6 Appendix

6.1 Attrition

The attrition rates for the different stages of the experiment are as follows:

Table 4: Sample sizes and attrition at each stage of the experiment

	No. of participants allocated to group	Responses received	Attrition rate
Survey 2			
Control	314	259	18%
AppTreat	315	261	17%
Survey 3			
Control	259	236	9%
AppTreat	261	237	9%
Survey 4			
Control	236	232	2%
AppTreat	237	231	3%

Given that the main difference between the intervention the treatment and the control participants undertake is in the increased survey length for treatment participants, I would expect to see higher attrition in the treatment participants, if differential attrition surfaces. There does not appear to be differential attrition between the different participant groups.

6.2 Balance Table

Table 5: Balance table for app treatment

Variable	(1) Control Mean/SD	(2) Treatment Mean/SD	(3) T-test p-value (1)-(2)
Baseline Phone Usage (mins)	309.17	318.68	0.382
	(116.60)	(128.54)	
Baseline FB Usage (mins)	53.71	55.78	0.604
	(45.54)	(45.16)	
Baseline Insta Usage (mins)	29.43	28.10	0.601
	(29.63)	(22.77)	
Female	0.573	0.575	0.973
	(0.50)	(0.50)	
Age	23.56	23.30	0.462
	(4.41)	(4.42)	
College	0.506	0.514	0.843
	(0.50)	(0.50)	
White	0.704	0.708	0.910
	(0.46)	(0.46)	
Instagram	0.863	0.832	0.276
	(0.34)	(0.38)	
Observations	314	315	
F-test of joint significance (p-value)			0.8106

Notes: Columns (1) and (2) present the demographics for the Control and App Treatment groups. Column (3) details the p-values of the difference-in-means between the two groups.

6.3 Model and Proofs

Note that this is only one such model that could explain the effectiveness of commitment devices, and that I do not use the empirical data from my experiment to estimate the parameters of this model.

Consider an individual i with a horizon of infinite periods, $t = 0, 1, 2\ldots$ and so on. At $t = 0$, the agent is faced with the option of whether to adopt a soft commitment device (e.g., an app limit). At any time t, an individual i's lifetime utility can be represented by:

$$u_{it} = R_i(x_{it}) - A_{it}\mathbb{1}(x_{it} > y_i)Q_i(x_{it} - y_i) + \beta_i \sum_{n=1}^{\infty} \left[\delta_i^n [R_i(x_{i,t+n}) - A_{i,t+n}\mathbb{1}(x_{i,t+n} > y_i)Q_i(x_{i,t+n} - y_i) - \delta_i^{k-1} C_i(x_{i,t+n-1})]\right]$$

where $x_{i,t}$ is the amount of time that individual i chooses to spend on the given platform on an average day in period t, and y_i is the time limit that individual i sets for the platform, where $y_i = 0$ if they choose not to adopt app limits. Individuals are assumed to exhibit quasi-hyperbolic discounting, where β_i, $\delta_i < 1$ for all individuals i. $R_i(\cdot)$ is a function that denotes the returns to platform usage that is assumed to be concave and increasing in x. $C_i(\cdot)$, on the other hand, is a increasing function of x that characterizes the costs of platform usage. Returns to platform usage are assumed to be immediate (e.g., Fulfillment at having shared a post with friends, the benefits of consuming media, etc.) and costs are incurred at a later time period (e.g., Putting off work or school, leading to pejorative impacts on future long-term goal fulfillment, etc.) – specifically $k \in \mathbb{Z}^+$ time periods after the returns to social media are realized. Thus, costs are therefore further discounted by δ_i^k. $A_{i,t}$ is a binary variable denoting whether or not individual i had an app limit instituted at time t, where $A_{i,t} = 1$ if they did.

$Q_i(\cdot)$ is an increasing function of x that represents the additional cost that the individual incurs if they exceed the time limit specified by their adopted soft commitment device (e.g., Psychological cost of exceeding a pre-set commitment and ignoring a reminder, the actual cost of having to press "Ignore Limit", etc). I assume that participants are unaware of the app limit option so they choose $A_i = 0$ regardless of their preferences; my treatment makes all treated participants aware.

Now, consider an arbitrary time period $t' > 0$. Denote the decision-maker's optimal choice for time t' from the perspective of the $t = 0$ self by:

$$x^0_{i,t'} = \underset{x_{i,t'} \in \mathcal{X}}{\operatorname{argmax}}\, u_{i,0}$$

Note that as $u_{i,0}$ is additively separable across time, the individual's optimal decision at t' doesn't vary depending on their decision at $t' - 1$ and so on. Assume that if the agent sets an app limit for themselves at $t = 0$, they will set the magnitude of the limit at their optimal choice $x^{0^*}_{i,t'}$. Further denote the agent's optimal choice for time t' should they *not* adopt a time limit, and make the decision at t', by:

$$x^{t',A=0}_{i,t'} = \{\underset{x_{i,t'} \in \mathcal{X}}{\operatorname{argmax}}\, u_{i,t'} | A_{i,t'} = 0\}$$

Conversely, $x^{t',A=1}_{i,t'} = \{\underset{x_{i,t'} \in \mathcal{X}}{\operatorname{argmax}}\, u_{i,t'} | A_{i,t'} = 1\}$ denotes the choice at t' if the agent adopts a time limit.

Proposition 1: In absence of app limits, agents spend more time at time t' than they would've chosen to for that time period in $t = 0$; i.e. $x^0_{i,t'} < x^{t',A=0}_{i,t'}$. *Assumptions:* $\beta_i, \delta_i < 1$, as future outcomes are discounted; $R_i(\cdot)$ is concave and increasing in x due to diminishing marginal utility of social media use.

Proof: First, consider the optimisation problem that yields $x^0_{i,t'} = \underset{x_{i,t'} \in \mathcal{X}}{\operatorname{argmax}} u_{i,0}$. Since $u_{i,t}$ is additively separable across time (i.e., the individual's optimal decision at t' doesn't vary depending on their decision at $t'-1$, etc.), we have that:

$$\begin{aligned}
\frac{\partial u_{i,0}}{\partial x^0_{i,t'}} &= \frac{\partial(\beta_i \delta_i^{t'} R_i(x_{i,t'}) - \beta_i \delta_i^{k+t'} C_i(x_{i,t'}))}{\partial x_{i,t'}} = 0 \\
&\Rightarrow \frac{\partial(R_i(x_{i,t'}) - \delta_i^k C_i(x_{i,t'}))}{\partial x_{i,t'}} = 0 \\
&\Rightarrow \frac{\partial C_i(x_{i,t'})}{\partial x_{i,t'}} = \frac{1}{\delta_i^k} \times \frac{\partial R_i(x_{i,t'})}{\partial x_{i,t'}}
\end{aligned}$$

Similarly, consider the optimisation problem that yields $x^{t',A=0}_{i,t'} = \{\underset{x_{i,t'} \in \mathcal{X}}{\operatorname{argmax}} u_{i,t'} | A_{i,t'} = 0\}$. We have that:

$$\begin{aligned}
\frac{\partial u_{i,t'}}{\partial x^{t',A=0}_{i,t'}} &= \frac{\partial(R_i(x_{i,t'}) - \beta_i \delta_i^k C_i(x_{i,t'}))}{\partial x_{i,t'}} = 0 \\
&\Rightarrow \frac{\partial C_i(x_{i,t'})}{\partial x_{i,t'}} = \frac{1}{\beta_i \delta_i^k} \times \frac{\partial R_i(x_{i,t'})}{\partial x_{i,t'}}
\end{aligned}$$

Recall that $\beta_i < 1$ and $R_i''(x_{i,t}) \leq 0$. Thus, it must be that $x^0_{i,t'} < x^{t',A=0}_{i,t'}$.

Proposition 2: Agents spend less (or equal) time on the platform at time t' if they implement an app limit than if they did not; i.e., $x^{t',A=1}_{i,t'} \leq x^{t',A=0}_{i,t'}$. *Assumptions:* $\beta_i, \delta_i < 1$, as future outcomes are discounted; $R_i(\cdot)$ is concave and increasing in x due to diminishing marginal utility of social media use; $\frac{\partial Q_i(\cdot)}{\partial x} > 0$, as the additional costs of exceeding the time limit increase with the amount of time individuals spend on the platform (e.g. the more an individual exceeds their pre-set limit, the higher the psychological costs).

Proof: First, consider the optimisation problem that yields $x^{t',A=1}_{i,t'} = \{\underset{x_{i,t'} \in \mathcal{X}}{\operatorname{argmax}} u_{i,t'} | A_{i,t'} = 1\}$. We have that:

$$\begin{aligned}
\frac{\partial u_{i,t'}}{\partial x^{t',A=1}_{i,t'}} &= \frac{\partial(R_i(x_{i,t'}) - \mathbb{1}(x_{i,t'} > y_i) Q_i(x_{i,t'} - y_i) - \beta_i \delta_i^k C_i(x_{i,t'}))}{\partial x_{i,t'}} = 0 \\
&\Rightarrow \frac{\partial C_i(x_{i,t'})}{\partial x_{i,t'}} = \frac{1}{\beta_i \delta_i^k} \left[\frac{\partial R_i(x_{i,t'})}{\partial x_{i,t'}} - \frac{\partial(Q_i(x_{i,t'} - y_i)\mathbb{1}(x_{i,t'} > y_i))}{\partial x_{i,t'}} \right]
\end{aligned}$$

Comparing this to the expression of $\frac{\partial u_{i,t'}}{\partial x^{t',A=0}_{i,t'}}$ in the proof for Proposition 1, we know that since $Q_i'(x_{i,t} - y_i) > 0$ and $R_i''(x_{i,t}) \leq 0$, it must be the case that $x^{t',A=1}_{i,t'} \leq x^{t',A=0}_{i,t'}$, where $x^{t',A=1}_{i,t'} = x^{t',A=0}_{i,t'}$ if $x_{i,t'} > y_i$ and $x^{t',A=1}_{i,t'} < x^{t',A=0}_{i,t'}$ otherwise.

6.4 Sample Questions

Figure 7: Daily screen time input

Now, please enter the **daily screen time for each individual day**, found in the content of the same screenshot. To do so, please follow these instructions:

1. Hold down on the bar for the first day (leftmost bar) until the bar turns green
2. Record the number of minutes displayed for that day (e.g. for the example below, you would record 2h52m = **172m** for **Day 2**, as seen in the red box)
3. Repeat for all 7 days (from left to right)

Please input your answer in minutes (one hour = 60 minutes). If you do not have a time displayed for a certain day, please enter '0'.

Here's an example of what the page would look like:

	Number of minutes spent on phone
Day 1	
Day 2	
Day 3	
Day 4	
Day 5	
Day 6	
Day 7	

Estimate On average, how many minutes do you estimate you spend on your phone/Facebook/Instagram daily? (one hour = 60 mins)

Estimate (Peers) On average, how many minutes do you estimate your closest friends spend on their phones/ Facebook/Instagram daily?

Ideal and Predicted

Recall that:

- You spent a daily average of XX minutes on Facebook, in the last 7 days
- Last week, you expressed that you ideally wanted to spend XX minutes on Facebook daily
- Last week, you predicted that you would spend XX minutes on Facebook daily

In the next 7 days, how much time would you ideally want to spend on your phone/Facebook/Instagram per day, in minutes?

In the next 7 days, how much time do you predict you will actually spend on your phone per day, in minutes?

Tangney Brief Self-Control Scale

For the questions below, please indicate how much each of the following statements reflects how you typically are in general (not just in context of social media), using the scale provided. *[Scale options from 1 (Not at all) to 5 (Very Much)]*

1. I am good at resisting temptation.
2. I have a hard time breaking bad habits.
3. I am lazy.
4. I say inappropriate things.
5. I do certain things that are bad for me, if they are fun.
6. I refuse things that are bad for me.
7. I wish I had more self-discipline.
8. People would say that I have iron self- discipline.
9. Pleasure and fun sometimes keep me from getting work done.
10. I have trouble concentrating.
11. I am able to work effectively toward long-term goals.
12. Sometimes I can't stop myself from doing something, even if I know it's wrong.
13. I often act without thinking through all the alternatives.

Present-bias parameter elicitation

In the following questions, assume hypothetically that you have been offered a choice of receiving \$20 today. We will ask you how much you would be willing to accept at a later date, instead of \$20 today.

[Note that these questions are hypothetical and are unrelated to your actual survey payment for this study. As long as you complete all four surveys, you will receive your $20 payment within a couple days of Survey 4 completion.]

If we paid you in one/two/four month, what's the lowest amount that you would be willing to accept (instead of receiving $20 today)?

App limit encouragement nudge

Earlier, you said that you would ideally want to spend:

- XX minutes on your phone daily
- XX minutes on Facebook daily

iOS 12's "Screen Time" has a feature called "App Limits" that allows you to set time limits for apps. If you set an app limit, a reminder will pop up on your phone once you have reached your daily time limit. At that point, you can choose to ignore and dismiss the message: the limit will not prevent you from using our phone if you need to continue using it.

Would you be willing to set time limits on your phone, equivalent to the ideal time(s) you specified previously (or lower, if you wish)? (Note that agreeing to set time limits is not a requirement for you to be able to continue to participate in the study, though we highly encourage you to do so. Remember that setting a time limit is not binding - you can always choose to ignore the limit later.)

☐ Yes, and I currently do not have time limits set on my phone

☐ I already have time limits set on my phone

☐ No

(If participant already had time limits set on your phone) If you already have time limits set on your phone, kindly detail (i) which apps you have time limits set for, and (ii) the amount of minutes each of the limits are set at: If you do not want to set a time limit, kindly detail your reasons for not doing so.

(If participant chose not to set a limit) If you do not want to set a time limit, kindly detail your reasons for not doing so.

(If participant agreed to setting limits) First, let's set an app limit for your phone time usage. Please go to Settings >> Screen Time >> App Limits >> Add Limit.

Select "All Apps Categories", and then click "Add" on the top right corner of the page. Please set the limit to XX minutes or less, as you previously stated as your ideal phone time usage.

Once you have set your limit, please enter the number of minutes you set as your "All Apps Categories" limit.

Please upload a screenshot of your "App Limits" page, displaying all the limits you have set.

6.5 Descriptive Time Statistics

Table 6a: Average daily time usage data, phone

	n	Mean	Std Error	[95% Confidence Int.]	
Baseline					
Estimated, self	629	245.11	5.33	234.65	255.58
Estimated, peers	629	244.24	5.20	234.03	254.46
Actual (Week 0)	507	319.62	5.73	308.37	330.87
Week 1					
Ideal (Week 1)	629	159.82	3.79	152.38	167.26
Predicted (Week 1)	629	260.72	5.67	249.58	271.86
Actual (Week 1)	510	313.92	5.43	303.25	324.60
Week 2					
Ideal (Week 2)	518	189.88	4.53	180.97	198.78
Predicted (Week 2)	518	261.08	5.85	249.59	272.58
Actual (Week 2)	470	304.33	5.52	293.48	315.18
Week 6					
Ideal (Week 6)	473	202.51	4.75	193.18	211.84
Predicted (Week 4)	473	258.66	5.35	248.16	269.17
Actual (Week 6)	451	307.18	5.65	296.08	318.29

Table 6b: Difference-in-means results, phone

	n	Mean	Std Err	Std Dev	[95% Conf Int]		p-value, $\mu_a - \mu_b > 0$
Misperceptions							
Actual - Estd (Baseline)	507	74.59	5.83	128.94	63.13	86.05	0.000***
Actual - Pred (Week 1)	510	55.67	4.94	111.54	45.97	65.38	0.000***
Actual - Pred (Week 2)	469	45.26	5.14	111.22	35.17	55.36	0.000***
Actual - Pred (Week 6)	451	48.81	4.10	87.05	40.75	56.86	0.000***
Self-Control							
Actual - Ideal (Week 1)	510	154.07	5.26	118.86	143.73	164.41	0.000***
Actual - Ideal (Week 2)	469	116.46	4.58	99.26	107.45	125.46	0.000***
Actual - Ideal (Week 6)	451	104.52	4.77	101.37	95.14	113.90	0.000***

Notes: The former table presents phone time statistics (both self-reports and direct-measurements) across all four surveys. The latter table presents difference-in-means results in phone times corresponding to misperceptions (difference between actual and estimated/predicted times) and self-control issues (difference between actual and self-reported ideal times for that week).

Table 7a: Average daily time usage data, Facebook

	n	Mean	Std Error	[95% Confidence Int.]	
Baseline					
Estimated, self	629	80.36	3.09	74.28	86.43
Estimated, peers	629	85.45	3.65	78.27	92.63
Actual (Week 0)	509	57.54	2.20	53.23	61.86
Week 1					
Ideal (Week 1)	629	41.06	1.91	37.31	44.81
Predicted (Week 1)	629	71.93	3.79	64.47	79.38
Actual (Week 1)	516	54.74	2.00	50.82	58.66
Week 2					
Ideal (Week 2)	518	40.98	2.19	36.68	45.28
Predicted (Week 2)	518	60.78	3.60	53.71	67.85
Actual (Week 2)	469	48.96	1.86	45.30	52.62
Week 6					
Ideal (Week 6)	473	40.28	2.04	36.28	44.28
Predicted (Week 6)	473	53.18	2.66	47.95	58.41
Actual (Week 6)	457	48.00	1.96	44.15	51.85

Table 7b: Difference-in-means results, Facebook

	n	Mean	Std Err	Std Dev	[95% Conf Int]		p-value $\mu_a - \mu_b > 0$
Misperceptions							
Actual - Estd (Baseline)	509	-20.91	3.02	68.22	-26.85	-14.97	1.000
Actual - Pred (Week 1)	516	-11.70	3.18	72.30	-17.96	-5.45	1.000
Actual - Pred (Week 2)	468	-9.01	2.72	58.84	-14.36	-3.67	0.999
Actual - Pred (Week 6)	457	-4.31	2.28	48.73	-8.79	0.174	0.970
Self-Control							
Actual - Ideal (Week 1)	516	17.25	1.98	45.04	13.35	21.14	0.000***
Actual - Ideal (Week 2)	468	10.30	1.91	41.35	6.54	14.05	0.000***
Actual - Ideal (Week 6)	457	8.21	2.01	42.87	4.27	12.15	0.000***

Notes: The former table presents Facebook time statistics (both self-reports and direct-measurements) across all four surveys. The latter table presents difference-in-means results in Facebook times corresponding to misperceptions (difference between actual and estimated/predicted times) and self-control issues (difference between actual and self-reported times for that week).

Table 8a: Average daily time usage data, Instagram

	n	Mean	Std Error	[95% Conf Int]	
Baseline					
Estimated, self	533	61.15	2.96	55.34	66.96
Estimated, peers	533	87.64	3.79	80.19	95.09
Actual (Week 0)	426	32.02	1.44	29.18	34.85
Week 1					
Ideal (Week 1)	533	30.55	1.69	27.23	33.88
Predicted (Week 1)	533	49.10	3.16	42.89	55.31
Actual (Week 1)	436	28.79	1.27	26.29	31.28
Week 2					
Ideal (Week 2)	438	26.24	1.95	22.42	30.07
Predicted (Week 2)	438	36.47	2.91	30.76	42.18
Actual (Week 2)	393	27.45	1.21	25.08	29.83
Week 6					
Ideal (Week 6)	398	25.07	1.53	22.05	28.08
Predicted (Week 6)	398	33.61	2.21	29.27	37.96
Actual (Week 6)	386	27.48	1.29	24.95	30.01

Table 8b: Difference-in-means results, Instagram

	n	Mean	Std Err	Std Dev	[95% Conf Int]		p-value $\mu_a - \mu_b > 0$
Misperceptions							
Actual - Estd (Baseline)	426	-29.27	3.07	63.41	-35.31	-23.24	1.000
Actual - Pred (Week 1)	436	-19.18	3.30	68.98	-25.67	-12.69	1.000
Actual - Pred (Week 2)	393	-6.42	1.62	32.14	-9.60	-3.23	1.000
Actual - Pred (Week 6)	386	-6.12	1.98	38.98	-10.03	-2.23	0.999
Self-Control							
Actual - Ideal (Week 1)	436	-1.04	1.81	37.88	-4.60	2.53	0.716
Actual - Ideal (Week 2)	393	2.94	1.36	27.05	0.26	5.62	0.016**
Actual - Ideal (Week 6)	386	2.37	1.52	29.95	-0.63	5.37	0.061*

Notes: The former table presents Instagram time statistics (both self-reports and direct-measurements) across all four surveys. The latter table presents difference-in-means results in Instagram times corresponding to misperceptions (difference between actual and estimated/predicted times) and self-control issues (difference between actual and self-reported times for that week).

Table 9: Difference-in-means results, Baseline Actual - Week 1 Actual

	n	Mean	Std Err	Std Dev	[95% Conf Int]		p-value $\mu_a - \mu_b > 0$
Phone	426	4.013	3.27	67.48	-2.41	10.44	0.1102
Facebook	432	1.74	1.28	26.66	-0.78	4.26	0.088*
Instagram	364	2.19	0.823	15.80	0.57	3.82	0.0042***

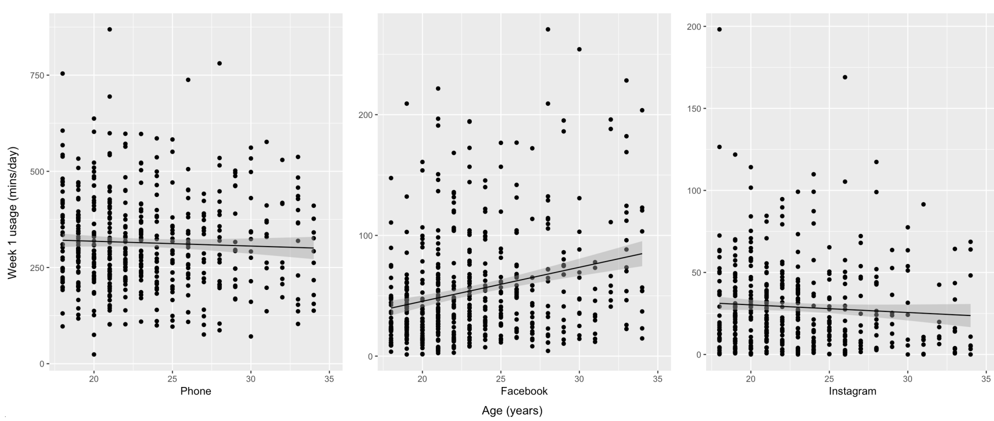
$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Notes: This table presents difference-in-means results between baseline "Screen Time" (collected in Survey 1) and Week 1 "Screen Time" on each of the phone, Facebook and Instagram platforms.

Usage by Demographic Characteristics

Given substantial variation in baseline usage of smartphones, Facebook, and Instagram, I break down the level of baseline usage of each platform by age, gender and education. As one would expect, Figure 8 shows a negative relationship between average daily phone/Instagram usage and age, though the relationship is not statistically significant. I omit participants aged 35 and over, as there are only two observations. Interestingly, there is a highly strong and significant positive relationship between age and the time respondents spend on Facebook. Given that the Facebook platform started in 2007 with the intention of targeting then-college students, the majority of highly active Facebook users would now be in their 30s. It is possible that the younger generation of teenagers and college-students are less reliant on Facebook as a mainstay of their social media (despite being active Facebook users), relying more heavily on other social media apps.

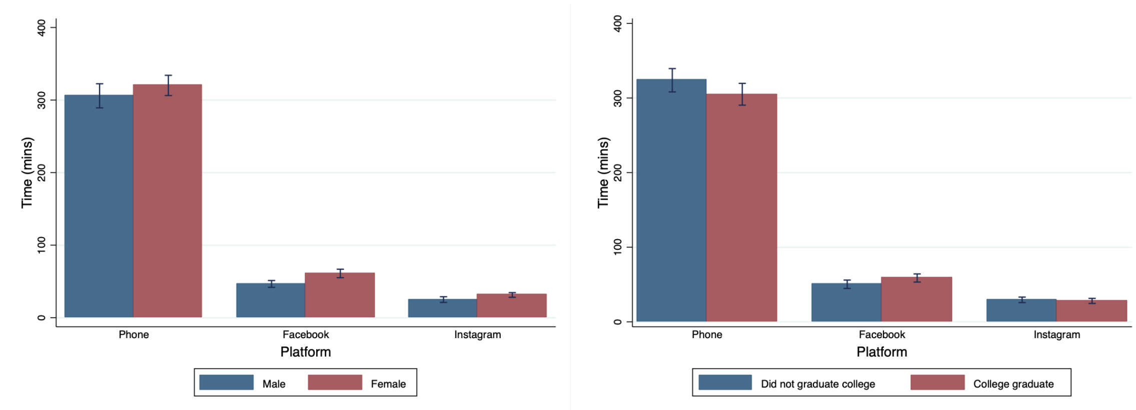
Figure 8: Relationship between age, and platform use



Notes: The black dots represent the binned scatterpoints, generated using binscatter with 20 equal-sized bins over age. The grey dots represent each individual observation with available Week 1 "Screen Time" data. The range of the y-axis varies by graph. Shaded area reflects 95 percent confidence intervals.

Figure 9 shows that, on average, females spend more time across all platforms than males. Furthermore, college graduates spend less time on their phones than non-college graduates, but spend more time on Facebook. This is consistent with the view that college-graduates are more active Facebook users – after all, they were the original target demographic of the platform.

Figure 9: Relationship between gender/education, and platform use



Notes: The y-axis represents pre-treatment Week 1 "Screen Time" actual usage data. The graph on the left presents average phone, Facebook, and Instagram use by gender; the graph on the right presents average phone and social media use by education level. Error bars represent 95% confidence intervals.

6.6 Treatment Effects, App Intervention

Table 10: First-stage between treatment assignment and actual treatment

	Short-term		*Long-term*	
	(1)	(2)	(3)	(4)
Variables	w/o Controls	w/ Controls	w/o Controls	w/ Controls
Apptreat	0.506***	0.505***	0.376***	0.370***
	(0.0299)	(0.0297)	(0.0308)	(0.0308)
Education		-0.0225		0.000213
		(0.0208)		(0.0215)
Female		0.0156		0.0357
		(0.0302)		(0.0312)
White		-0.0102		-0.000282
		(0.0332)		(0.0345)
Age		0.0107**		0.00795*
		(0.00422)		(0.00431)
Instagram		-0.0914**		-0.0976**
		(0.0418)		(0.0427)
Constant	0	-0.190*	0	-0.123
	(0.0212)	(0.0992)	(0.0217)	(0.101)
Observations	518	518	461	461
R-squared	0.357	0.377	0.244	0.265

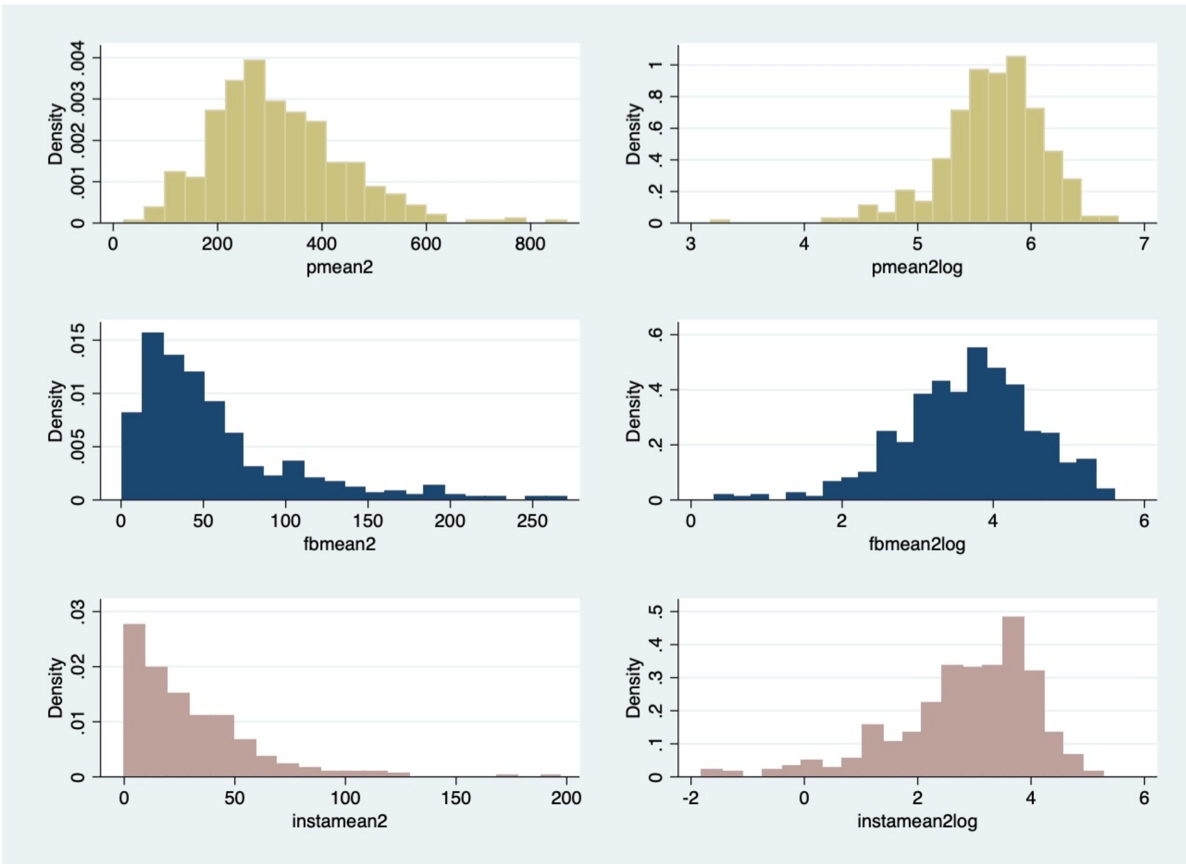
Robust standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Note: Columns 1, 3 present the first-stage relationship between treatment assignment and actual treatment without controls, whilst Columns 2, 4 present the relationship when controls are added.

Using *Time* as an outcome variable instead of $\log Time$

Due to the expectation that the app treatment would have a constant proportional rather than a constant level effect on individuals, I use $\log Time$ as the outcome variable in the main specification. For diagnostic purposes, I plot the histogram of both $Time$ and $\log Time$ and find that the distribution of $\log Time$ more approximately resembles a normal distribution.

Figure 10: Histogram of Week 1 platform use, absolute time (in minutes) and log time (in minutes)



Notes: Figures on the left side present the distributions of average Week 1 daily phone, Facebook and Instagram use, in minutes. Figures on the right present the distributions of average Week 1 daily phone, Facebook and Instagram use in log minutes.

Here, I present the intention-to-treat effects of the app limit assignment on the amount of time spent on social media (in minutes). Table 11a presents the main results of the impacts of app limits. Assignment into the app limit treatment has a significant negative impact on the amount of time participants spend on their phones and Facebook, reducing overall phone usage by 11 minutes and Facebook usage by 6 minutes respectively. It also reduces Instagram usage by roughly 2 minutes a day, although this is not significant, possibly because there were less observations for the Instagram platform.

Table 11a: Intention-to-treat effects of app limits

	Short-term			*Long-term*		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone	Facebook	Instagram	Phone	Facebook	Instagram
Apptreat	-11.14**	-5.635***	-2.007*	-3.908	-5.565***	-1.810
	(5.110)	(1.803)	(1.107)	(6.135)	(2.029)	(1.262)
Constant	319.5***	54.38***	28.31***	323.0***	55.20***	29.46***
	(3.127)	(1.100)	(0.677)	(3.092)	(1.082)	(0.692)
Observations	5,345	5,415	4,558	5,178	5,278	4,478
R-squared	0.004	0.006	0.008	0.004	0.006	0.005
Number of id	455	456	385	445	445	376

Notes: Columns (1) through (3) present the short-term intention-to-treat effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (over the course of five weeks, between Survey 2 and Survey 4). *Apptreat* is the binary variable indicating treatment assignment.

Table 11b presents the treatment-on-treated (ToT) effects of app limits on the amount of time spent on social media (in minutes). Overall, the estimated ToT effect of the app limit intervention is relatively large and significant – the app treatment significantly decreases phone usage by 21 minutes and Facebook usage by 11 minutes. The ToT effect on Instagram is a reduction of Instagram daily usage by 4 minutes, though the result is not significant.

Table 11b: Treatment-on-treated effects of app limits

	Short-term			*Long-term*		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone	Facebook	Instagram	Phone	Facebook	Instagram
Actual treatment	-20.82**	-10.59***	-4.069*	-9.730	-13.76***	-4.970
	(9.342)	(3.268)	(2.193)	(15.23)	(5.019)	(3.441)
Constant	319.5***	54.31***	28.31***	323.0***	55.20***	29.47***
	(3.125)	(1.095)	(0.677)	(3.096)	(1.084)	(0.696)
Observations	5,339	5,409	4,558	5,178	5,278	4,478
Number of id	454	455	385	445	445	376

Notes: Columns (1) through (3) present the short-term treatment-on-treated effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (over the course of five weeks, between Survey 2 and Survey 4).

Average time spent on platform daily, by app intervention treatment group

Figure 11a: Average time spent on smartphone (mins)

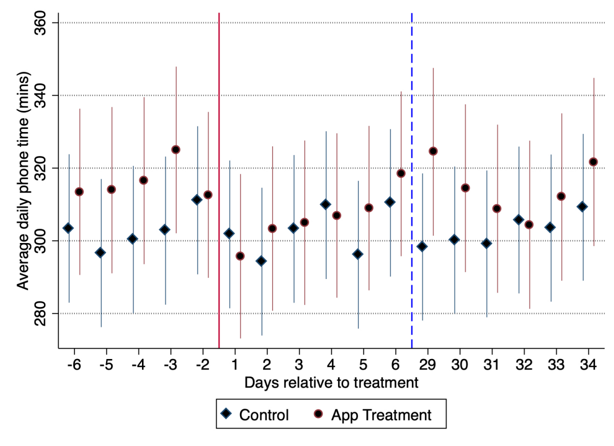


Figure 11b: Average time spent on Facebook (mins)

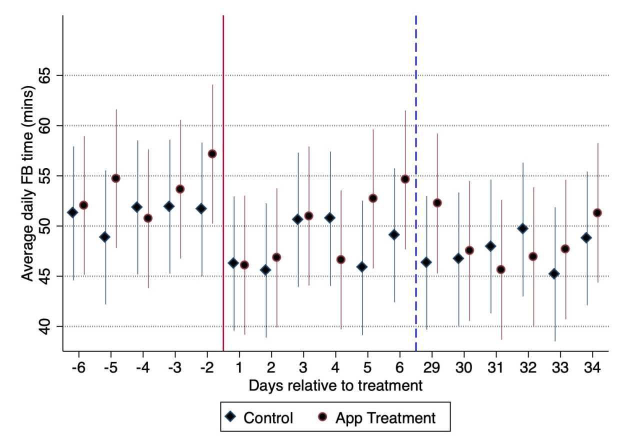
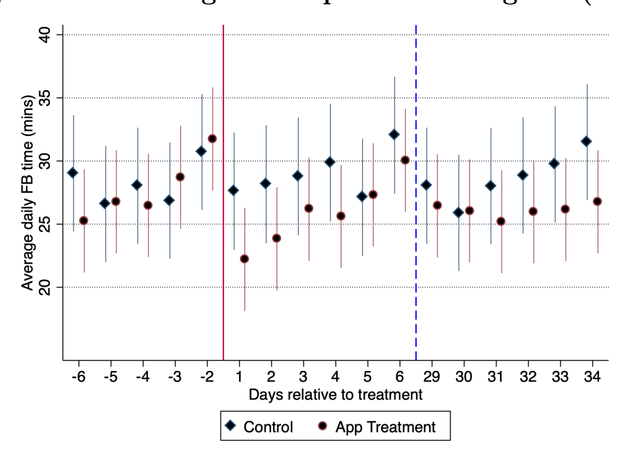


Figure 11c: Average time spent on Instagram (mins)



Notes: Red dots represent the mean time the app treatment group spent on the platform on any given day relative to the app treatment, whilst the blue dots represent the mean time the control group spent on the platform on any given day. Error bars represent 95 percent confidence intervals. The red line denotes the time of the app treatment, and the dotted blue line demarcates the privacy treatment.

6.7 Heterogeneous Treatment Effects

By Baseline Discrepancy Between Actual and Ideal Usage (Lack of Self-Control)

Figure 12a: Heterogeneous treatment effects of app limits on phone usage

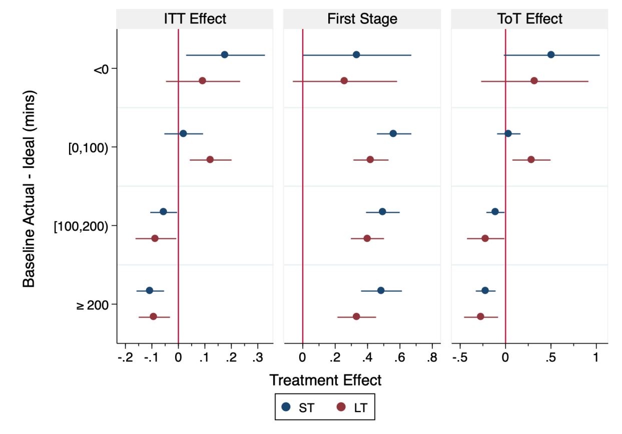


Figure 12b: Heterogeneous treatment effects of app limits on Facebook usage

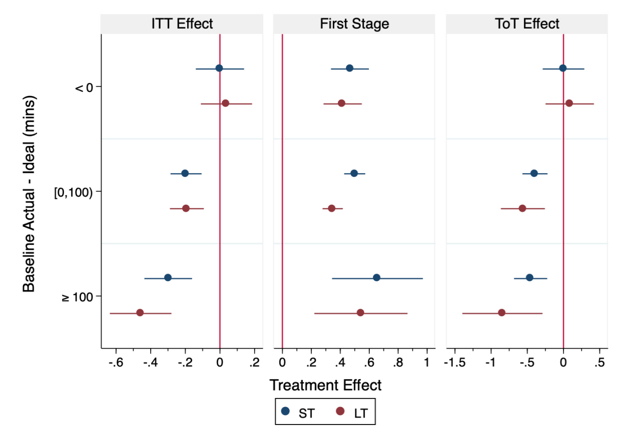
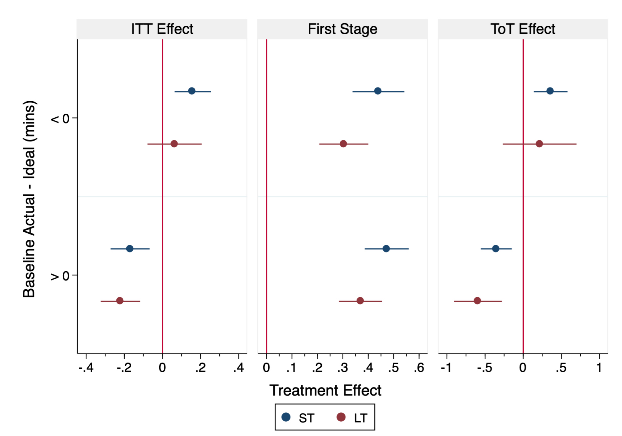


Figure 12c: Heterogeneous treatment effects of app limits on Instagram usage



Notes: This figure presents intention-to-treat, first-stage, and treatment-on-treated effects of the app limit intervention, by **baseline discrepancy between actual and ideal usage**. Short-term effects (duration of one week, between Survey 2 and 3) are in blue, whilst long-term effects (over the course of four weeks, between Survey 3 and 4) are in red. Error bars reflect 95 percent confidence intervals.

Table 12a: Regression results – Interaction of baseline lack of self-control (actual - ideal) with app treatment assignment

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone, ST	Phone, LT	FB, ST	FB, LT	Insta, ST	Insta, LT
Apptreat	0.0507*	0.0765**	-0.0985***	-0.0773*	-0.0439	-0.110**
	(0.0303)	(0.0337)	(0.0370)	(0.0453)	(0.0388)	(0.0451)
Apptreat * Actual-Ideal	-0.000522***	-0.000602***				
(Phone)	(0.000128)	(0.000146)				
Apptreat * Actual-Ideal			-0.00274***	-0.00277***		
(FB)			(0.000848)	(0.000889)		
Apptreat * Actual-Ideal					-0.00266**	-0.00223**
(Insta)					(0.00120)	(0.000930)
Constant	5.643***	5.665***	3.547***	3.566***	2.876***	2.945***
	(0.0106)	(0.0111)	(0.0219)	(0.0243)	(0.0263)	(0.0259)
Observations	5,173	5,127	5,195	5,170	4,116	4,073
R-squared	0.008	0.010	0.015	0.011	0.010	0.010
Number of id	437	437	442	442	370	370

Table 12b: Regression results – Interaction of baseline lack of self-control (actual - ideal) binary variable with app treatment assignment

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone, ST	Phone, LT	FB, ST	FB, LT	Insta, ST	Insta, LT
Apptreat	0.0263	0.0627*	-0.0317	0.00997	0.168***	0.0408
	(0.0292)	(0.0337)	(0.0480)	(0.0572)	(0.0499)	(0.0694)
Apptreat * Actual-Ideal	-0.107***	-0.151***				
(Phone, Above Median)	(0.0358)	(0.0421)				
Apptreat * Actual-Ideal			-0.228***	-0.273***		
(FB, Above Median)			(0.0747)	(0.0800)		
Apptreat * Actual-Ideal					-0.400***	-0.282***
(Insta, Above Median)					(0.0726)	(0.0891)
Constant	5.644***	5.661***	3.546***	3.565***	2.878***	2.945***
	(0.0106)	(0.0110)	(0.0219)	(0.0243)	(0.0262)	(0.0258)
Observations	5,205	5,175	5,213	5,188	4,128	4,085
R-squared	0.006	0.009	0.015	0.012	0.017	0.012
Number of id	444	445	445	445	372	372

Notes: The former table presents regression results for the interaction of the discrepancy between actual and ideal platform usage with app treatment. The latter table presents regression results for the interaction of a binary variable (=1 when the discrepancy between actual and ideal is above median) with app treatment assignment. Columns (1) through (3) present the short-term intention-to-treat effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (over the course of five weeks, between Survey 2 and Survey 4).

By Platform Usage

Figure 13a: Heterogeneous treatment effects of app limits on phone usage

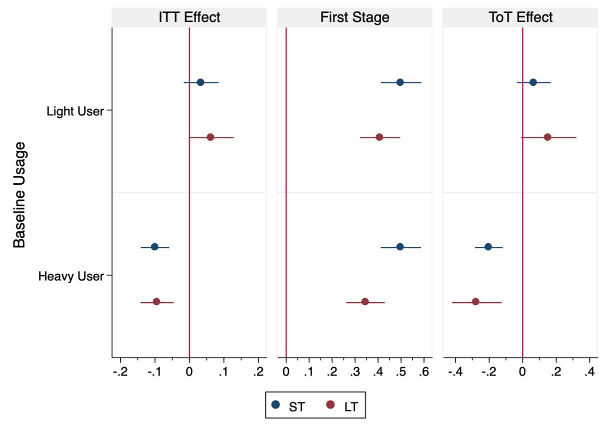


Figure 13b: Heterogeneous treatment effects of app limits on Facebook usage

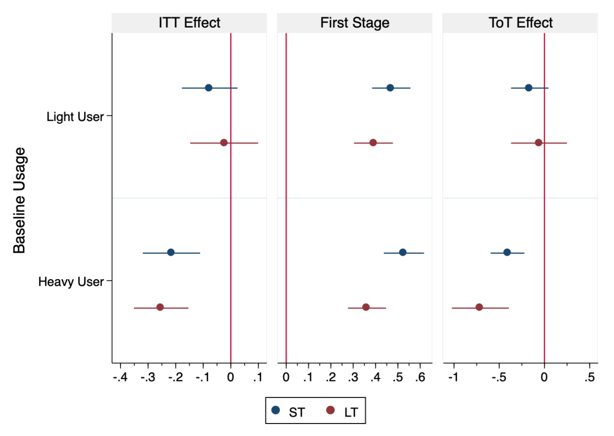
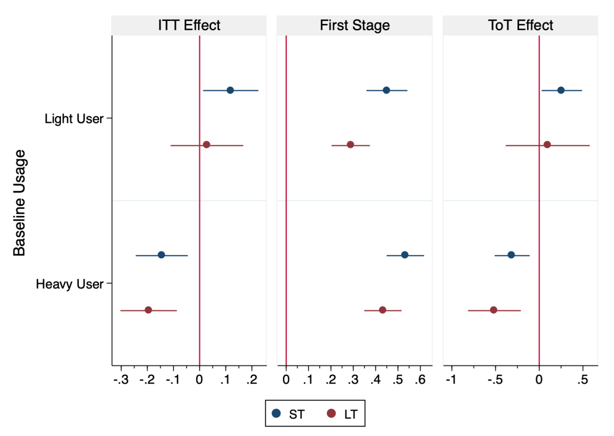


Figure 13c: Heterogeneous treatment effects of app limits on Instagram usage



Notes: This figure presents intention-to-treat, first-stage, and treatment-on-treated effects of the app limit intervention, by **Week 1 usage**. Short-term effects (duration of one week, between Survey 2 and 3) are in blue, whilst long-term effects (over the course of four weeks, between Survey 3 and 4) are in red. Error bars reflect 95 percent confidence intervals.

Table 13a: Regression results – Interaction of Week 1 platform usage with app treatment assignment

Variables	(1) Phone, ST	(2) Phone, LT	(3) FB, ST	(4) FB, LT	(5) Insta, ST	(6) Insta, LT
Apptreat	0.145***	0.188***	-0.0627	0.0272	0.194***	0.0173
	(0.0547)	(0.0623)	(0.0530)	(0.0626)	(0.0644)	(0.0764)
Apptreat * Phone Use	-0.000548***	-0.000640***				
	(0.000145)	(0.000163)				
Apptreat * FB Use			-0.00154**	-0.00275***		
			(0.000668)	(0.000715)		
Apptreat * Insta Use					-0.00769***	-0.00400**
					(0.00172)	(0.00161)
Constant	5.643***	5.665***	3.547***	3.566***	2.877***	2.945***
	(0.0106)	(0.0110)	(0.0220)	(0.0243)	(0.0264)	(0.0259)
Observations	5,173	5,127	5,195	5,170	4,116	4,073
R-squared	0.009	0.011	0.012	0.012	0.015	0.010
Number of id	437	437	442	442	370	370

Robust standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table 13b: Regression results – Interaction of Week 1 platform usage (below/above median) with app treatment assignment

Variables	(1) Phone, ST	(2) Phone, LT	(3) FB, ST	(4) FB, LT	(5) Insta, ST	(6) Insta, LT
Apptreat	0.0423	0.0694**	-0.0668	-0.00239	0.107*	0.00622
	(0.0281)	(0.0337)	(0.0517)	(0.0637)	(0.0550)	(0.0710)
Apptreat * Phone (Above Median)	-0.138***	-0.162***				
	(0.0351)	(0.0421)				
Apptreat * FB (Above Median)			-0.152**	-0.236***		
			(0.0754)	(0.0815)		
Apptreat * Insta (Above Median)					-0.261***	-0.202**
					(0.0765)	(0.0912)
Constant	5.644***	5.661***	3.546***	3.565***	2.878***	2.946***
	(0.0106)	(0.0110)	(0.0221)	(0.0243)	(0.0262)	(0.0259)
Observations	5,205	5,175	5,213	5,188	4,128	4,085
R-squared	0.008	0.010	0.013	0.011	0.011	0.010
Number of id	444	445	445	445	372	372

Robust standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Notes: The former table presents regression results for the interaction of Week 1 "Screen Time" with app treatment. The latter table presents regression results for the interaction of a binary variable (=1 when the Week 1 "Screen Time" is above median) with app treatment assignment. Columns (1) through (3) present the short-term intention-to-treat effects (i.e. the duration of one week, between Survey 2 and 3), and Columns (4) through (6) present the longer-term effects (duration of four weeks, between Survey 3 and 4).

By Self-Control Measures

Figure 14a: Heterogeneous treatment effects of app limits on phone usage

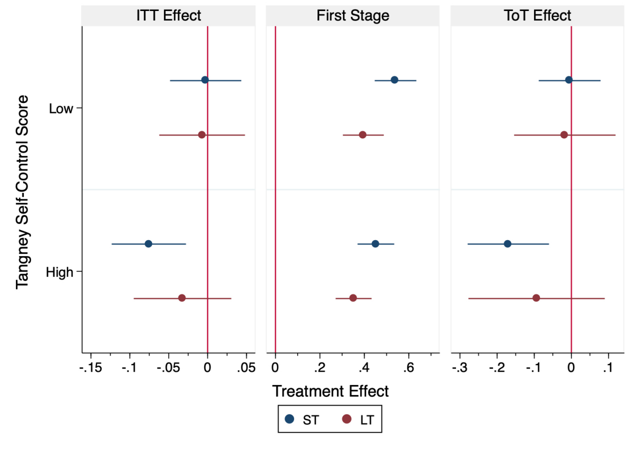


Figure 14b: Heterogeneous treatment effects of app limits on Facebook usage

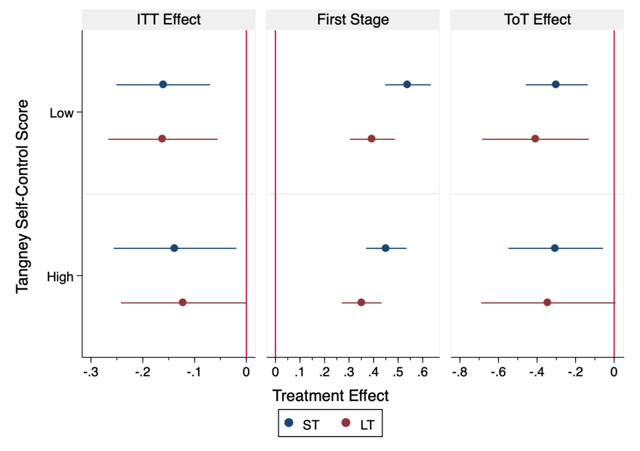
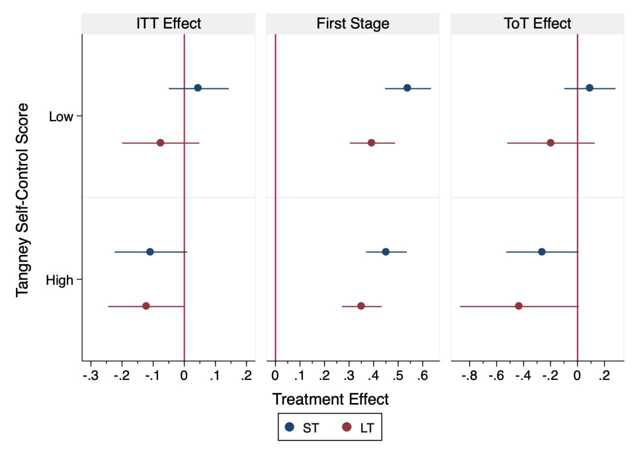


Figure 14c: Heterogeneous treatment effects of app limits on Instagram usage



Notes: This figure presents intention-to-treat, first-stage, and treatment-on-treated effects of the app limit intervention, by **Tangney self-control score**. Short-term effects (duration of one week, between Survey 2 and 3) are in blue, whilst long-term effects (over the course of four weeks, between Survey 3 and 4) are in red. Error bars reflect 95 percent confidence intervals.

Table 14a: Regression results – Interaction of Tangney self-control score with app treatment assignment

Variables	(1) Phone, ST	(2) Phone, LT	(3) FB, ST	(4) FB, LT	(5) Insta, ST	(6) Insta, LT
Apptreat	0.116	-0.0250	-0.382**	-0.160	0.417**	-0.00796
	(0.0796)	(0.102)	(0.186)	(0.193)	(0.164)	(0.202)
Apptreat * Self-Control Score	-0.0474*	0.00237	0.0745	0.0107	-0.145***	-0.0315
	(0.0253)	(0.0325)	(0.0604)	(0.0614)	(0.0521)	(0.0611)
Constant	5.644***	5.661***	3.546***	3.565***	2.878***	2.946***
	(0.0107)	(0.0112)	(0.0221)	(0.0245)	(0.0262)	(0.0258)
Observations	5,205	5,175	5,213	5,188	4,128	4,085
R-squared	0.004	0.003	0.012	0.008	0.010	0.008
Number of id	444	445	445	445	372	372

Robust standard errors in parentheses, $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$

Table 14b: Regression results – Interaction of binary Tangney self-control score (below/above median) with app treatment assignment

Variables	(1) Phone, ST	(2) Phone, LT	(3) FB, ST	(4) FB, LT	(5) Insta, ST	(6) Insta, LT
Apptreat	0.00295	-0.00633	-0.151***	-0.148***	0.0538	-0.0776
	(0.0233)	(0.0278)	(0.0466)	(0.0549)	(0.0497)	(0.0641)
Apptreat * Self-Control Score	-0.0801**	-0.0259	0.00564	0.0477	-0.203**	-0.0661
(Above Median)	(0.0355)	(0.0435)	(0.0791)	(0.0829)	(0.0798)	(0.0901)
Constant	5.643***	5.661***	3.546***	3.566***	2.878***	2.946***
	(0.0107)	(0.0112)	(0.0221)	(0.0245)	(0.0262)	(0.0259)
Observations	5,205	5,175	5,213	5,188	4,128	4,085
R-squared	0.004	0.003	0.011	0.008	0.010	0.008
Number of id	444	445	445	445	372	372

Robust standard errors in parentheses, $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$

Notes: The former table presents regression results for the interaction of Tangney's brief self-control score with app treatment. The latter table presents regression results for the interaction of a binary variable (=1 when the score is above median) with app treatment assignment. Columns (1) through (3) present the short-term intention-to-treat effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (duration of five weeks, between Survey 2 and Survey 4).

Figure 15a: Heterogeneous treatment effects of app limits on phone usage

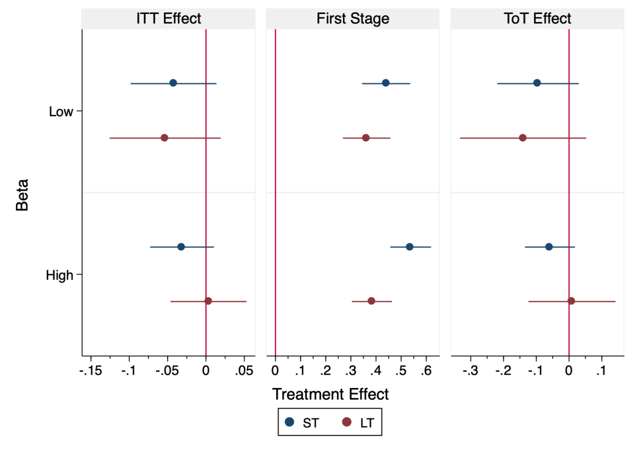


Figure 15b: Heterogeneous treatment effects of app limits on Facebook usage

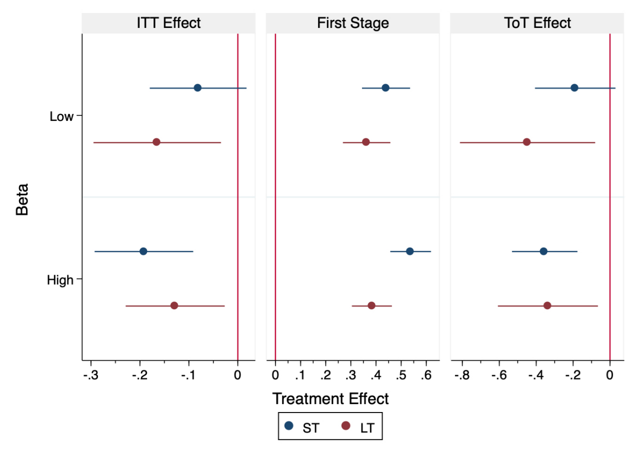
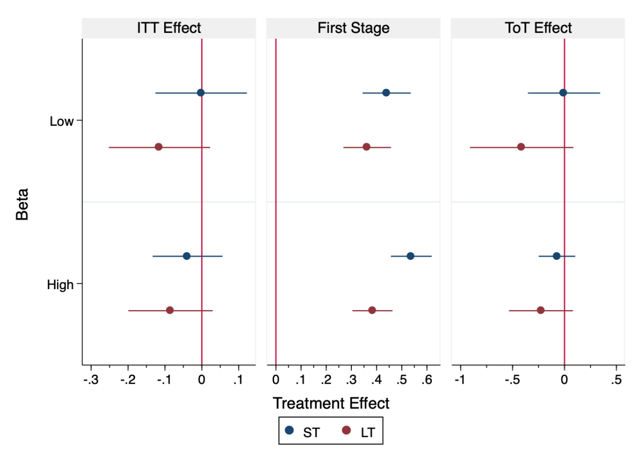


Figure 15c: Heterogeneous treatment effects of app limits on Instagram usage



Notes: This figure presents intention-to-treat, first-stage, and treatment-on-treated effects of the app limit intervention, by **present-bias parameter**. Short-term effects (duration of one week, between Survey 2 and 3) are in blue, whilst long-term effects (over the course of four weeks, between Survey 3 and 4) are in red. Error bars reflect 95 percent confidence intervals.

Table 15a: Regression results – Interaction of present-bias parameter with app treatment assignment

	(1)	(2)	(3)	(4)	(7)	(8)
Variables	Phone, ST	Phone, LT	FB, ST	FB, LT	Insta, ST	Insta, LT
Apptreat	-0.0339*	-0.0198	-0.147***	-0.119***	-0.0431	-0.0919**
	(0.0185)	(0.0222)	(0.0393)	(0.0423)	(0.0407)	(0.0464)
Apptreat * Beta	0.00146***	0.000578	-0.00137	-0.00661***	0.00607***	-0.00206
	(0.000419)	(0.000528)	(0.00175)	(0.00132)	(0.00197)	(0.00157)
Constant	5.643***	5.660***	3.541***	3.560***	2.884***	2.943***
	(0.0108)	(0.0113)	(0.0222)	(0.0247)	(0.0263)	(0.0259)
Observations	5,151	5,121	5,153	5,128	4,081	4,037
R-squared	0.003	0.003	0.011	0.009	0.008	0.007
Number of id	439	440	440	440	368	368

Robust standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table 15b: Regression results – Interaction of binary present-bias parameter (below/above median) with app treatment assignment

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone, ST	Phone, LT	FB, ST	FB, LT	Insta, ST	Insta, LT
Apptreat	-0.0346	-0.0564	-0.0886*	-0.176***	-0.00733	-0.123*
	(0.0285)	(0.0378)	(0.0508)	(0.0676)	(0.0667)	(0.0728)
Apptreat * Beta	0.00497	0.0644	-0.0985	0.0816	-0.0449	0.0275
(Above Median)	(0.0365)	(0.0454)	(0.0739)	(0.0850)	(0.0829)	(0.0934)
Constant	5.643***	5.661***	3.546***	3.566***	2.878***	2.946***
	(0.0107)	(0.0112)	(0.0221)	(0.0245)	(0.0262)	(0.0258)
Observations	5,205	5,175	5,213	5,188	4,128	4,085
R-squared	0.002	0.004	0.012	0.008	0.007	0.008
Number of id	444	445	445	445	372	372

Robust standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Notes: The former table presents regression results for the interaction of the present-bias parameter with app treatment. The latter table presents regression results for the interaction of a binary variable (=1 when the present-bias parameter is above median) with app treatment assignment. Columns (1) through (3) present the short-term intention-to-treat effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (duration of five weeks, between Survey 2 and Survey 4).

By Gender

Figure 16a: Heterogeneous treatment effects of app limits on phone usage, by gender

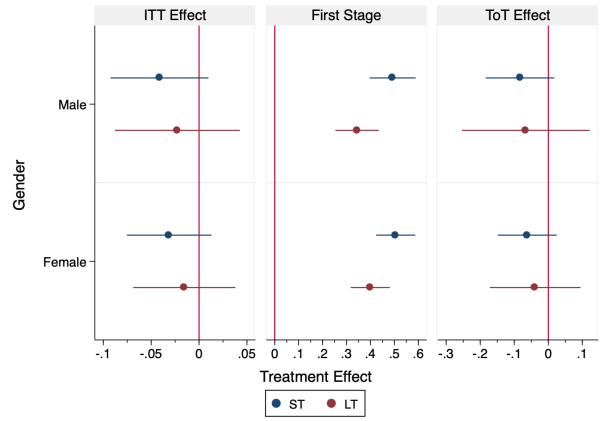


Figure 16b: Heterogeneous treatment effects of app limits on Facebook usage, by gender

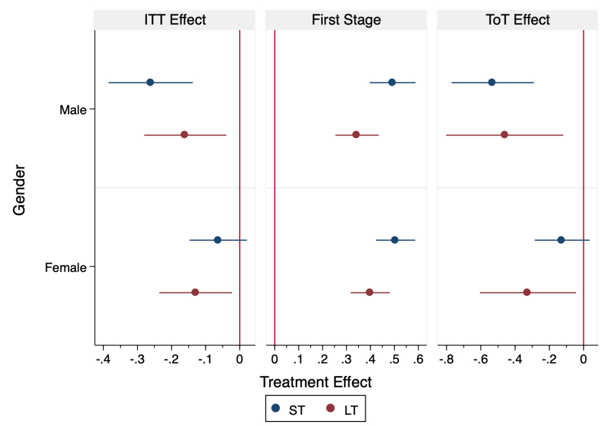
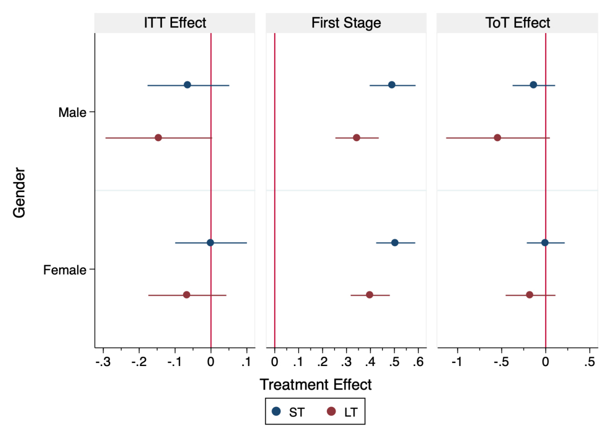


Figure 16c: Heterogeneous treatment effects of app limits on Instagram usage, by gender



Notes: This figure presents intention-to-treat, first-stage, and treatment-on-treated effects of the app limit intervention, by **gender**. Short-term effects (duration of one week, between Survey 2 and 3) are in blue, whilst long-term effects (over the course of four weeks, between Survey 3 and 4) are in red. Error bars reflect 95 percent confidence intervals.

Table 16: Regression results – Interaction of gender with app treatment assignment

Variables	(1) Phone, ST	(2) Phone, LT	(3) FB, ST	(4) FB, LT	(5) Insta, ST	(6) Insta, LT
Apptreat	-0.0430	-0.0311	-0.248***	-0.147**	-0.0647	-0.159**
	(0.0272)	(0.0344)	(0.0640)	(0.0618)	(0.0596)	(0.0771)
Apptreat * Female	0.0204	0.0245	0.178**	0.0375	0.0497	0.0878
	(0.0359)	(0.0437)	(0.0781)	(0.0828)	(0.0798)	(0.0954)
Constant	5.643***	5.661***	3.545***	3.565***	2.879***	2.946***
	(0.0107)	(0.0112)	(0.0220)	(0.0245)	(0.0262)	(0.0259)
Observations	5,198	5,168	5,206	5,181	4,123	4,078
R-squared	0.002	0.003	0.013	0.008	0.007	0.008
Number of id	443	444	444	444	371	371

Robust standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Notes: This table presents regression results for the interaction of a binary variable indicating gender (where Female = 1) with app treatment assignment. Columns (1) through (3) present the short-term intention-to-treat effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (duration of five weeks, between Survey 2 and Survey 4).

By Age

Figure 17a: Heterogeneous treatment effects of app limits on phone usage, by age

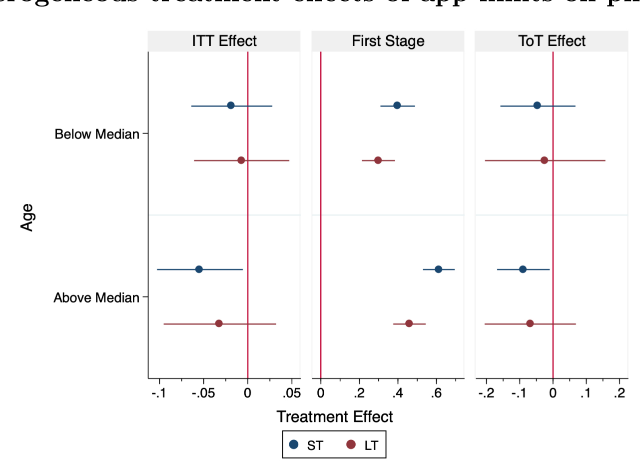


Figure 17b: Heterogeneous treatment effects of app limits on Facebook usage, by age

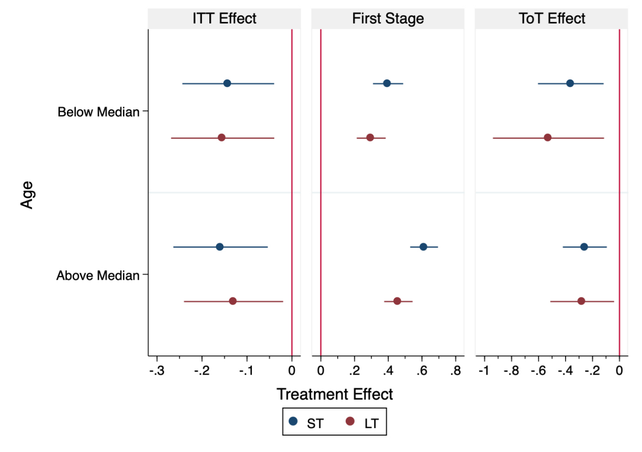
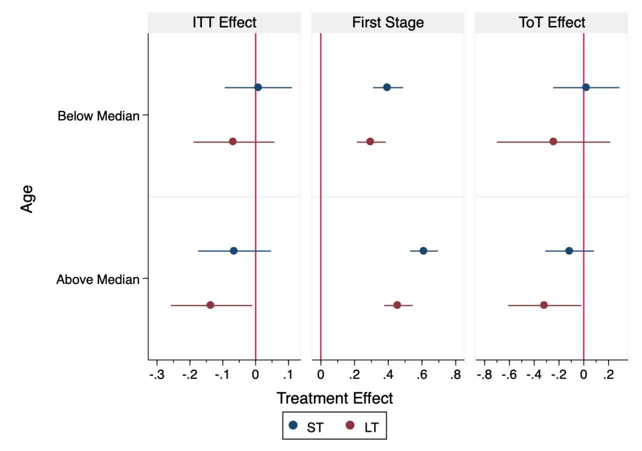


Figure 17c: Heterogeneous treatment effects of app limits on Instagram usage, by age



Notes: This figure presents intention-to-treat, first-stage, and treatment-on-treated effects of the app limit intervention, by **age**. Short-term effects (duration of one week, between Survey 2 and 3) are in blue, whilst long-term effects (over the course of four weeks, between Survey 3 and 4) are in red. Error bars reflect 95 percent confidence intervals.

Table 17a: Regression results – Interaction of age with app treatment assignment

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone, ST	Phone, LT	FB, ST	FB, LT	Insta, ST	Insta, LT
Apptreat	-0.00669	-0.0671	-0.195	-0.490**	0.0128	-0.104
	(0.0922)	(0.114)	(0.204)	(0.207)	(0.199)	(0.254)
Apptreat * Age	-0.00105	0.00210	0.00197	0.0154*	-0.00207	-0.00007
	(0.00375)	(0.00475)	(0.00825)	(0.00829)	(0.00833)	(0.0106)
Constant	5.643***	5.661***	3.546***	3.566***	2.878***	2.946***
	(0.0107)	(0.0112)	(0.0222)	(0.0245)	(0.0262)	(0.0258)
Observations	5,205	5,175	5,213	5,188	4,128	4,085
R-squared	0.002	0.003	0.011	0.009	0.007	0.008
Number of id	444	445	445	445	372	372

Robust standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table 17b: Regression results – Interaction of binary age (below/above median) with app treatment assignment

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone, ST	Phone, LT	FB, ST	FB, LT	Insta, ST	Insta, LT
Apptreat	-0.0122	-0.0138	-0.129**	-0.135**	-0.00987	-0.0855
	(0.0249)	(0.0281)	(0.0541)	(0.0601)	(0.0553)	(0.0643)
Apptreat * Age (Above Median)	-0.0389	-0.00777	-0.0390	0.0179	-0.0537	-0.0444
	(0.0355)	(0.0430)	(0.0767)	(0.0822)	(0.0794)	(0.0909)
Constant	5.643***	5.661***	3.546***	3.566***	2.878***	2.946***
	(0.0107)	(0.0112)	(0.0222)	(0.0245)	(0.0262)	(0.0258)
Observations	5,205	5,175	5,213	5,188	4,128	4,085
R-squared	0.003	0.003	0.011	0.008	0.007	0.008
Number of id	444	445	445	445	372	372

Robust standard errors in parentheses, $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Notes: The former table presents regression results for the interaction of age with app treatment. The latter table presents regression results for the interaction of a binary variable (=1 when age is above median) with app treatment assignment. Columns (1) through (3) present the short-term intention-to-treat effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (duration of five weeks, between Survey 2 and Survey 4).

6.7.1 Estimating Treatment Heterogeneity Using Causal Forests, Facebook use

Following Athey and Wager's (2019) application of causal forests to the National Study of Learning Mindsets dataset, I apply causal forests to estimate conditional average treatment effects (Athey, Tibshirani, and Wager, 2019; Nie and Wager, 2017). I base my analysis on data from $n = 459$ individuals who completed the third survey with valid Facebook time usage results.

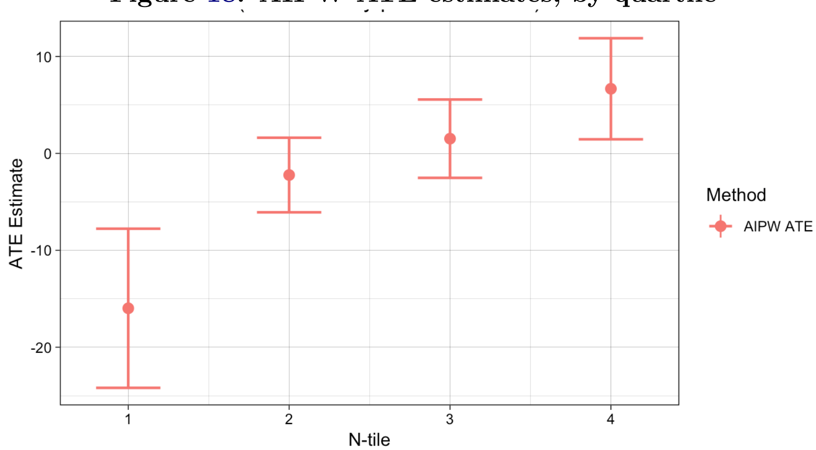
Assessing Treatment Heterogeneity Across Subgroups

I first test for heterogeneity by creating quartiled subpopulations based on predicted treatment effect strength. I compute the average treatment effect within each subgroup through the augmented inverse-propensity weighted (AIPW) average treatment effect, using doubly robust scores. Note that since the app treatment is randomized, the method will yield unbiased estimates of the group-specific treatment effect. This gives some indication of heterogeneity in the data: the ATE is clearly not constant across the quartiles. Using the Wald test, I reject the null hypothesis that the ATE is constant across quartiles.

Table 18: AIPW ATE estimates, by quartile

Quartile 1	-15.982
	(4.189)
Quartile 2	-2.232
	(1.963)
Quartile 3	1.519
	(2.062)
Quartile 4	6.668
	(2.660)
p-value	0.0003

Figure 18: AIPW ATE estimates, by quartile



Assessing Treatment Heterogeneity Across Covariates

Another way of checking heterogeneity is by checking if different groups have varying average covariate levels across quartiles of estimated conditional average treatment effects (CATEs).

Table 19: Average covariate values in each quartile

Covariate	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Female	0.663	0.514	0.598	0.497
	(0.036)	(0.036)	(0.036)	(0.036)
Age	23.88	23.03	23.88	23.03
	(0.327)	(0.328)	(0.327)	(0.328)
White	0.75	0.716	0.706	0.71
	(0.033)	(0.033)	(0.033)	(0.033)
Education	1.647	1.519	1.668	1.71
	(0.065)	(0.066)	(0.065)	(0.066)
Present-bias parameter	0.936	0.881	3.059	3.22
	(0.013)	(0.013)	(0.013)	(0.013)
Tangney Self-control Score	3.059	3.220	3.213	3.352
	(0.05)	(0.05)	(0.05)	(0.05)
Week 1 FB Usage	94.10	39.15	38.72	43.45
	(2.912)	(2.92)	(2.912)	(2.92)
Week 1 Act-Ideal, FB Usage	50.78	8.14	4.21	6.84
	(3.023)	(3.032)	(3.023)	(3.032)
Privacy Violations	0.174	0.200	0.082	0.008
	(0.015)	(0.015)	(0.015)	(0.015)
Clicked on social media links	0.002	0.033	0.027	0.033
	(0.011)	(0.011)	(0.011)	(0.011)

Table 20: Covariate variation across quartiles

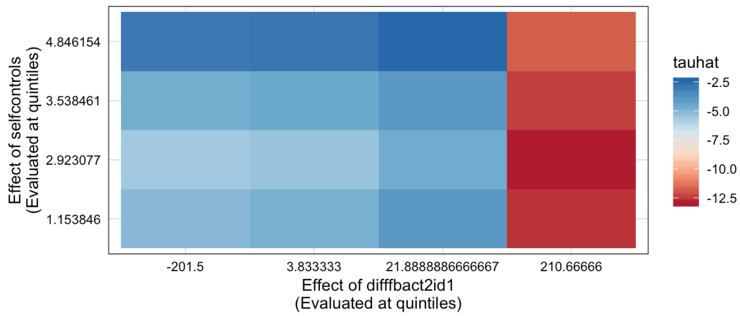
Week 1 FB Usage	0.2593
Week 1 Act-Ideal, FB Usage	0.1817
Privacy Violations	0.1256
Present-bias parameter	0.0387
Tangney Self-control Score	0.0233

There is clearly some heterogeneity across covariates: Table 20 displays the top 5 covariates with variation across quartiles. It makes sense that treatment strength would vary with pre-treatment Facebook usage, pre-treatment discrepancies between ideal and actual usage, and also with self-control proxies. Interestingly, there is also varying levels of privacy violations across quartiles of estimated CATEs, suggesting that people who have experienced privacy violations on Facebook before are more responsive to treatment (i.e., are more likely to experience stronger negative treatment effects).

Interaction between baseline Facebook usage and Tangney self-control score

In the body of my paper, I note that whilst heavy users experience stronger treatment effects, participants with *higher* Tangney self-control scores appear to benefit more from the treatment: thus, it would be instructive to look at the interaction of these two variables and their relationship with the conditional average treatment effect. As seen in Figure 19 below, I find that treatment effects are predicted to be strongest for those with very high discrepancies between their professed ideal times and the time they actually spent on Facebook, and that had a relatively moderate level of self-control.

Figure 19: Interaction between self-control score and pre-treatment discrepancies between ideal and actual time spent on Facebook



Lastly, we can evaluate the quality of our causal forest heterogeneity estimates using a method that estimates the best linear predictor of CATE using out-of-bag predictions (Chernozhukov et al., 2018). Using an omnibus test for the presence of heterogeneity, I reject the null of lack of heterogeneity.

Table 21: Omnibus test for heterogeneity

	Estimate	Std. Error	t-value	Pr(>t)
Mean Forest Prediction	0.985	0.637	1.545	0.0614
Differential Forest Prediction	2.584	0.925	2.795	0.0027

Notes: This table displays the best linear fit using forest predictions (on held-out data) as well as the mean forest prediction as regressors, along with one-sided heteroskedasticity-robust (HC3) standard errors.

6.8 Self-control Measures

Table 22: Regression results: Discrepancy between actual and ideal time on Tangney Brief Self-Control Score

	(1)	(2)	(3)
Variables	Phone	Facebook	Instagram
selfcontrols	-0.399***	-0.225**	-0.0715
	(0.132)	(0.0934)	(0.0959)
Constant	2.825***	1.570***	0.576*
	(0.432)	(0.305)	(0.312)
Observations	509	511	424
R-squared	0.018	0.011	0.001

Standard errors in parentheses, $^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$

Notes: This table presents regression results of the discrepancy between actual and ideal platform usage on Tangney's Brief Self-Control Score. Column (1) details results for phone usage, Column (2) for Facebook usage, and Column (3) for Instagram use.

Table 23: Regression results: Discrepancy between actual and ideal time on present-bias parameter

	(1)	(2)	(3)
Variables	Phone	Facebook	Instagram
beta	-0.128	0.322	0.479
	(0.575)	(0.344)	(0.417)
Constant	1.720***	0.514*	-0.0238
	(0.484)	(0.288)	(0.350)
Observations	332	333	272
R-squared	0.000	0.003	0.005

Standard errors in parentheses, $^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$

Notes: This table presents regression results of the discrepancy between actual and ideal platform usage on the present-bias parameter. Column (1) details results for phone usage, Column (2) for Facebook usage, and Column (3) for Instagram use.

6.9 Robustness Checks

6.9.1 Robustness to the inclusion of participants that already have an app limit prior to experiment

Note that in the body of the paper, users that already had existing app limits were omitted from the analysis. Here, I perform a robustness check, assigning participants that already have an app limit prior to experiment a treatment value of 1. Though some of the treatment effects are slightly diminished, there are no changes to the significance of the effects.

Table 25: Treatment-on-treated effects of app limits on logtime spent on Phone/FB/Insta

		Short-term			*Long-term*	
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone	Facebook	Instagram	Phone	Facebook	Instagram
Actual Treatment	-0.0589*	-0.277***	-0.0710	-0.0438	-0.314***	-0.292**
	(0.0327)	(0.0678)	(0.0794)	(0.0532)	(0.104)	(0.129)
Constant	5.643***	3.545***	2.878***	5.661***	3.565***	2.946***
	(0.0107)	(0.0220)	(0.0262)	(0.0112)	(0.0245)	(0.0261)
Observations	5,199	5,207	4,128	5,175	5,188	4,085
Number of id	443	444	372	445	445	372

Notes: Columns (1) through (3) present the short-term treatment-on-treated effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (over the course of five weeks, between Survey 2 and Survey 4). Note that actual treatment is coded as 1 for participants that already have an app limit prior to experiment.

6.9.2 Robustness check omitting participants that submitted inaccurate data

Recall that in every survey, I asked participants to input their phone/Facebook/Instagram usage times for the previous 7 days, as well as their weekly average as it appears on the "Screen Time" page. Additionally, participants submitted screenshots of their "Screen Time" page. To double-check for accuracy, I computed the average time over the 7 days and made sure that it was equal to the average specified by the participant. If there was a discrepancy between the two figures, I manually checked the uploaded screenshot to ensure that data has been inputted accurately. About 10% of participants had daily averages that did not match their inputted averages; upon manually checking their uploaded screenshots, these were mostly due to accidental misreporting. A small handful of participants consistently reported times that were at odds with their scrernshotted "Screen Time". Where possible, I manually inputted their real "Screen Time" as harvested from their screenshots, or otherwise omitted their observations.

Here, I run a robustness check omitting participants who ever had inaccurate data from the the analysis, yielding similar results.

Table 25: Treatment-on-treated effects of app limits on logtime spent on Phone/FB/Insta

	Short-term			*Long-term*		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Phone	Facebook	Instagram	Phone	Facebook	Instagram
Actual Treatment	-0.0733**	-0.338***	-0.0687	-0.0388	-0.361***	-0.398***
	(0.0358)	(0.0784)	(0.0895)	(0.0592)	(0.106)	(0.134)
Constant	5.626***	3.510***	2.863***	5.644***	3.526***	2.923***
	(0.0120)	(0.0256)	(0.0286)	(0.0124)	(0.0272)	(0.0298)
Observations	4,319	4,332	3,418	4,201	4,214	3,307
Number of id	369	370	309	362	362	302

Notes: Columns (1) through (3) present the short-term treatment-on-treated effects (i.e. the duration of one week, between Survey 2 and Survey 3), and Columns (4) through (6) present the longer-term effects (over the course of five weeks, between Survey 2 and Survey 4). Note that actual treatment is coded as 1 for participants that already have an app limit prior to experiment.

7 Bibliography

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow. 2019. "The Welfare Effects of Social Media," *NBER Working Paper*, w25514.

Ambec, S., and N. Treich. 2007. "ROSCAs as financial agreements to cope with self-control problems," *Journal of Development Economics*, 82(1): 120–37.

Andreassen, C.S., T. Torsheim, G.S. Brunborg, and S. Pallesen. 2012. "Development of a Facebook addiction scale," *Psychological Reports*, 110(2012): 501-517.

Andreassen, C.S., T. Torsheim, M.D. Griffiths, S.R. Gjertsen, E. Krossbakken, S. Kvam, and S. Pallesen. 2013. "The relationship between behavioral addictions and the five-factor model of personality," *Journal of Behavioral Addictions*, 2(2013): 90-99.

Andreassen, C.S and S. Pallesen. 2014. "Social Network Site Addiction – An Overview," *Current Pharmaceutical Design*, 20(25): 4053-4061.

Andreassen, C.S. 2015. "Online Social Network Site Addiction: A Comprehensive Review," *Technology and Addiction*, 2(2): 175–184.

Andreassen, C.S., S. Pallesen, and M.D. Griffiths. 2017. "The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey," *Addictive Behaviors*, 64: 287-293.

Angrist, J. D. and G. W. Imbens. 1995. "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American Statistical Association*, 90(430): 431-442.

Ashraf N., D. Karlan, and W. Yin. 2006. "Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines." *Quarterly Journal of Economics*, 121(2): 635–72.

Augenblick, N., M. Niederle, M. and C. Sprenger. 2015. "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*, 130 (3): 1067-1115.

Augenblick, N., and M. Rabin. 2018. "An experiment on time preference and misprediction in unpleasant tasks," *Review of Economic Studies*, 86(3): 941-975.

Basu, K. 2008. "Hyperbolic Discounting and the Sustainability of Rotational Savings Arrangements," *American Economic Journal: Microeconomics*, 3(4): 143-71.

Becker, G. S., and K. M. Murphy. 1988. "A Theory of Rational Addiction," *Journal of Political Economy*, 96 (4): 675-700.

Bénabou, R., and J. Tirole. 2004. "Willpower and personal rules," *Journal of Political Economy*, 112(4): 848-886.

Benartzi, S., and R. Thaler. 2004. "Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, 112(1): 164-187.

Bernheim, D. and A. Rangel. 2007. "Toward Choice-Theoretic Foundations for Behavioral Welfare Economics," *American Economic Review*, 97(2): 464-470.

Beshears, J., J.J. Choi, D. Laibson, and B. Madrian. 2008. "How are preferences revealed?" *Journal of Public Economics*, 92(8-9): 1787-1794.

Breman, A. 2009. "Give more tomorrow: two field experiments on altruism and intertemporal choices," *Journal of Public Economics*, 95(11): 1349-1357.

Bryan, G., D. Karlan, and S. Nelson. 2010. "Commitment devices," *Annual Review of Economics*, 2(1), 671– 698.

Chetty, R. 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80: 969–1018.

Deursen, A., C.L. Bolle, S.M. Hegner, and P.A.M. Kommers. 2015. "Modeling habitual and addictive smartphone behavior," *Computers in Human Behavior*, 45: 411-420.

Duflo E., M. Kremer, and J. Robinson. 2009. "Nudging farmers to use fertilizer: theory and experimental evidence from Kenya," *NBER Working Paper* w15131.

Duggan, M. and J. Brenner. 2013. *The demographics of social media users.* Washington, DC: Pew Research Center's Internet & American Life Project.

eMarketer, Squarespace. 2018. "Number of Facebook users by age in the U.S. as of January 2018 (in millions)." *Statista - The Statistics Portal.* https://www.statista.com/statistics/398136/us-facebook-user-age-groups, Accessed February 27 2019

Elphinston, R. A. and P. Noller. 2011. "Facebook intrusion and the implications for romantic jealousy and relationship satisfaction," *Cyberpsychology, Behavior, and Social Networking*, 14: 631–635.

Felisoni, D. and A.S. Godoi. 2018. "Cell phone usage and academic performance: An experiment," *Computers and Education*, 117: 175-187.

Fudenberg, D., and D. Levine. 2006. "A Dual-Self Model of Impulse Control," *The American Economic Review*, 96(5): 1449-1476.

Freeman, D., P. Manzini, M. Mariotti, and L. Mittone. 2016. "Procedures for Eliciting Time Preferences," *IZA DP*, No. 9857.

Giné, X., D. Karlan, and J. Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation," *American Economic Journal: Applied Economics*, 2(4): 213-35.

GlobalWebIndex. 2017. "Daily Time Spent on Social Networking by Internet Users Worldwide from 2012 to 2017 (in Minutes)." *Statista - The Statistics Portal*, www.statista.com/statistics/433871/daily-social-media-usage-worldwide/, Accessed 22 Sep 2018

Griffiths, M.D. 2005. "A components model of addiction within a biopsychosocial framework," *Journal of Substance Use*, 10(2005): 191-197.

Guardian. 2018. "Why Hunt's screen time limits for kids are scientific nonsense," *The Guardian.* https://www.bbc.com/news/uk-43853678, Accessed April 27 2019

Gugerty, M. 2007. "You Can't Save Alone: Commitment in Rotating Savings and Credit Associations in Kenya," *Economic Development and Cultural Change*, 55(2), 251-282.

Hong, F., D. Huang, H. Lin, and S. Chiu. 2014. "Analysis of the psychological traits, Facebook usage, and Facebook addiction model of Taiwanese university students," *Telemat-*

ics and Informatics, 31: 597–606.

Hoong, Ruru. 2019. "The Behavioral Economics of Social Media: A Study of Self Commitment Devices and the Facebook Privacy Paradox," *Stanford Digital Repository*. Available at: https://purl.stanford.edu/qt999vz1100

Hunt, M. G., R. Marx, C. Lipson, and J. Young. 2018. "No More FOMO: Limiting Social Media Decreases Loneliness and Depression," *Journal of Social and Clinical Psychology*, 37(10): 751–768.

Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2): 263–291.

Karaiskos, D., E. Tzavellas, B. Georgia and T. Paparrigopoulos. 2010. "Social network addiction: A new clinical disorder?" *European Psychiatry* 25: 855-855.

Kan, K. 2007. "Cigarette smoking and self-control," *Journal of Health Economics*, 26(1): 61-81.

Kim, B.E., D. Seligman and J. W. Kable. 2012. "Preference reversals in decision making under risk are accompanied by changes in attention to different attributes," *Frontiers in Neuroscience*, 6: 1-10.

Ko, C. H., J. Y. Yen, C. C. Chen, S. H. Chen, K. Wu, and C. F. Yen. 2006. "Tridimensional personality of adolescents with internet addiction and substance use experience." *The Canadian Journal of Psychiatry*, 51(14): 887-894.

Koc, M. and S. Gulyagci. 2013. "Facebook addiction among Turkish college students: The role of psychological health, demographic, and usage characteristics." *CyberPsychology, Behavior, and Social Networking*, 16: 279–284.

Kuss, D. and M.D. Griffiths. 2011. "Online Social Networking and Addiction – A Review of the Psychological Literature." *International Journal of Environmental Research and Public Health*, 8: 3528-52.

Laibson, D. 1997. "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112(2): 443-477.

Laibson, D., A. Repetto, and J. Tobacman. 1998. "Self-Control and Saving for Retirement," *Brookings Papers on Economic Activity*, 1998(1): 91-196.

Lee, Y.K., C. T. Chang, Y. Lin, and Z. H. Cheng. 2014. "The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress," *Computers in Human Behavior*, 31: 373-383.

Lenhart, A., and M. Madden. 2007. "Social networking websites and teens: An overview," *Working Paper.*

Madrian, B. C., and D.F. Shea. 2001. "The power of suggestion: Inertia in 401 (k) participation and savings behavior," *The Quarterly Journal of Economics* 116 (4), 1149–1187.

Marotta, V. and A. Acquisti. 2017. "Online Distractions, Website Blockers, and Economic Productivity: A Randomized Field Experiment," *Preliminary Draft.*

Madden, G. J., Raiff, B. R., Lagorio, C. H., Begotka, A. M., Mueller, A. M., Hehli, D. J., Wegener, A. A. 2004. "Delay Discounting of Potentially Real and Hypothetical Rewards: II. Between- and Within-Subject Comparisons," *Experimental and Clinical Psychopharmacology*, 12(4): 251–261.

Mosquera, R., M. Odunowo, T. McNamara, X. Guo, and R. Petrie. 2018. "The Economic Effects of Facebook," *Working Paper, Texas A&M University.*

O'Donoghue T., and M. Rabin. 1999. "Doing it now or later," *American Economic Review*, 89: 103–24.

O'Donoghue, T., and M. Rabin. 2001. "Choice and Procrastination," *Quarterly Journal of Economics*, 116(1): 121-160.

O'Donoghue, T., and M. Rabin. 2006. "Incentives and Self-Control. Advances," *Economics and Econometrics: Theory and Applications*, Ninth World Congress, 215-245.

Pelling, E. L. and K.M. White. 2009. "The theory of planned behavior applied to young people's use of social networking Web sites." *Cyberpsychology and Behavior,* 12(6): 755-9.

Ryan, T., A. Chester, J. Reece, S. Xenos. 2014. "The uses and abuses of Facebook: A review of Facebook addiction, *Journal of Behavioral Addictions*, 3(3): 133–148.

Samuelson, W., and R. Zeckhauser. 1988. "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1: 7-59.

Social Media Addiction Reduction Technology Act, 116th Congress, 2019. https://www.hawley.senate.gov/ 07/Social-Media-Addiction-Reduction-Technology-Act.pdf

Strotz, R. H. 1956. "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, XXIII, 165–80.

Tangney, J. P., R. F. Baumeister, A. L. and Boone. 2004. "High self-control predicts good adjustment, less pathology, better grades, and interpersonal success," *Journal of Personality*, 72: 271–322.

Thaler, R. H., and H. M. Shefrin. 1981. "An Economic Theory of Self-Control," *Journal of Political Economy*, 89 (2): 392-406.

Thaler, R. H., C.R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness.* New Haven: Yale University Press.

Tromholt, M. 2016. "The Facebook Experiment: Quitting Facebook Leads to Higher Levels of Well-Being." *Cyberpsychology, Behavior, and Social Networking* 19(11): 661–666.

Twenge, J. M. 2018. *IGen.* New York, NY: Atria Paperback.

Vanman, Eric J., Rosemary Baker, and Stephanie J. Tobin. 2018. "The Burden of Online Friends: The Effects of Giving Up Facebook on Stress and Well-Being," *The Journal of Social Psychology* 158(4):496–507.

VentureBeat. 2019. "Over 90% of Facebook's advertising revenue now comes from mobile," https://venturebeat.com/2018/04/25/over-90-of-facebooks-advertising-revenue-now-comes-from-mobile, Accessed 28 April 2019.

Wilson, K., S. Fornasier and K. White. 2010. "Psychological predictors of young adults' use of social networking sites," *Cyberpsychology, Behavior, and Social Networking*, 13(2): 173-7.