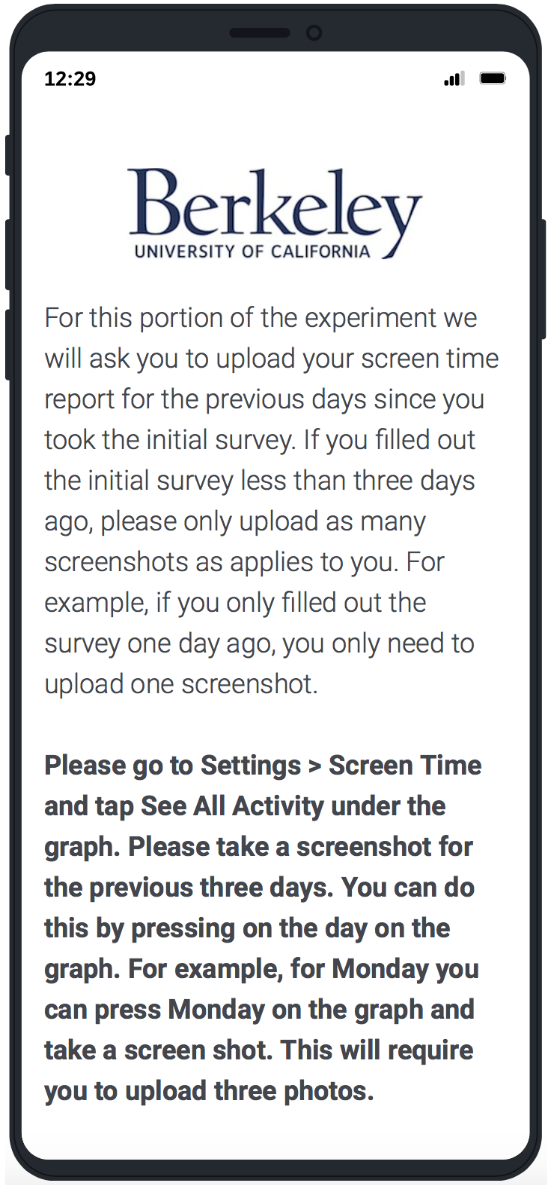
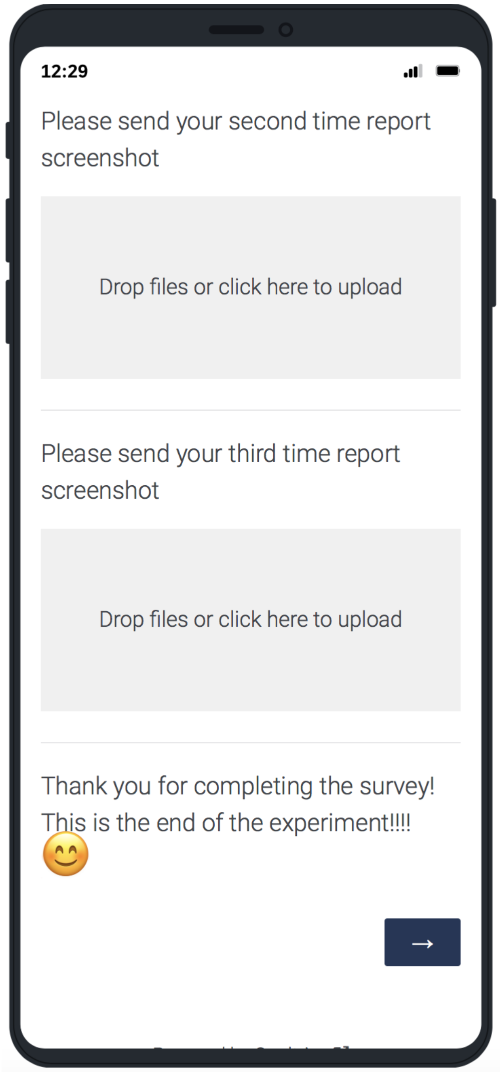

12:29
For this portion of the experiment we will ask you to upload your screen time report for the previous week. If you do not already have screen time report turned on, please send a screen shot of the current report.
Please go to Settings > Screen Time and tap See All Activity under the graph. Please take a screenshot of this page by simultaneously pressing the home button and the power button on IPhones older than the IPhone 10**.** For newer IPhones**, press the power and top volume button simultaneously.**
Please upload the photo to the media upload tab on this question.
Drop files or click here to upload

Image 6: Final Survey






12:29
Please send your second time report screenshot
Drop files or click here to upload
Please send your third time report screenshot
Drop files or click here to upload
Thank you for completing the survey! This is the end of the experiment!!!!
→