Online Appendix

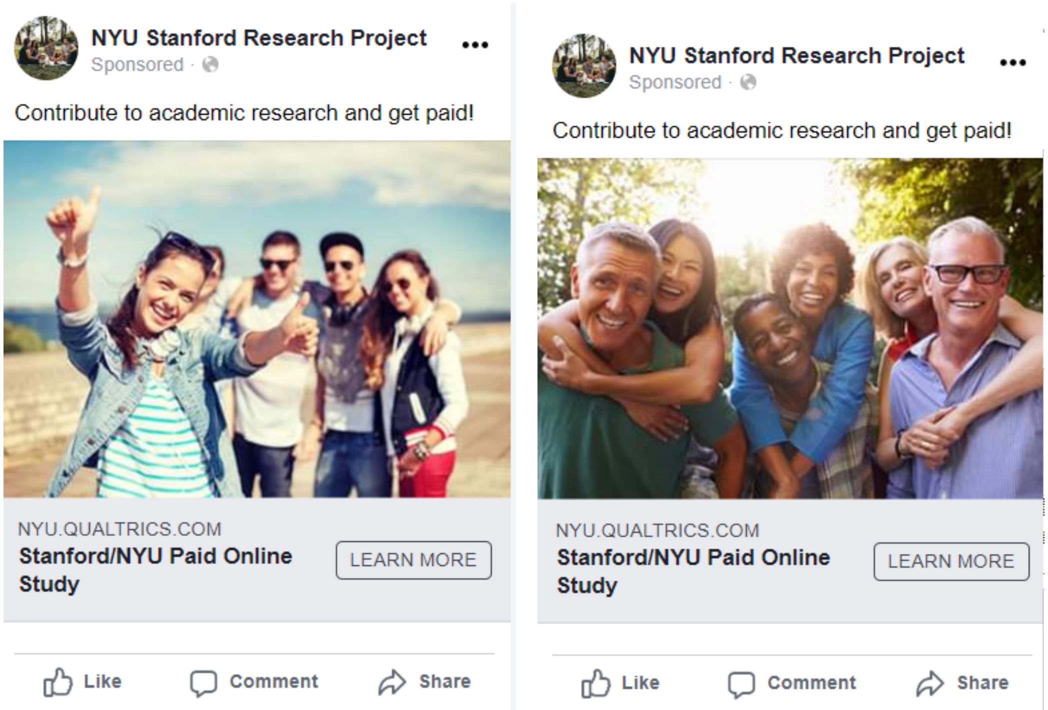
Digital Addiction

Hunt Allcott, Matthew Gentzkow, and Lena Song

Table of Contents

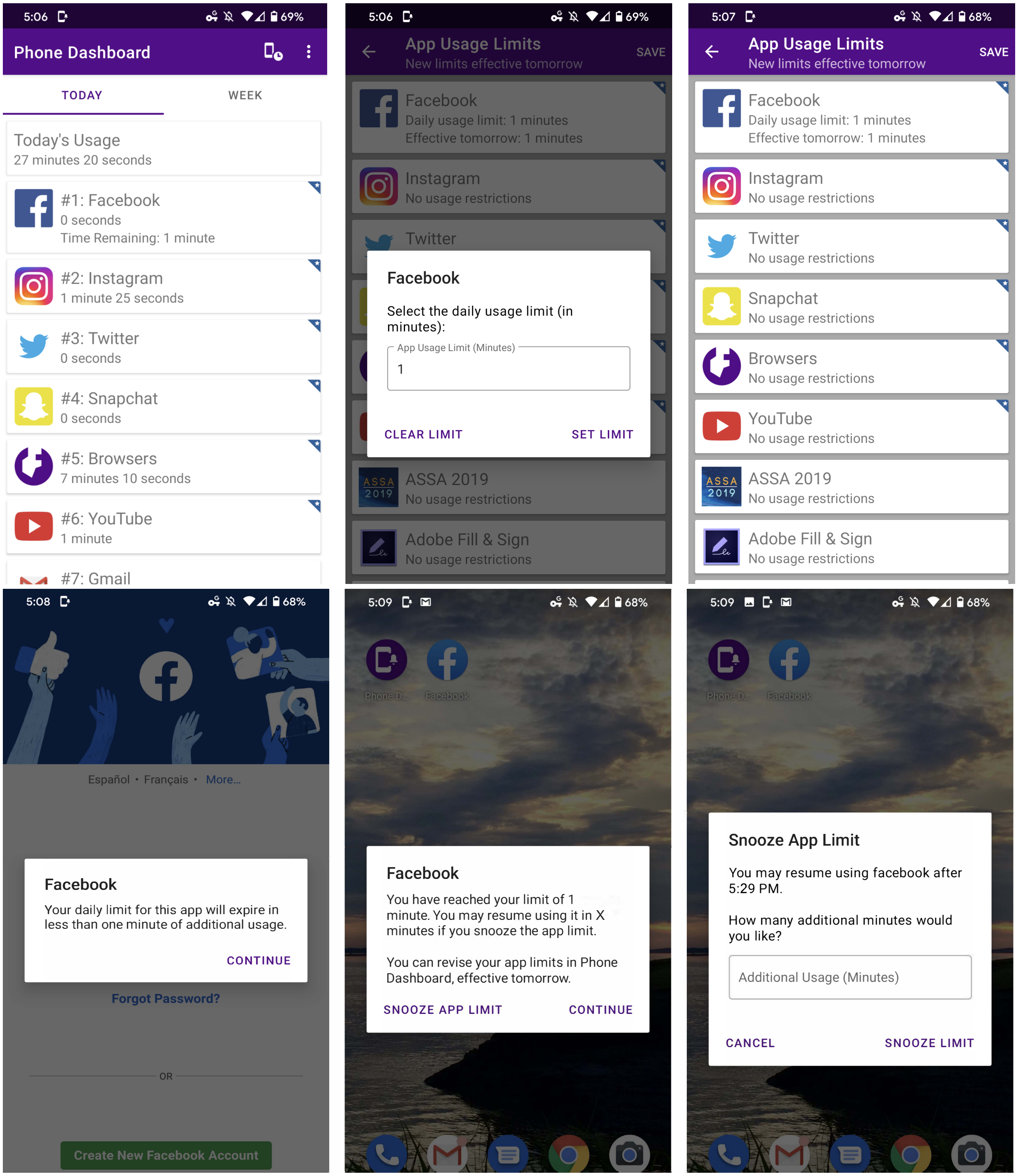
A Experimental Design Appendix

Figure A1: **Facebook Recruitment Ads**



Notes: The ads at left and right were shown to users aged 18–34 and 35–64, respectively.

Figure A2: **Phone Dashboard Screenshots**



Notes: This figure presents screenshots of the Phone Dashboard app. The top left presents the day's total usage by app. The top middle shows how a user can set daily a daily usage limit for each app, effective tomorrow. The top right shows the usage limits set for each app. The bottom left shows the warning users receive when they are within five minutes or one minute of their limit. The bottom middle shows the message users receive when they reach the limit. Users with the snooze functionality can resume using an app after a delay of $X \in \{0, 2, 5, 20\}$ minutes. The bottom right shows the option for a user to choose how many additional minutes to add to the daily limit after the snooze delay. All participants had the usage information in the top left panel, while only the Limit group had the time limit functionalities in the other panels.

A.1 Variable Definitions

Ideal use change. Some people say they use their smartphone too much and ideally would use it less. Other people are happy with their usage or would ideally use it more. How do you feel about your overall smartphone use over the past 3 weeks?

- I used my smartphone too much.
- I used my smartphone the right amount.
- I used my smartphone too little.

Relative to your actual use over the past 3 weeks, by how much would you ideally have [if "too much": reduced. If "too little": increased] your smartphone use? Please give a number in percent. ____ %

Addiction scale. Over the past 3 weeks, how often have you...

- Been worried about missing out on things online when not checking your phone?
- Checked social media, text messages, or email immediately after waking up?
- Used your phone longer than intended?
- Found yourself saying "just a few more minutes" when using your phone?
- Used your phone to distract yourself from personal problems?
- Used your phone to distract yourself from feelings of guilt, anxiety, helplessness, or depression?
- Used your phone to relax in order to go to sleep?
- Tried to reduce your phone use without success?
- Experienced that people close to you are concerned about the amount of time you use your phone?
- Felt anxious when you don't have your phone?
- Found it difficult to switch off or put down your phone?
- Been annoyed or bothered when people interrupt you while you use your phone?
- Felt your performance in school or at work suffers because of the amount of time you use your phone?
- Lost sleep due to using your phone late at night?
- Preferred to use your phone rather than interacting with your partner, friends, or family?
- Put off things you have to do by using your phone?

Never, Rarely, Sometimes, Often, Always

SMS addiction scale.

- In the past 24 hours, did you use your phone longer than intended?
- In the past 24 hours, did your performance at school or work suffer because of the amount of time you used your phone?
- In the past 24 hours, did you feel like you had an easy time controlling your screen time?
- In the past 24 hours, did you use your phone mindlessly?
- In the past 24 hours, did you use your phone because you were feeling down?
- In the past 24 hours, did using your phone keep you from working on something you needed to do?
- In the past 24 hours, would you ideally have used your phone less?
- Last night, did you lose sleep because of using your phone late at night?
- When you woke up today, did you immediately check social media, text messages, or email?

Please text back your answer on a scale from 1 (not at all) to 10 (definitely).

Phone makes life better. To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?

11-point scale from -5 (Makes my life worse) to 0 (Neutral) to 5 (Makes my life better)

Subjective well-being. Please tell us the extent to which you agree or disagree with each of the following statements. Over the past 3 weeks, ...

- … I was a happy person
- … I was satisfied with my life
- … I felt anxious
- … I felt depressed
- … I could concentrate on what I was doing
- … I was easily distracted
- … I slept well

7-point scale from strongly disagree to neutral to strongly agree

B Data Appendix

Table A1: **Response Rates**

(a) **Limit**

	(1) Control	(2) All limits	(3) Snooze 0	(4) Snooze 2	(5) Snooze 5	(6) Snooze 20	(7) No snooze	(8) F-test p-value
Completed survey 3	0.97	0.96	0.96	0.98	0.96	0.97	0.95	0.51
Completed survey 4	0.95	0.94	0.95	0.95	0.94	0.95	0.93	0.81
Have period 2 usage	1.00	1.00	0.99	0.99	1.00	1.00	1.00	0.23
Have period 3 usage	0.99	0.98	0.99	0.98	0.99	0.98	0.98	0.68
Have period 4 usage	0.98	0.97	0.99	0.98	0.97	0.97	0.96	0.37
Have period 5 usage	0.97	0.96	0.97	0.96	0.96	0.97	0.95	0.70

(b) **Bonus**

	(1) Control	(2) Treatment	(3) t-test p-value
Completed survey 3	0.97	0.96	0.74
Completed survey 4	0.95	0.95	0.64
Have period 2 usage	1.00	1.00	0.16
Have period 3 usage	0.98	0.98	0.95
Have period 4 usage	0.98	0.97	0.84
Have period 5 usage	0.96	0.96	0.85

Notes: Columns 1 and 2 of Panel (a) present present response rates for Limit and Limit Control groups. Columns 3–7 present response rates for each of the snooze delay conditions within the Limit group. Column 8 presents the p-value of an F-test of differences between the Limit Control and the separate snooze delay conditions. Columns 1 and 2 of Panel (b) present response rates for Bonus and Bonus Control groups. Column 3 presents the p-value of a t-test of differences between the Bonus and Bonus Control groups.

Table A2: **Covariate Balance**

(a) **Limit**

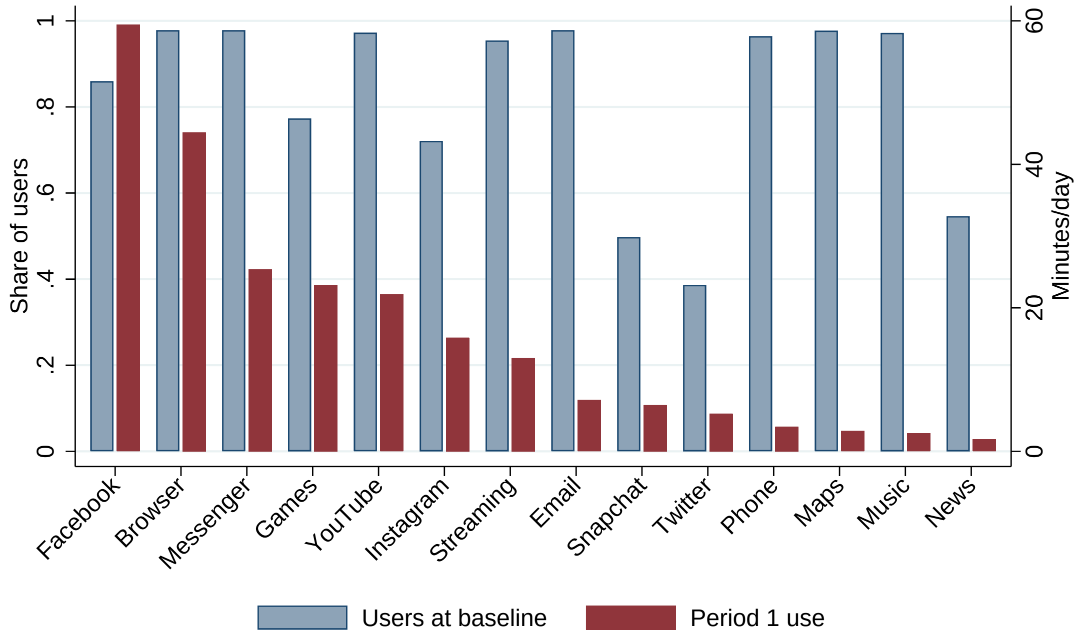
Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	t-test p-value (1)-(2)
Income ($000s)	40.15	41.76	0.35
	(36.22)	(37.84)	
College	0.67	0.67	0.72
	(0.47)	(0.47)	
Male	0.38	0.40	0.51
	(0.49)	(0.49)	
White	0.70	0.74	0.13
	(0.46)	(0.44)	
Age	33.61	33.79	0.76
	(12.33)	(12.35)	
Period 1 FITSBY use (minutes/day)	151.96	154.07	0.64
	(92.00)	(99.19)	
N	1150	783	
F-test of joint significance (p-value)			0.65
F-test, number of observations			1933

(b) **Bonus**

Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	t-test p-value (1)-(2)
Income ($000s)	41.26	40.65	0.76
	(39.16)	(36.11)	
College	0.67	0.67	0.75
	(0.47)	(0.47)	
Male	0.41	0.38	0.26
	(0.49)	(0.49)	
White	0.71	0.72	0.61
	(0.46)	(0.45)	
Age	33.53	33.73	0.76
	(12.17)	(12.40)	
Period 1 FITSBY use (minutes/day)	151.24	153.34	0.67
	(91.97)	(95.94)	
N	479	1454	
F-test of joint significance (p-value)			0.94
F-test, number of observations			1933

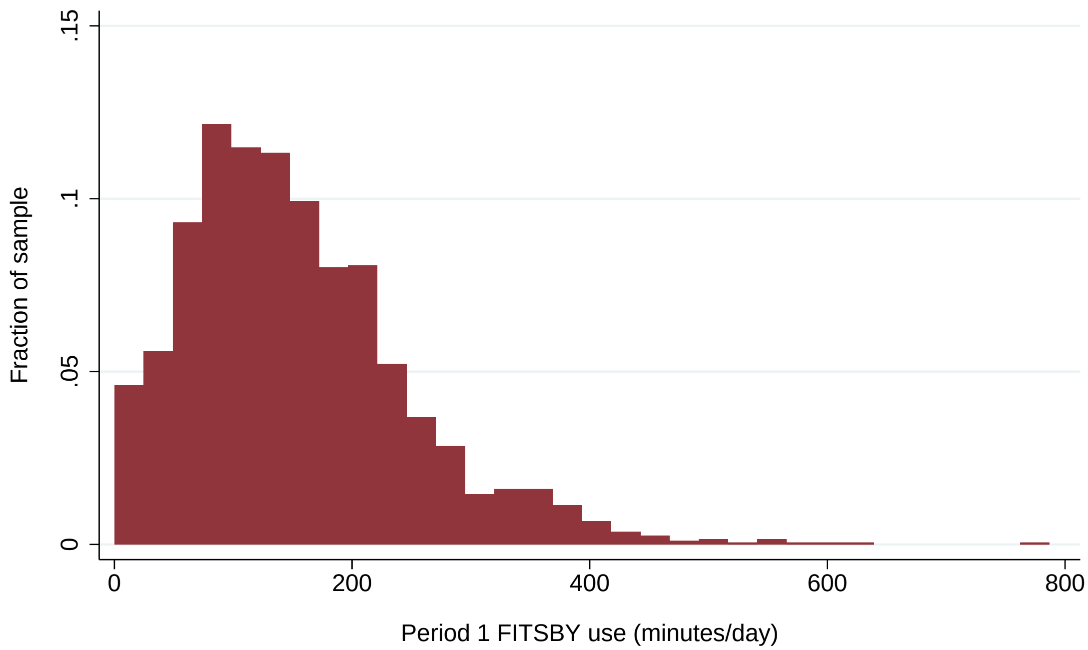
Notes: Panels (a) and (b) present tests of covariate balance for the Limit and Bonus treatment and control groups.

Figure A3: **Most Popular Apps**



Notes: This figure presents the share of users that have each app and the average daily screen time in period 1 (baseline). Period 1 use is across all users, not conditioning on whether or not they have the app.

Figure A4: **Distribution of Baseline FITSBY Use**



Notes: This figure presents a distribution of FITSBY use in period 1 (baseline). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

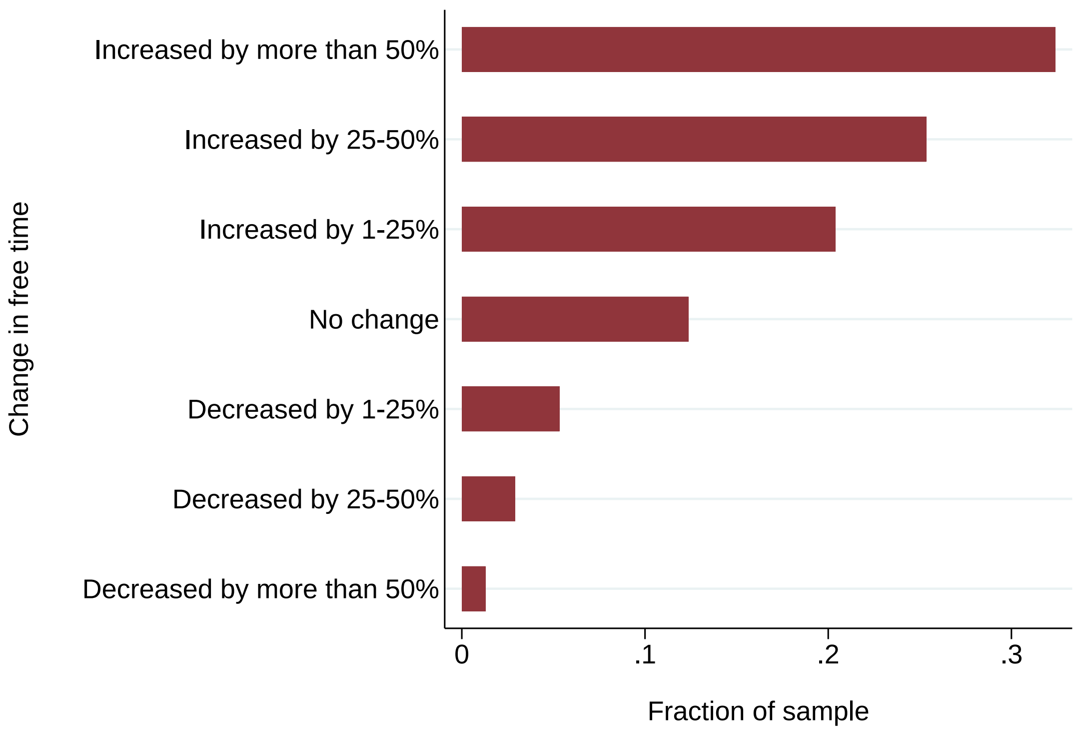
Table A3: **Descriptive Statistics for Survey Outcome Variables**

	Mean	Standard deviation	Minimum value	Maximum value
Ideal use change	-19.0	21.4	-100	70
Addiction scale x (-1)	-6.2	2.6	-16	0
SMS addiction scale x (-1)	1.7	3.1	-9	9
Phone makes life better	1.6	2.0	-5	5
Subjective well-being	0.2	2.5	-7	7

Notes: This table present descriptive statistics for the survey outcome variables at baseline.

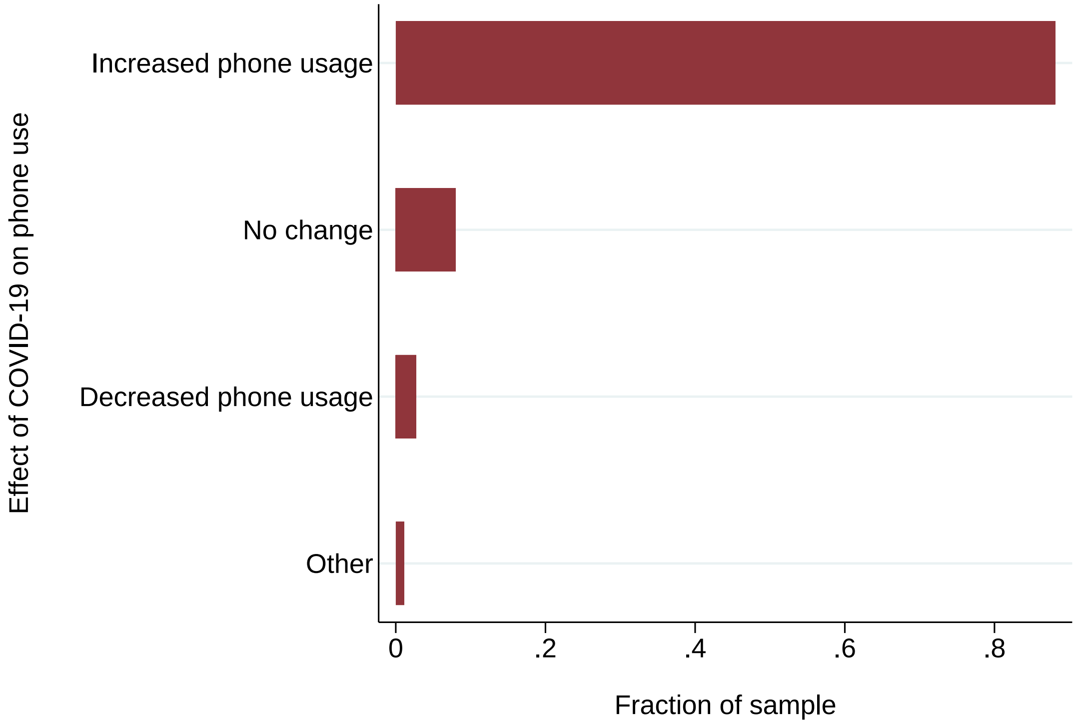
C Differences Between 2019 and the Study Period

Figure A5: **Effects of Coronavirus Outbreak on Free Time**



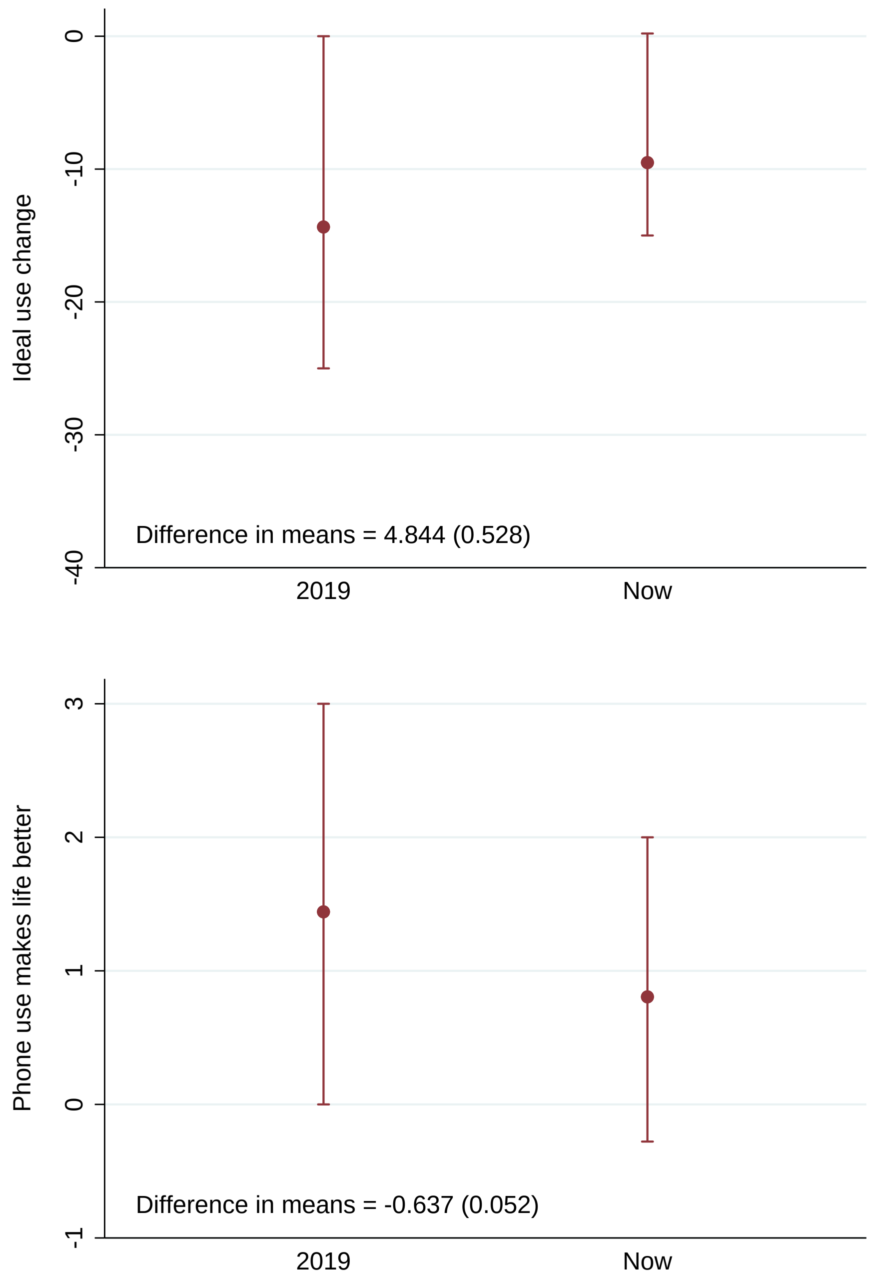
Notes: This figure presents the distribution of responses to the baseline survey question, "To what extent has the recent coronavirus outbreak changed how much free time you have?"

Figure A6: **Effects of Coronavirus on Smartphone Use**



Notes: The baseline survey asked, "How has the recent coronavirus outbreak changed how you use your smartphone?" We coded the responses as to whether they indicated increased, decreased, or unchanged smartphone use.

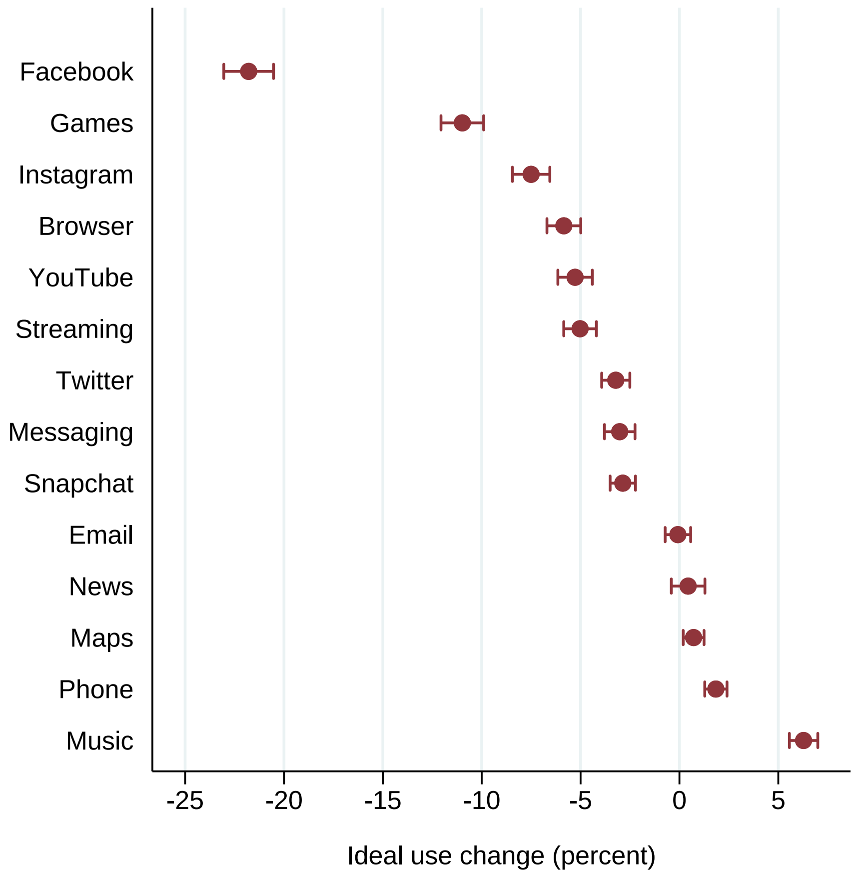
Figure A7: **Self-Control Problems in 2019 versus Now**



Notes: This figure presents the mean (dots) and 25th and 75th percentiles (spikes) of responses to *ideal use change* and *phone use makes life better* for 2019 and for the past 3 weeks, as reported on the baseline survey. *Ideal use change* is the answer to, “Relative to your actual use [in 2019 / over the past 3 weeks], by how much would you ideally have [reduced/increased] your screen time? *Phone use makes life better* is the answer to, “To what extent do you think your smartphone use made your life better or worse [in 2019 / over the past 3 weeks]?”

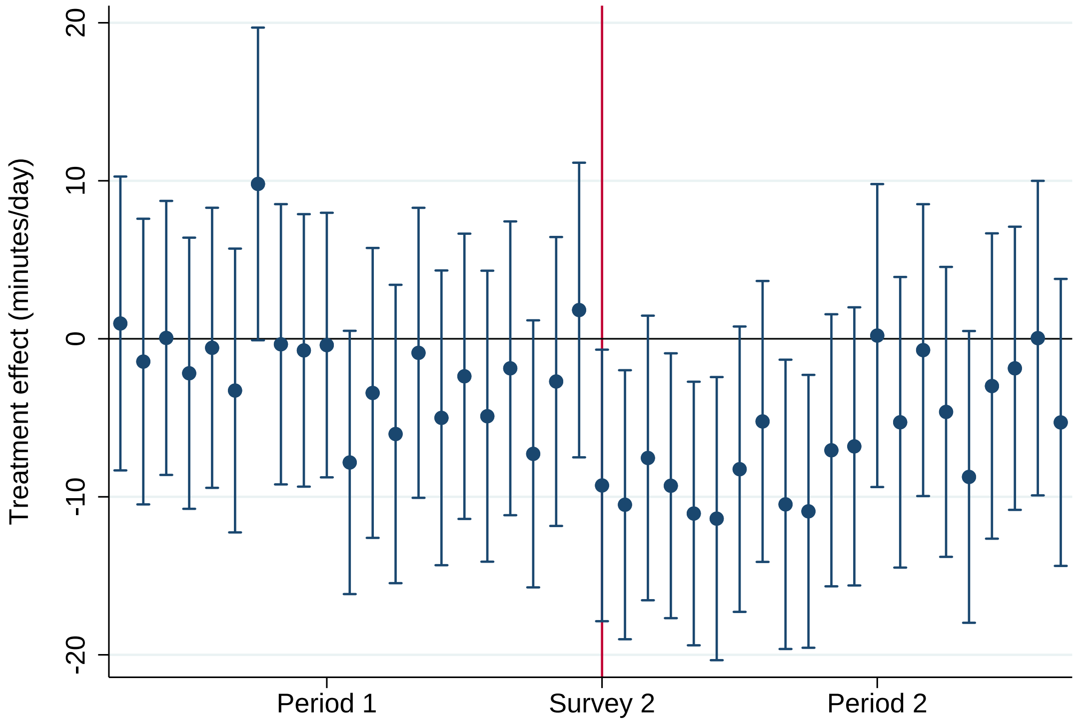
D Model-Free Results Appendix

Figure A8: **Ideal Use Change by App or Category**



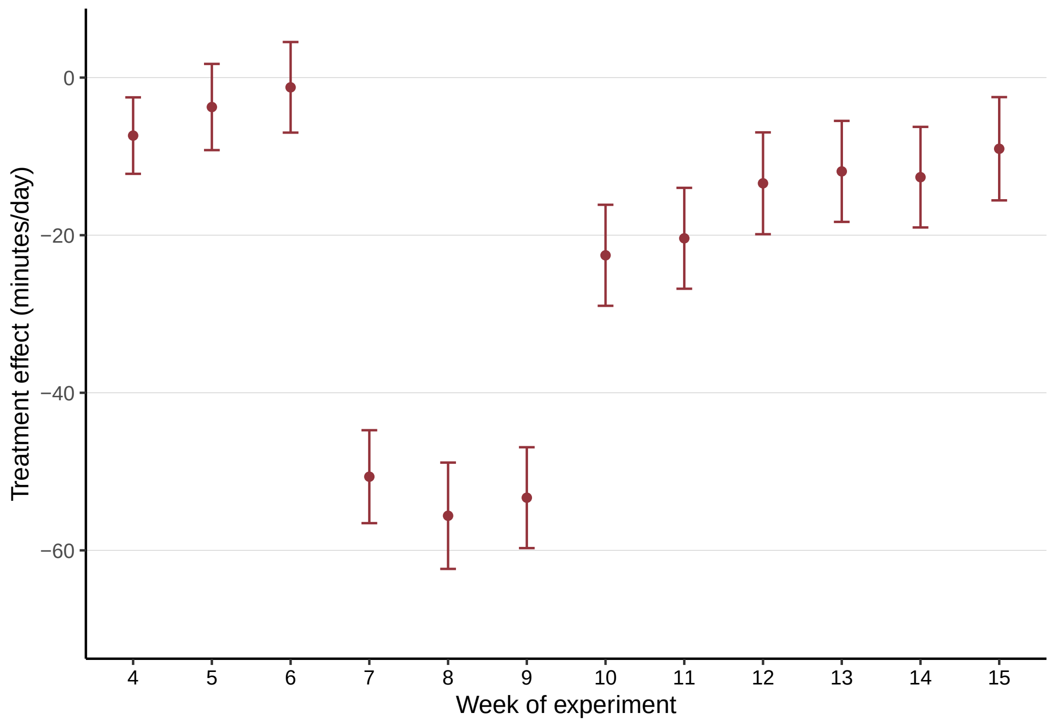
Notes: This figure presents mean *ideal use change* by app or app category at baseline. *Ideal use change* is the answer to, “Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?” We code “I don’t use this app at all” as 0, so these results reflect how much each app contributes to overall temptation, not how tempting each app is for the subset of people who use it.

Figure A9: **Effects of Bonus on FITSBY Use by Day for Periods 1 and 2**



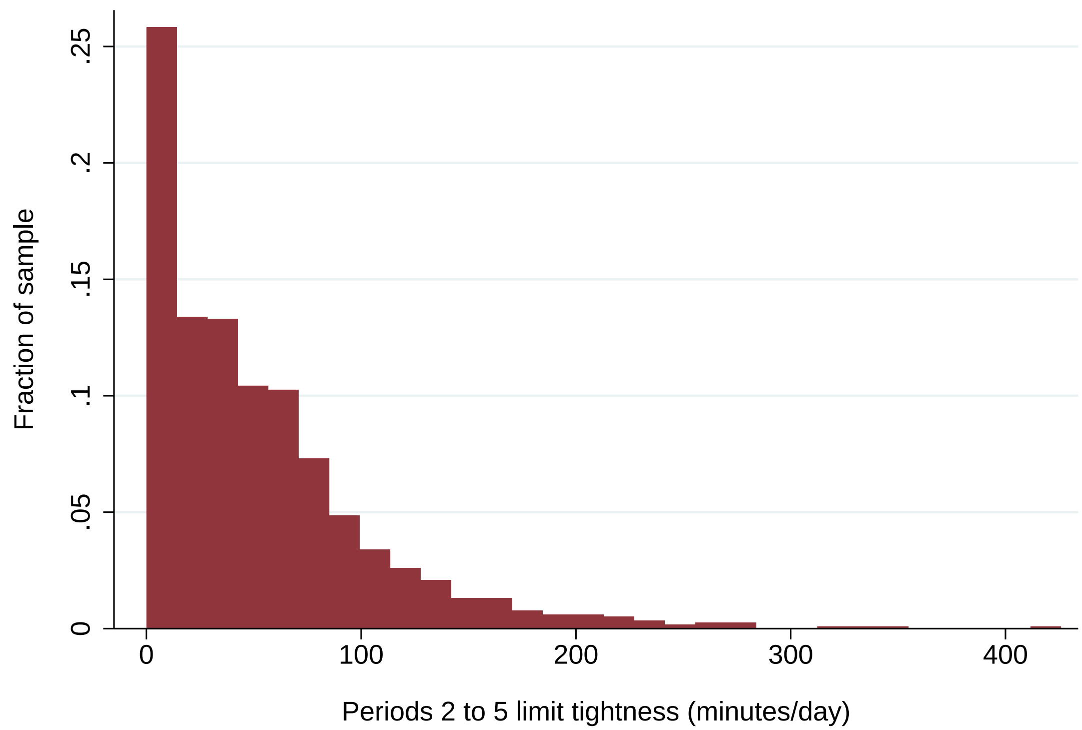
Notes: This figure presents differences in average FITSBY use between the Bonus and Bonus Control group for each day of periods 1 and 2. The vertical line indicates the day of survey 2, when the bonus was announced. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure A10: **Effects of Bonus on FITSBY Use by Week**



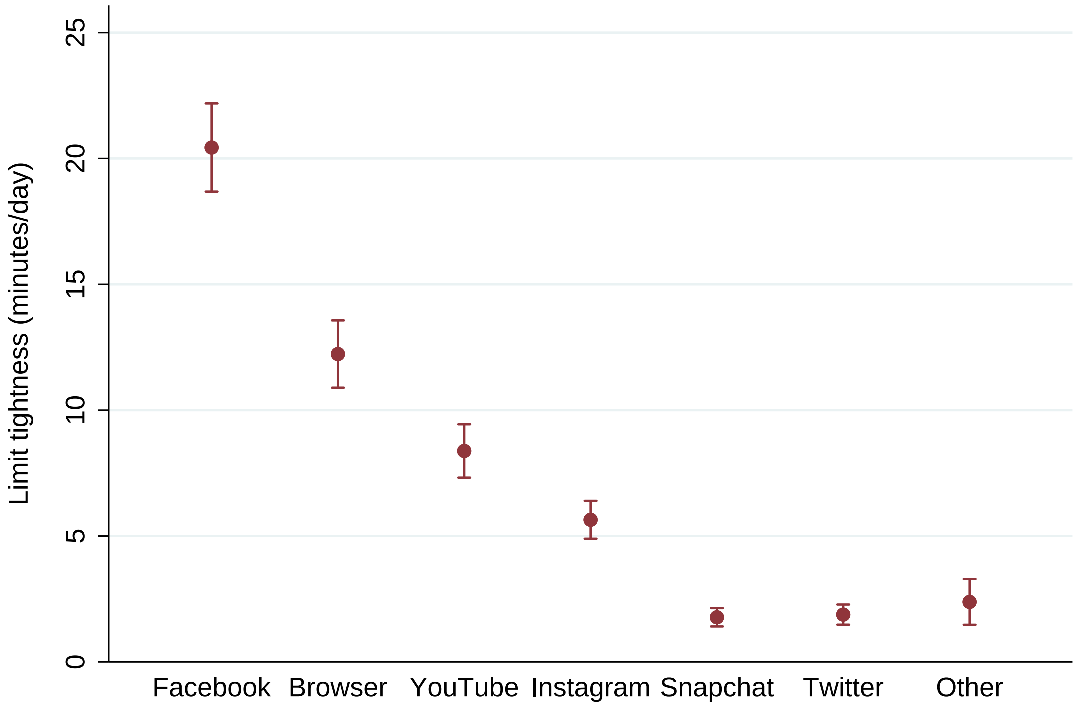
Notes: This figure presents effects of the bonus treatment on FITBSY use by week using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure A11: **Distribution of User-Level Limit Tightness**



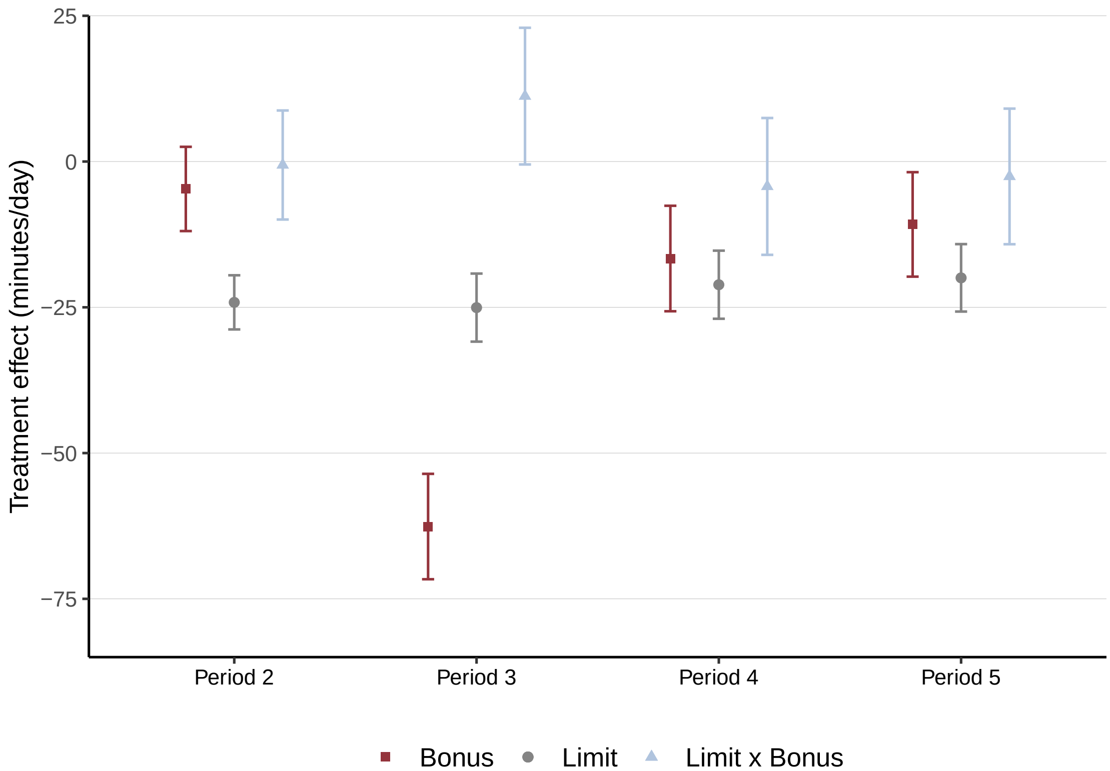
Notes: This figure presents mean *user-level limit tightness* over periods 2–5. *User-level limit tightness* is the amount by which a user's limits would have hypothetically reduced overall screen time if applied to their baseline use without snoozes; see equation (5).

Figure A12: **Average Limit Tightness by App**



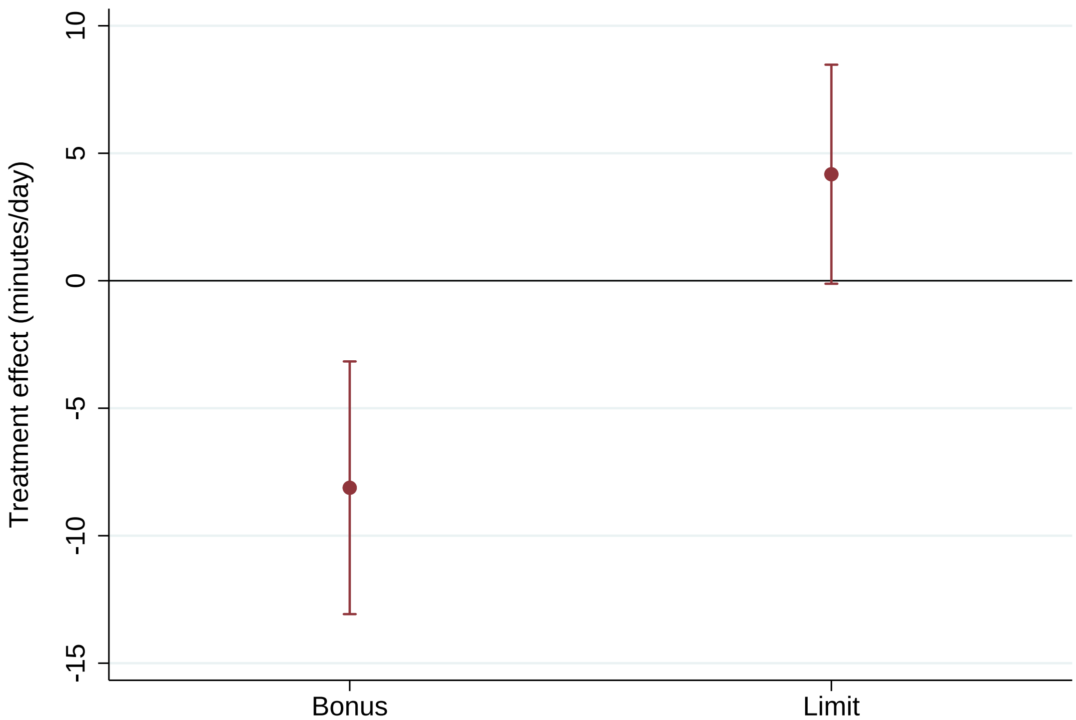
Notes: This figure presents average *limit tightness* by app over periods 2–5. *Limit tightness* is the amount by which a user's limits would have hypothetically reduced screen time if applied to their baseline use without snoozes; see equation (5). FITSBY apps are in order of decreasing period 1 use.

Figure A13: **Interaction Effects of Bonus and Limit by Period**



Notes: This figure presents effects of bonus and limit treatments on FITSBY use using equation (4) with an additional interaction term for participants in the intersection of the Limit and Bonus groups. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure A14: **Effects on Self-Reported FITSBY Use Change on Other Devices**



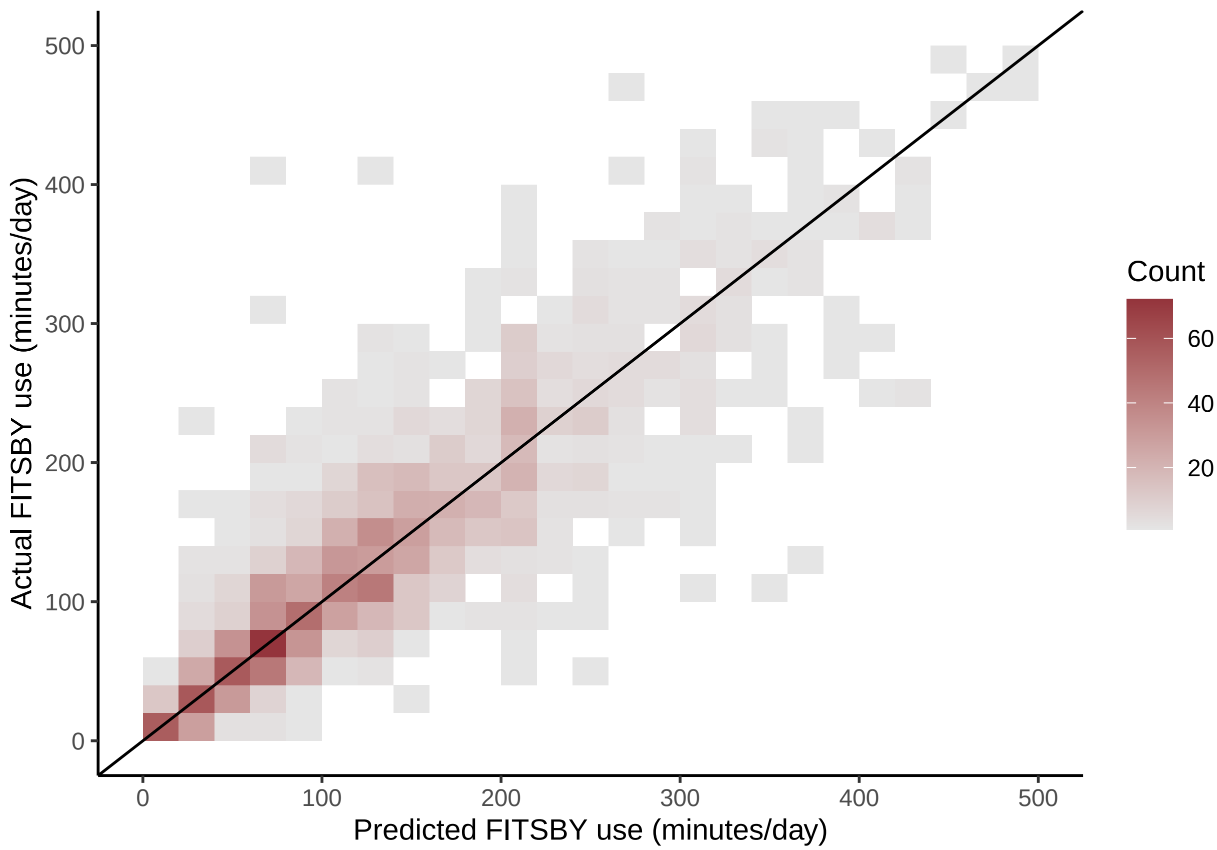
Notes: This figure presents the effects of bonus and limit treatments on self-reported change in FITSBY use on other devices relative to the three weeks before the study using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. Self-reported changes are winsorized at 150 minutes.

D.1 Validation of Predicted Use and Multiple Price List Responses

Predicted use lines up well with actual use; see Appendix Figures A15 and A16. The $5 (instead of $1) prediction accuracy reward slightly reduces the absolute value of the prediction error but has tightly estimated zero effects on predicted use, actual use, and the level of the prediction error; see Appendix Table A4.

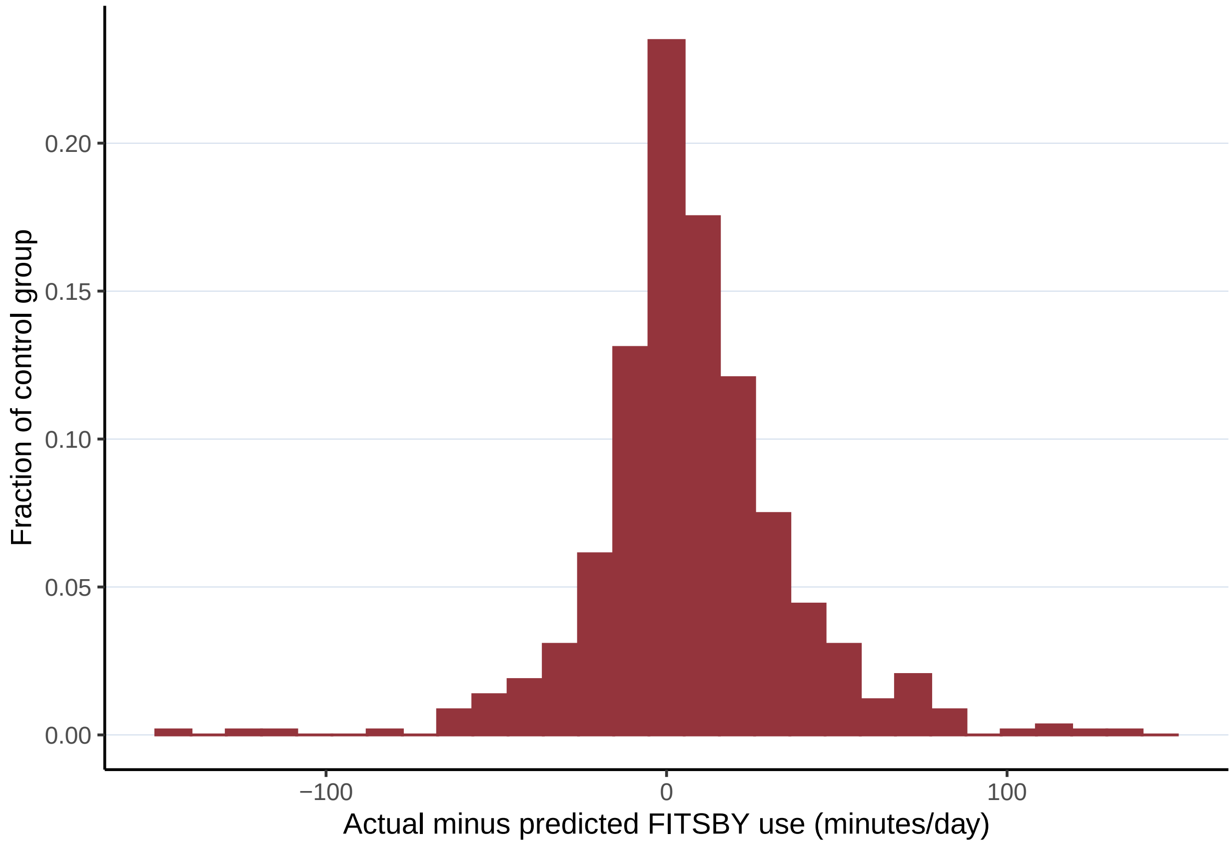
Multiple price lists are cognitively challenging, so we carry out several additional analyses to validate that these valuations are informative about people's preferences. First, participants' valuations of the bonus are correlated with the amount of money they could expect to earn; see Appendix Figure A19. Second, the limit valuation and the behavior change premium (defined in Section E.3) are correlated with each other and with *limit tightness*, *ideal use change*, *addiction scale*, *SMS addiction scale*, and other variables in expected ways; see Appendix Table A5. Third, after the bonus MPL, we asked people to "select the statement that best describes your thinking when trading off the Screen Time Bonus against the fixed payment." 24 percent responded that "I wanted to give myself an incentive to use my phone less over the next three weeks, even though it might result in a smaller payment," and this group had a substantially higher average behavior change premium; see Appendix Figures A20 and A21.

Figure A15: **Predicted vs. Actual FITSBY Use in Control**



Notes: This figure presents the number of Control group participants in each cell of actual and predicted FITSBY use across periods 2–4, using predictions from the survey just before each period. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure A16: **Histogram of Actual Minus Predicted FITSBY Use in Control Group**



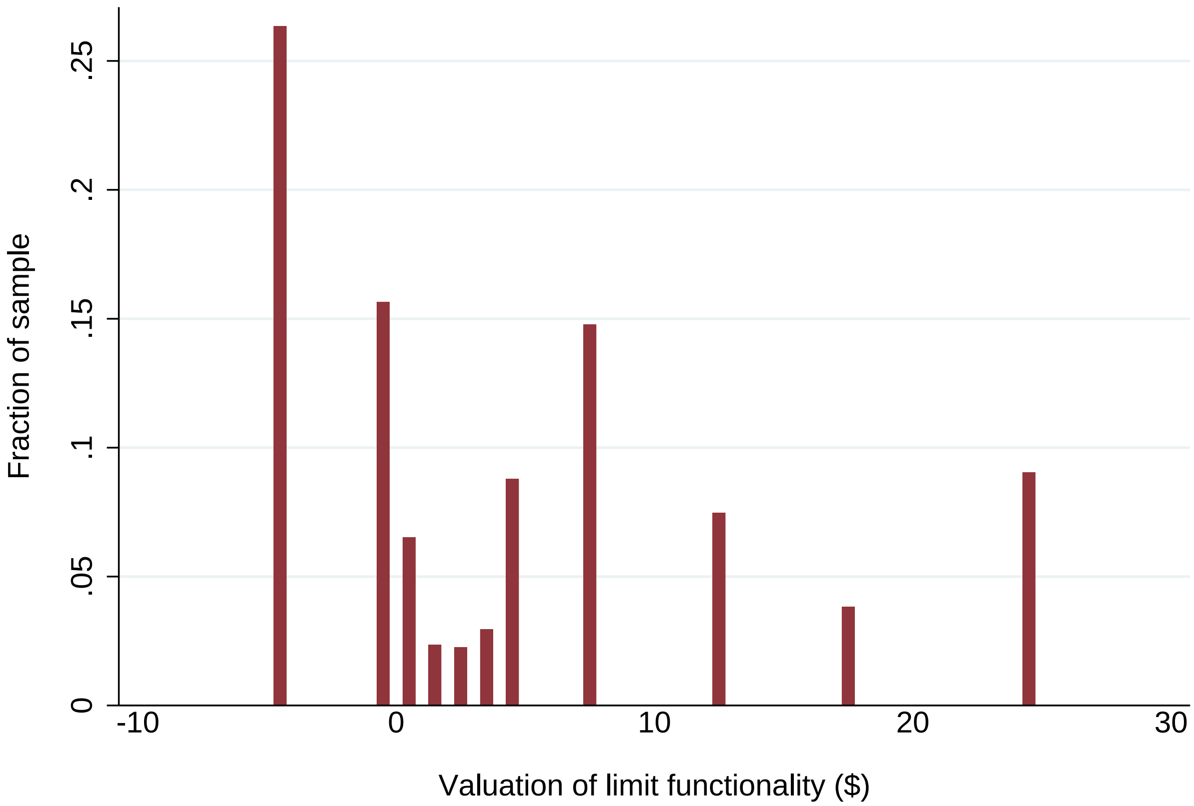
Notes: This figure presents the distribution of the difference between actual and predicted FITSBY use across periods 2–4 in the Control group, using predictions from the survey just before each period. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Table A4: **Effect of Prediction Accuracy Reward**

	(1)	(2)	(3)	(4)
	Predicted use	Actual use	Predicted - actual use	Absolute value of predicted - actual use
High prediction reward	1.219	3.343	-2.207	-2.379
	(2.582)	(2.386)	(1.691)	(1.435)
Constant	118.9	116.7	2.300	35.22
	(1.908)	(1.670)	(1.376)	(1.212)

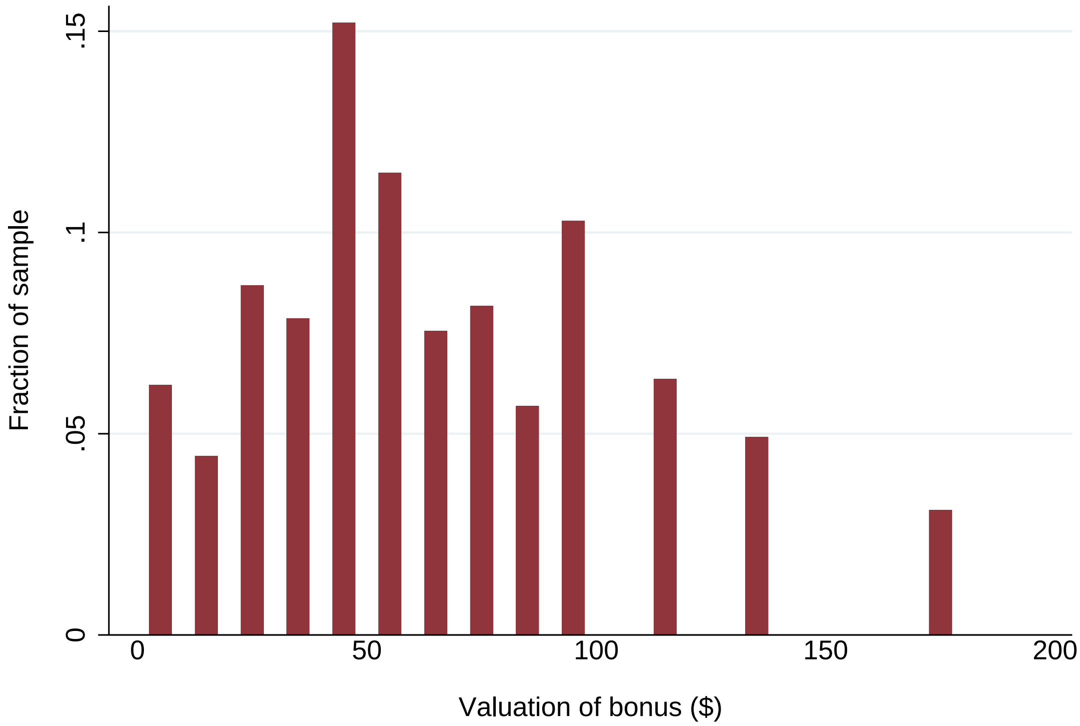
Notes: This table presents the effects of being offered the higher Prediction Reward ($5 instead of $1 for predicting within 15 minutes of actual screen time) on predicted and actual FITSBY use in minutes per day. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. Standard errors are in parentheses.

Figure A17: **Valuation of Limit Functionality**



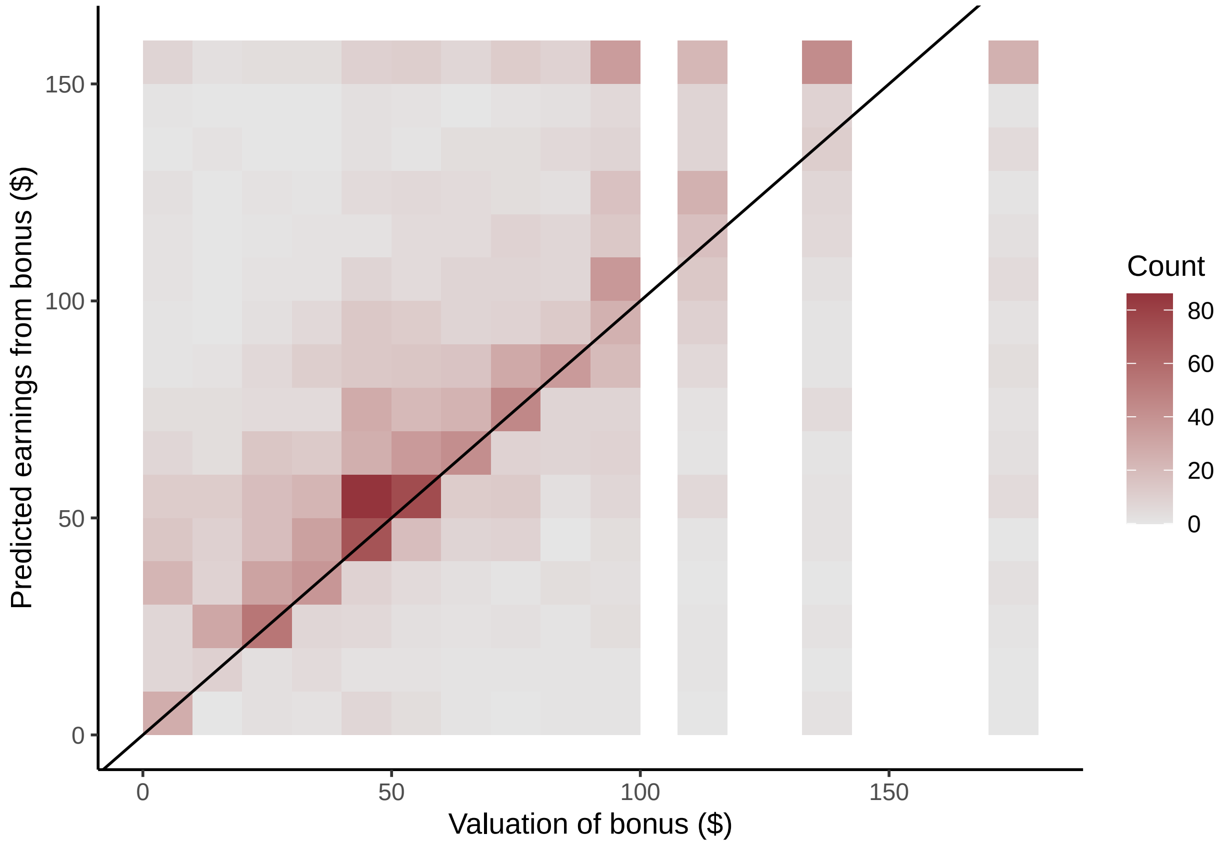
Notes: This figure presents the distribution of valuations of access to the limit functionality for the next three weeks, as elicited in a multiple price list on survey 3. Valuations above $20 are plotted at $25, and valuations below $-1 are plotted at $-5.

Figure A18: **Valuation of Screen Time Bonus**



Notes: This figure presents the distribution of valuations of the Screen Time Bonus incentive, as elicited on survey 2. Valuations above $150 are plotted at $175.

Figure A19: **Valuation of Bonus vs. Predicted Bonus Earnings**



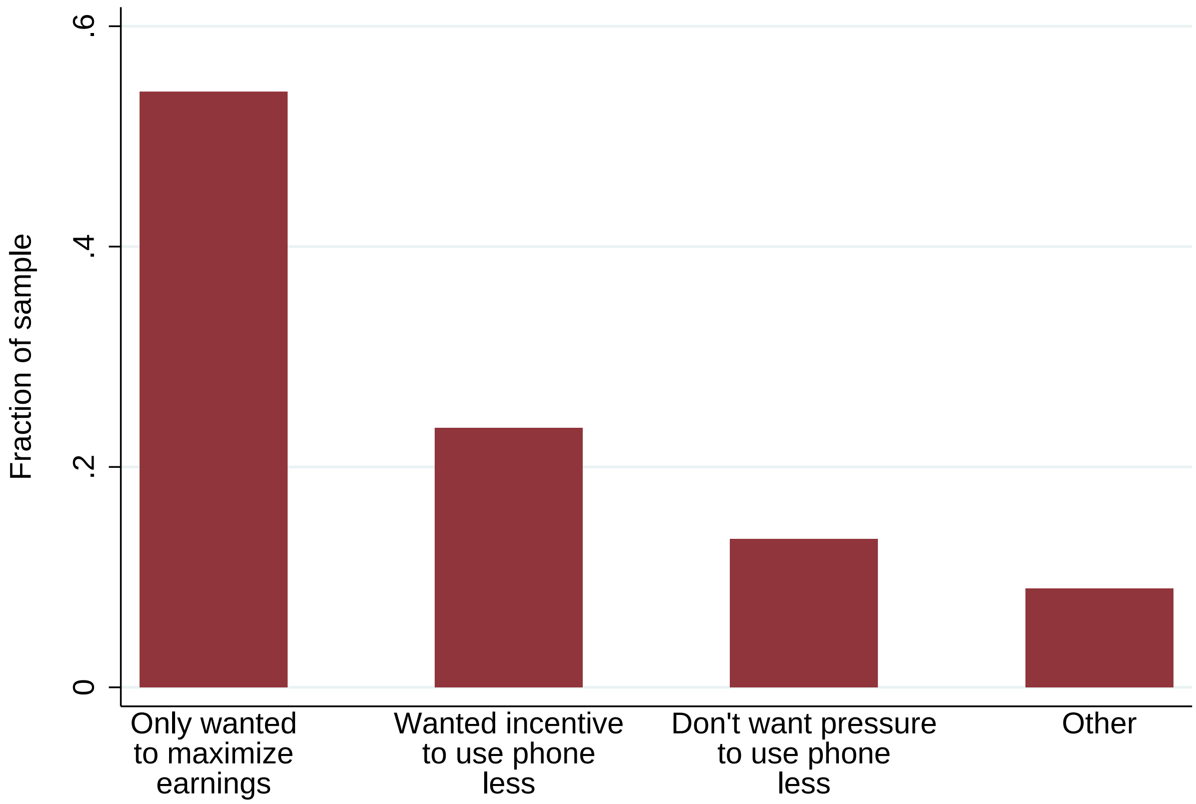
Notes: This figure presents the number of participants in each cell of predicted earnings from the Screen Time Bonus (given the participant's Bonus Benchmark and predicted FITSBY use) and valuation of the bonus, as elicited on survey 2.

Table A5: **Correlations between Temptation and Addiction Measures**

	Behavior change premium	Valuation of limit	Limit tightness	Interest in limits	Ideal use change × (-1)	Addiction scale	SMS addiction scale	Phone makes life better × (-1)
Behavior change premium	1							
Valuation of limit	0.116	1						
Limit tightness	0.471	0.199	1					
Interest in limits	0.032	0.146	0.204	1				
Ideal use change x (-1)	0.117	0.112	0.218	0.319	1			
Addiction scale	0.267	0.078	0.243	0.356	0.435	1		
SMS addiction scale	0.272	0.132	0.259	0.312	0.345	0.651	1	
Phone makes life better x (-1)	0.022	0.082	0.154	0.295	0.392	0.303	0.234	1

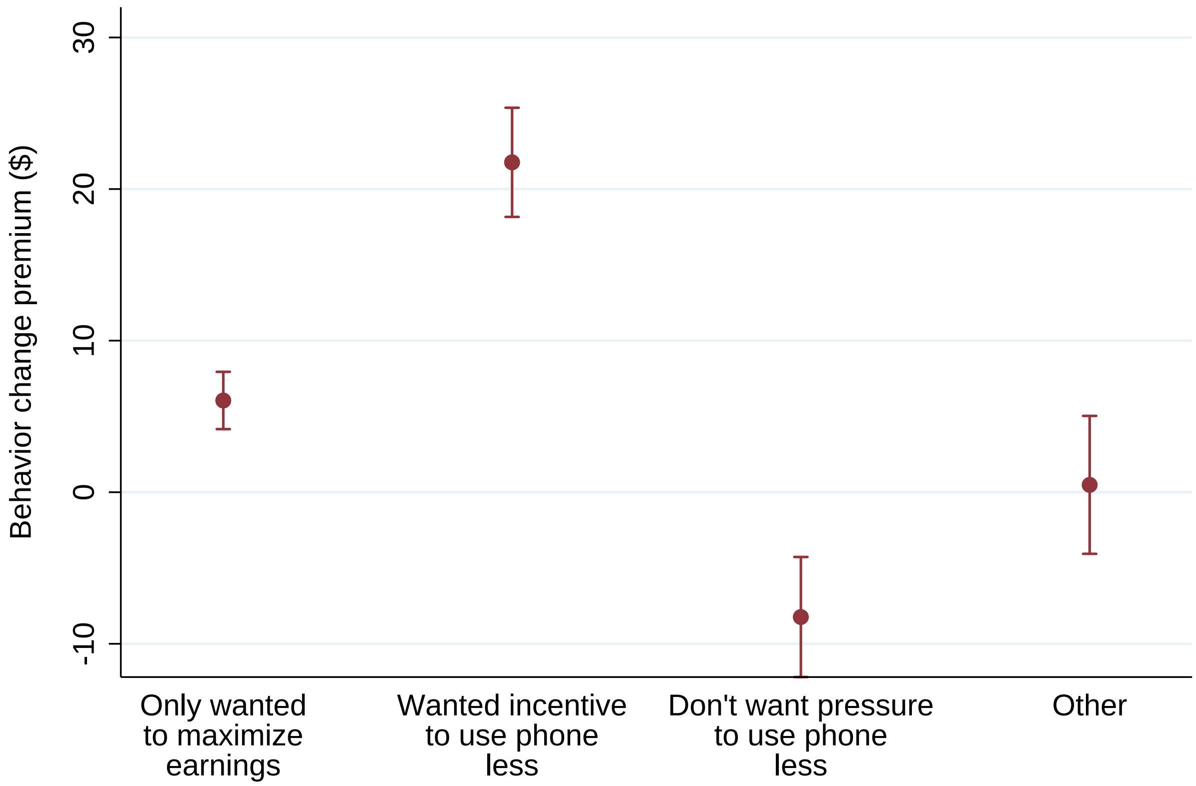
Note: The behavior change premium is the difference between the valuation of the Screen Time Bonus and the modeled valuation if the consumer believed herself to be time consistent. *Interest in limits*, *ideal use change*, *addiction scale*, *SMS addiction scale*, and *phone makes life better* are from survey 1.

Figure A20: **Reported Reasoning on Screen Time Bonus Multiple Price List**



Notes: After the bonus multiple price list, survey 2 asked participants to "select the statement that best describes your thinking when trading off the Screen Time Bonus against the fixed payment." This figure presents the share of participants who selected each answer.

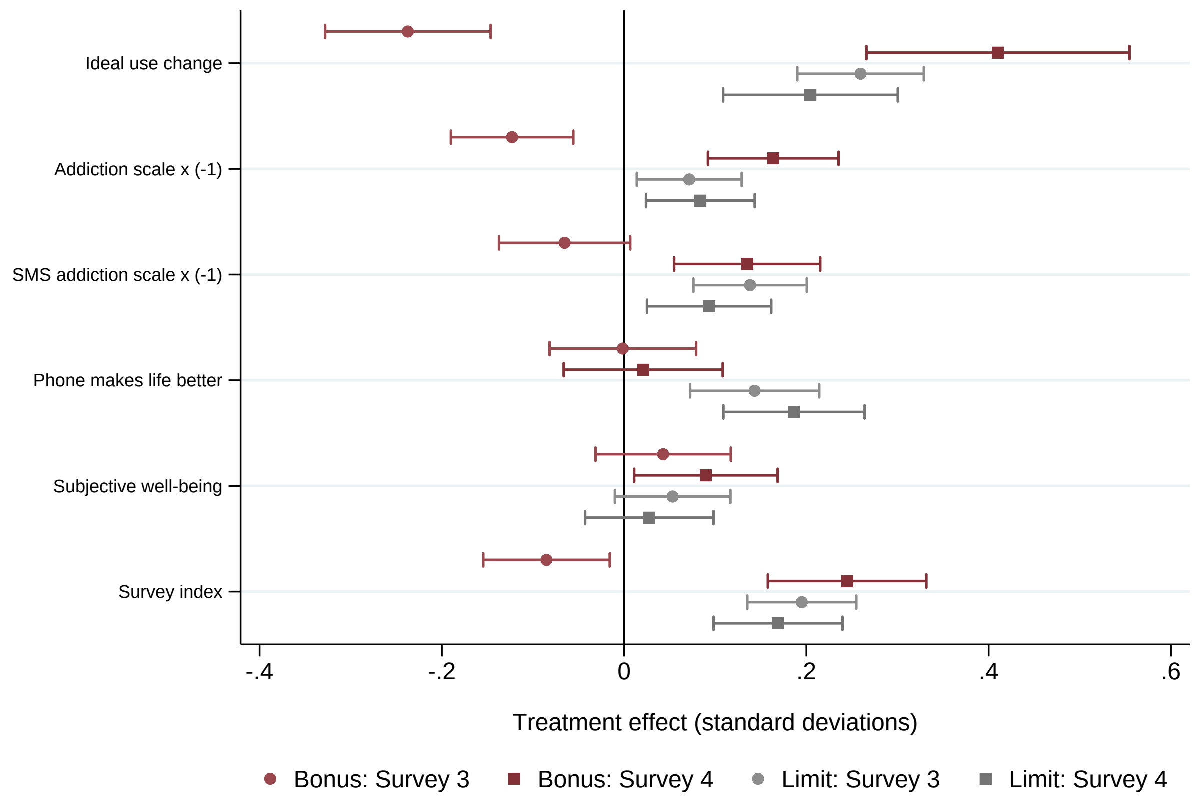
Figure A21: **Behavior Change Premium by Reported Reasoning**



Notes: The behavior change premium is the difference between the valuation of the Screen Time Bonus and the modeled valuation if the consumer believed herself to be time consistent. After the bonus multiple price list, survey 2 asked participants to "select the statement that best describes your thinking when trading off the Screen Time Bonus against the fixed payment." This figure presents means and 95 percent confidence intervals of the behavior change premium by responses to that question.

D.2 Additional Estimates of Effects on Survey Outcome Variables

Figure A22: **Effects of Limits and Bonus on Survey Outcomes on Surveys 3 and 4**



Notes: This figure presents effects of the bonus and limit treatment on survey outcome variables using equation (4), allowing separate coefficients for effects on surveys 3 vs. 4. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Table A6: **Treatment Effects**

(a) **Bonus**

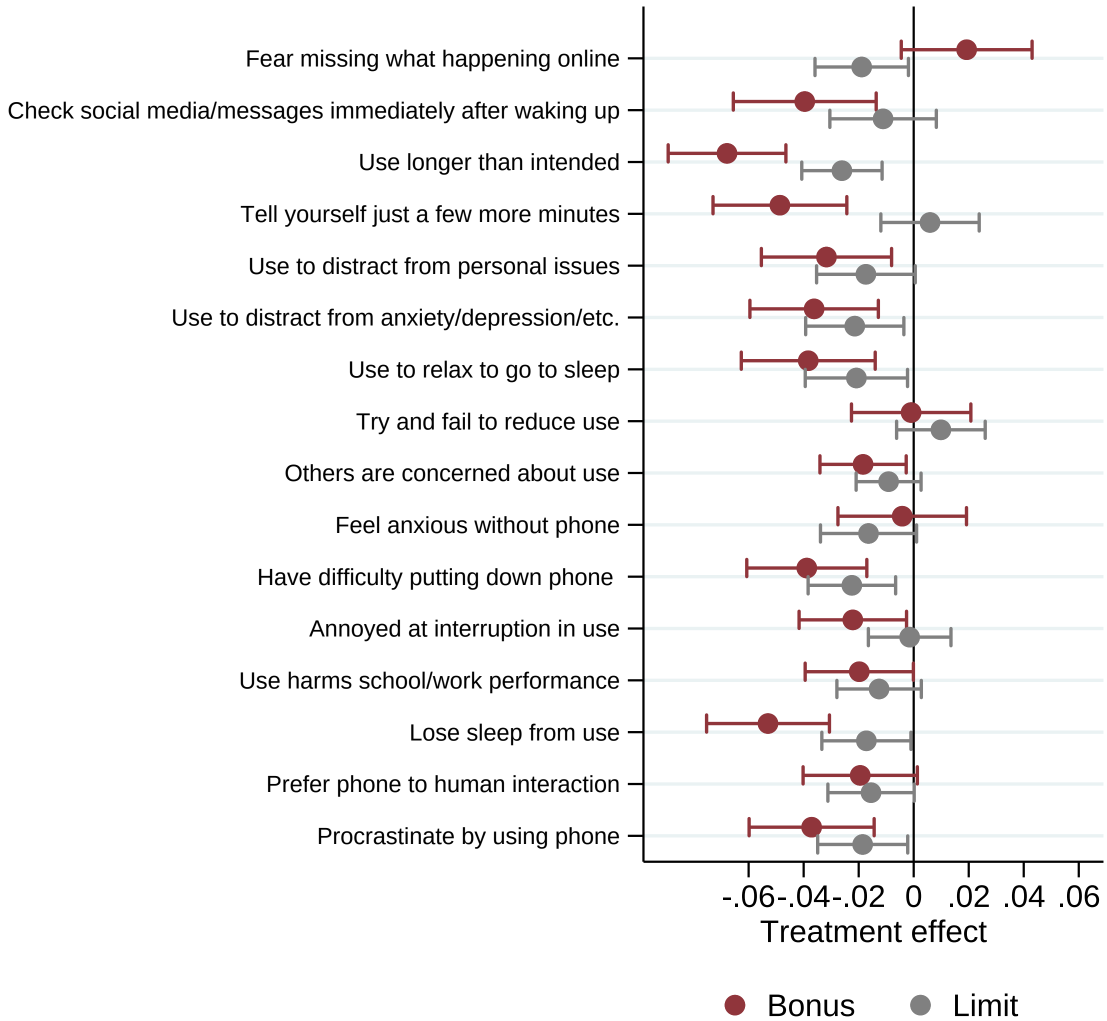
	(1)	(2)	(3)	(4)	(5)	(6)
	Treatment effect (original units)	Standard error (original units)	Treatment effect (SD units)	Standard error (SD units)	P-value	Sharpened FDR-adjusted q-value
Ideal use change	9.0	1.6	0.41	0.074	0.000	0.000
Addiction scale x (-1)	0.44	0.10	0.16	0.037	0.000	0.000
SMS addiction scale x (-1)	0.42	0.13	0.14	0.041	0.001	0.004
Phone makes life better	0.042	0.090	0.021	0.045	0.64	0.78
Subjective well-being	0.23	0.10	0.090	0.040	0.026	0.09
Survey index	0.17	0.031	0.24	0.044	0.000	0.000

(b) **Limit**

	(1)	(2)	(3)	(4)	(5)	(6)
	Treatment effect (original units)	Standard error (original units)	Treatment effect (SD units)	Standard error (SD units)	P-value	Sharpened FDR-adjusted q-value
Ideal use change	5.1	0.75	0.23	0.034	0.000	0.000
Addiction scale x (-1)	0.21	0.071	0.078	0.027	0.004	0.008
SMS addiction scale x (-1)	0.36	0.090	0.12	0.028	0.000	0.000
Phone makes life better	0.33	0.064	0.16	0.032	0.000	0.000
Subjective well-being	0.10	0.075	0.040	0.030	0.18	0.24
Survey index	0.13	0.020	0.18	0.029	0.000	0.000

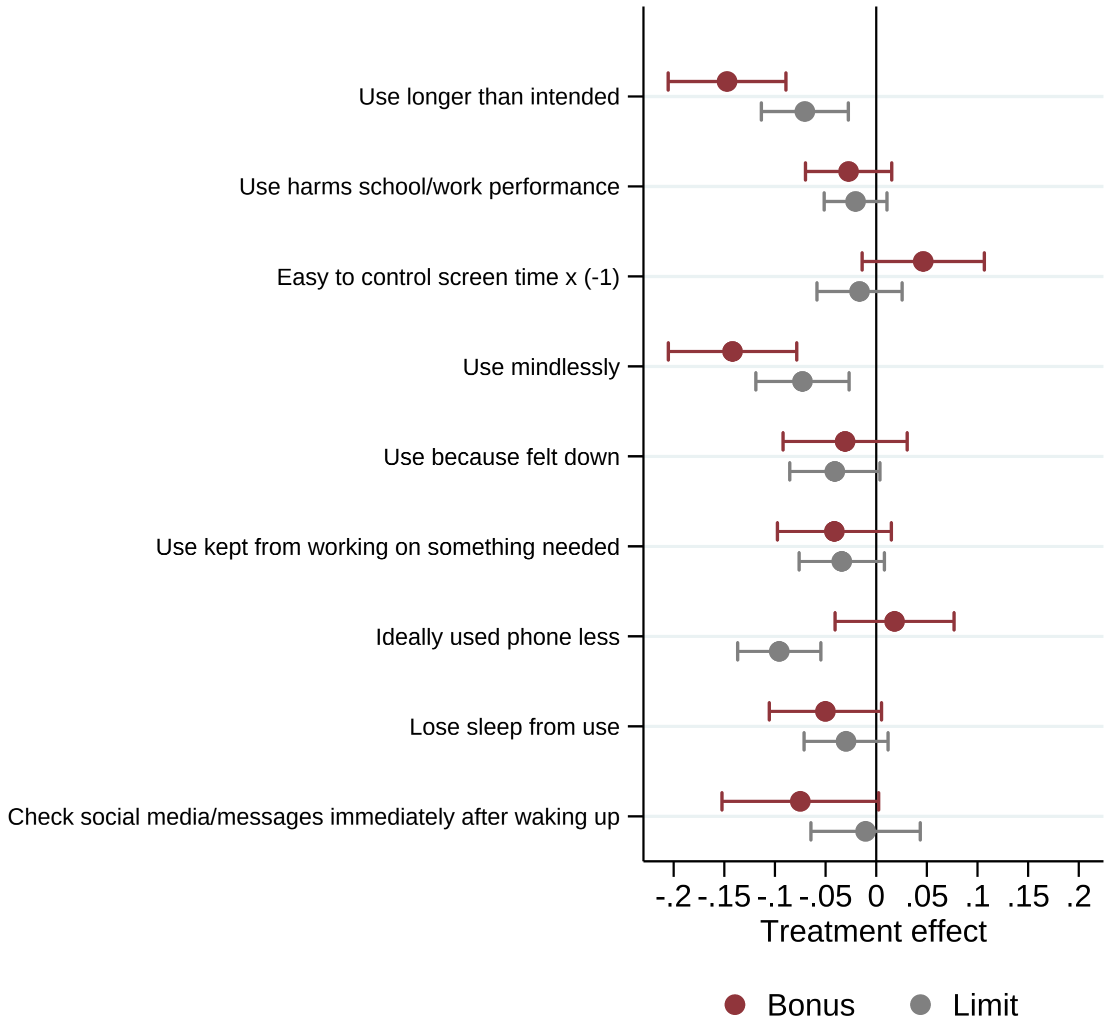
Notes: This table presents effects of the bonus and limit treatments on survey outcome variables using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline. The effects in standard deviation units in column 3 match those reported on Figure 8.

Figure A23: **Effects on Addiction Responses**



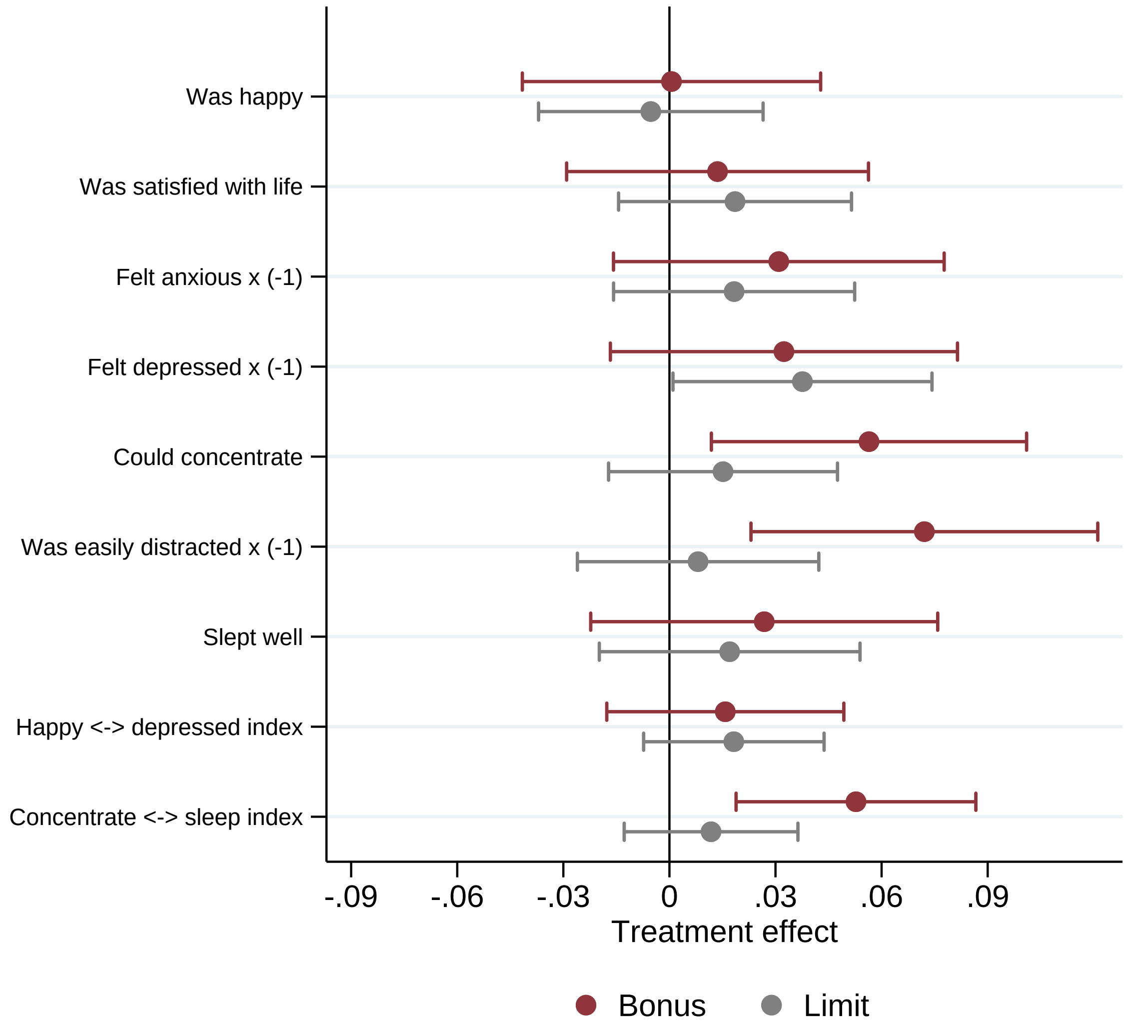
Notes: This figure presents the effects of the bonus and limit treatments on individual items in the *addiction scale* variable using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. The direction of the effects in this figure are opposite those in the main figures, because *addiction scale* is multiplied by -1 in those figures.

Figure A24: **Effects on SMS Addiction Responses**



Notes: This figure presents the effects of the bonus and limit treatments on individual items in the *SMS addiction scale* variable using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. The direction of the effects in this figure are opposite those in the main figures, because *SMS addiction scale* is multiplied by -1 in those figures.

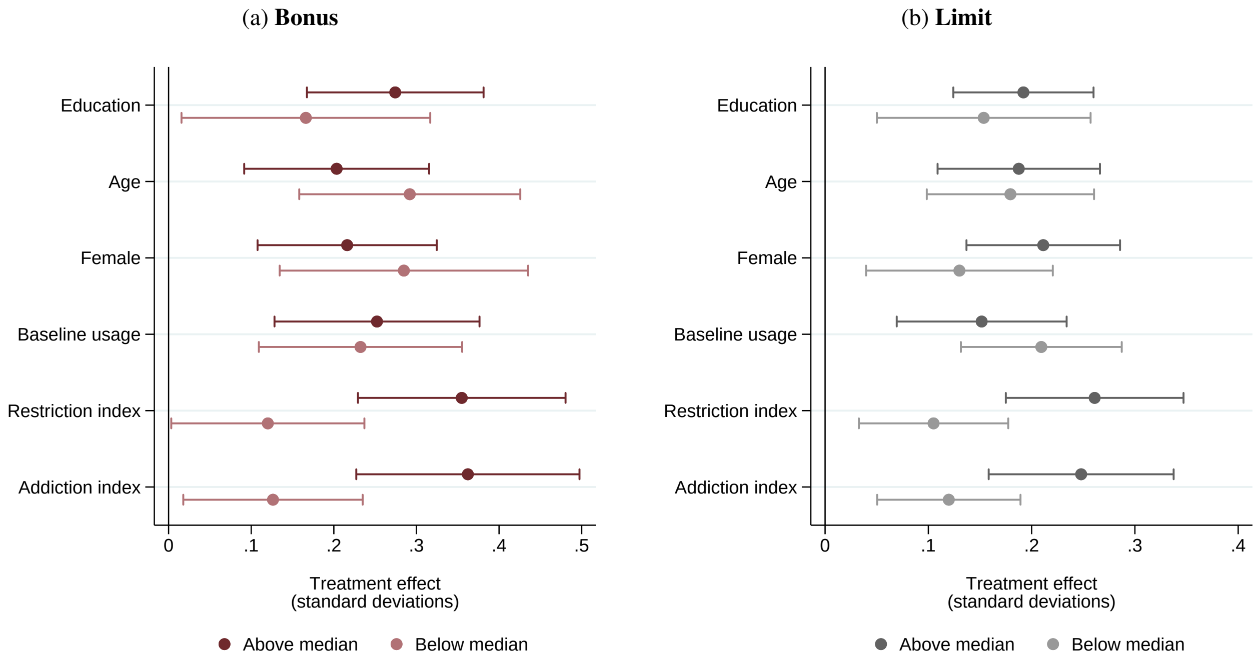
Figure A25: **Effects on Subjective Well-Being Responses**



Notes: This figure presents the effects of the bonus and limit treatments on individual items in the *subjective well-being* variable using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4.

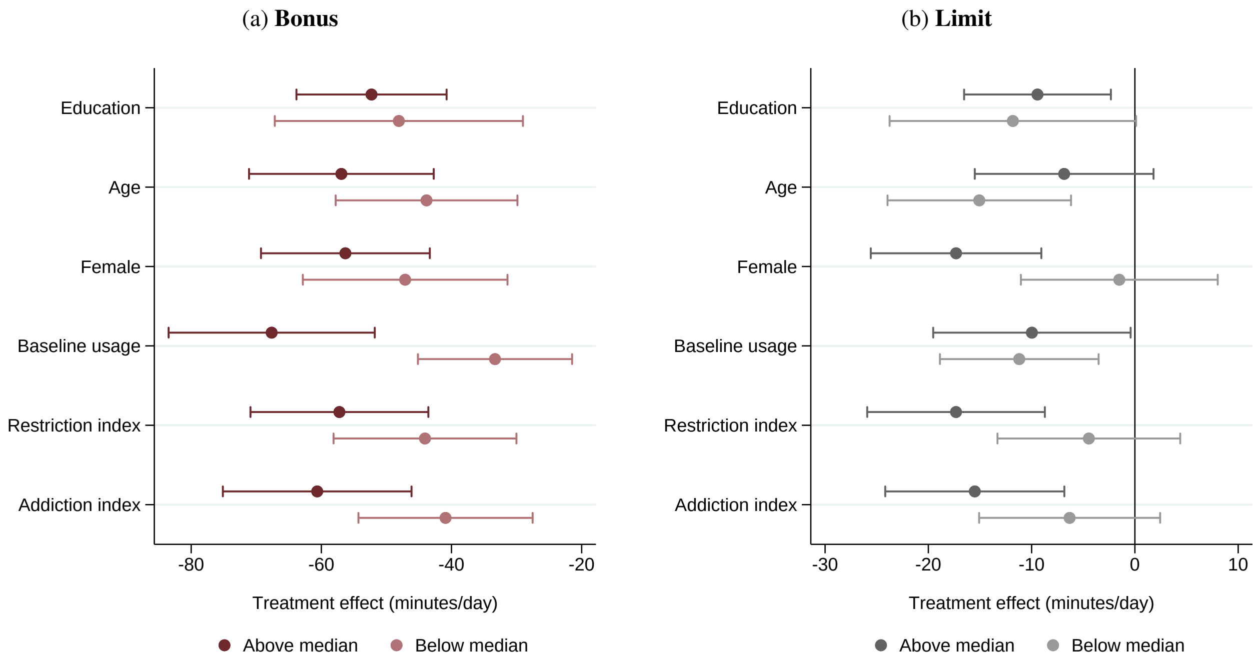
D.3 Heterogeneous Treatment Effects

Figure A26: **Heterogeneous Effects of Limits and Bonus on Survey Index**



Notes: This figure presents heterogeneous effects of the bonus and limit treatments on *survey index*, the inverse-covariance weighted average of five measures of smartphone addiction and subjective well-being, using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. Above-median education includes people with a college degree or more, above-median age includes people 30 and older, and median baseline FITSBY use is 137 minutes per day. *Restriction index* is a combination of *interest in limits* and *ideal use change*. *Addiction index* is a combination of *addiction scale* and *phone makes life better*.

Figure A27: **Heterogeneous Effects of Limits and Bonus on FITSBY Use**



Notes: This figure presents heterogeneous effects of the bonus and limit treatments on FITSBY use using equation (4). The bonus effects are measured in period 3, while the limit effects are measured in periods 2–5. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. Above-median education includes people with a college degree or more, above-median age includes people 30 and older, and median baseline FITSBY use is 137 minutes per day. *Restriction index* is a combination of *interest in limits* and *ideal use change*. *Addiction index* is a combination of *addiction scale* and *phone makes life better*.

D.4 Local Average Treatment Effects on Survey Outcomes

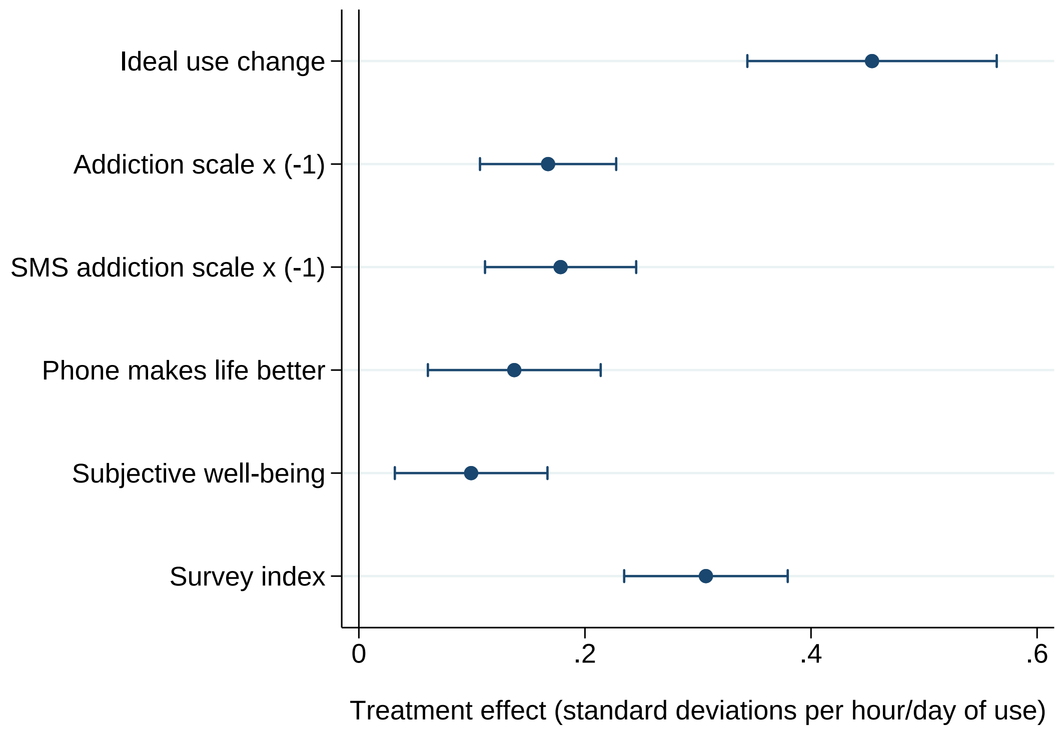
Our pre-analysis plan specified that we would also estimate instrumental variables (IV) regressions with previous period FITSBY use $x_{i,t-1}$ as the endogenous variable:

$$Y_{it} = \tau x_{i,t-1} + \beta_t \boldsymbol{X}_{i1} + \nu_{it} + \varepsilon_i, \tag{17}$$

instrumenting for $x_{i,t-1}$ with B_i and L_i interacted with $t = 3$ and $t = 4$ indicators. We combine data from surveys 3 and 4 and let all coefficients other than τ vary across the two periods. Conceptually, this regression combines the effects of the bonus and limit intervention, weighting the interventions by their effects on FITSBY use. Because the limit treatment could affect survey outcomes through channels other than reduced FITSBY use—for example, by giving people an increased feeling of control over their screen time—we do not claim that the IV exclusion restriction necessarily holds.

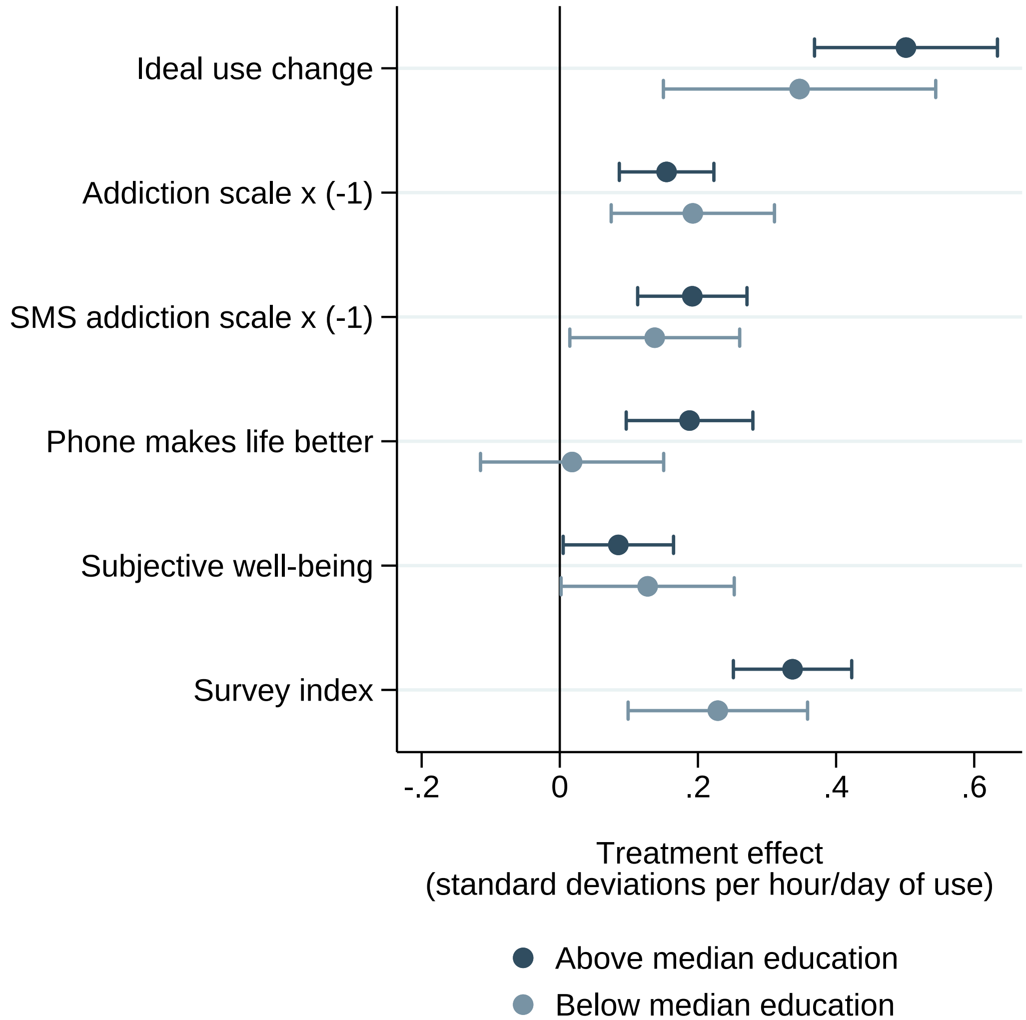
Appendix Figure A28 presents local average treatment effects estimated using equation (17), combining effects from both treatments. Appendix Figures A29–A34 study heterogeneity along the six pre-specified moderators. The results are qualitatively similar to Figures 8 and A26, except that the estimates are slightly more precise, as would be expected from combining effects of two interventions. Note that since the average effects of both interventions are about the same for people with low versus high baseline use (Figure A26), the local average treatment effects of reduced use are much larger for people with low baseline use (Appendix Figure A32).

Figure A28: **Local Average Treatment Effects of FITSBY Use on Survey Outcome Variables**



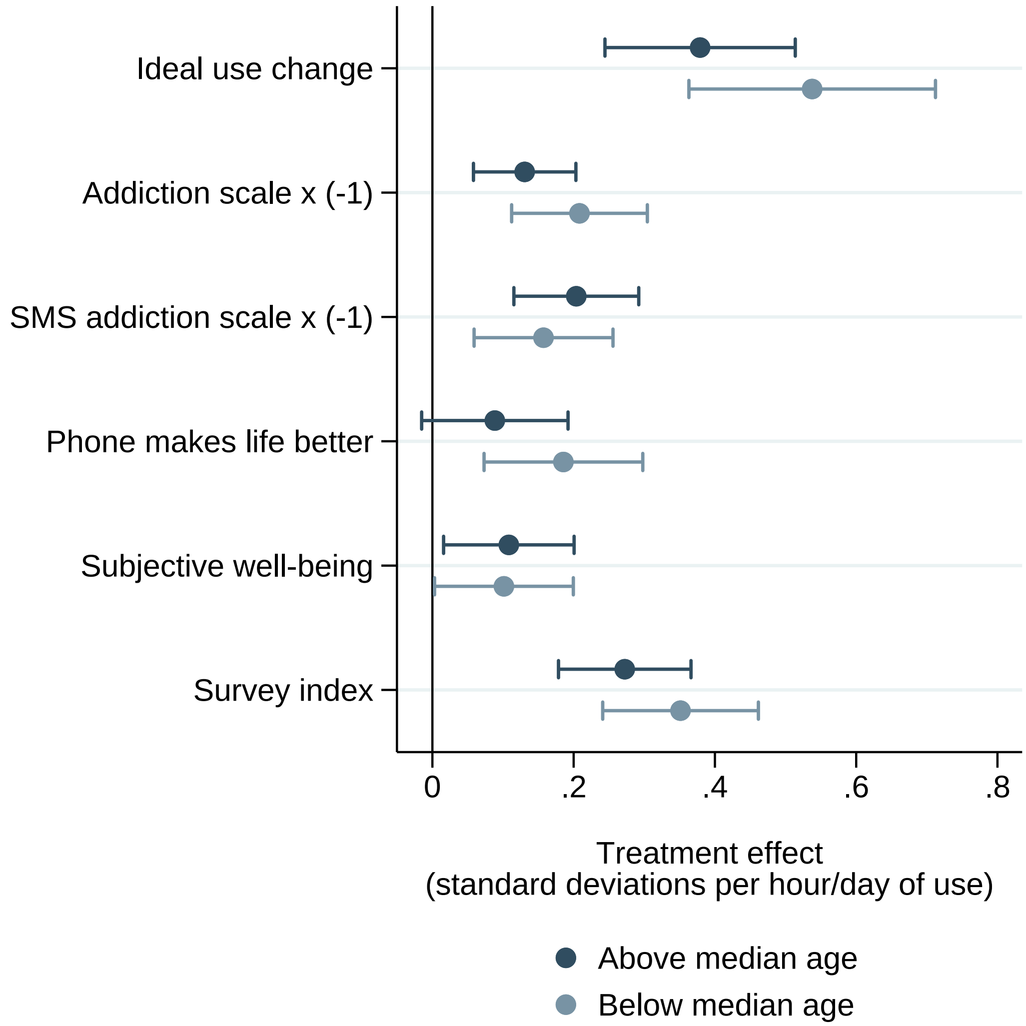
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17). We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, “Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?” *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, “To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?” *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A29: **Heterogeneous Effects on Survey Outcome Variables by Education**



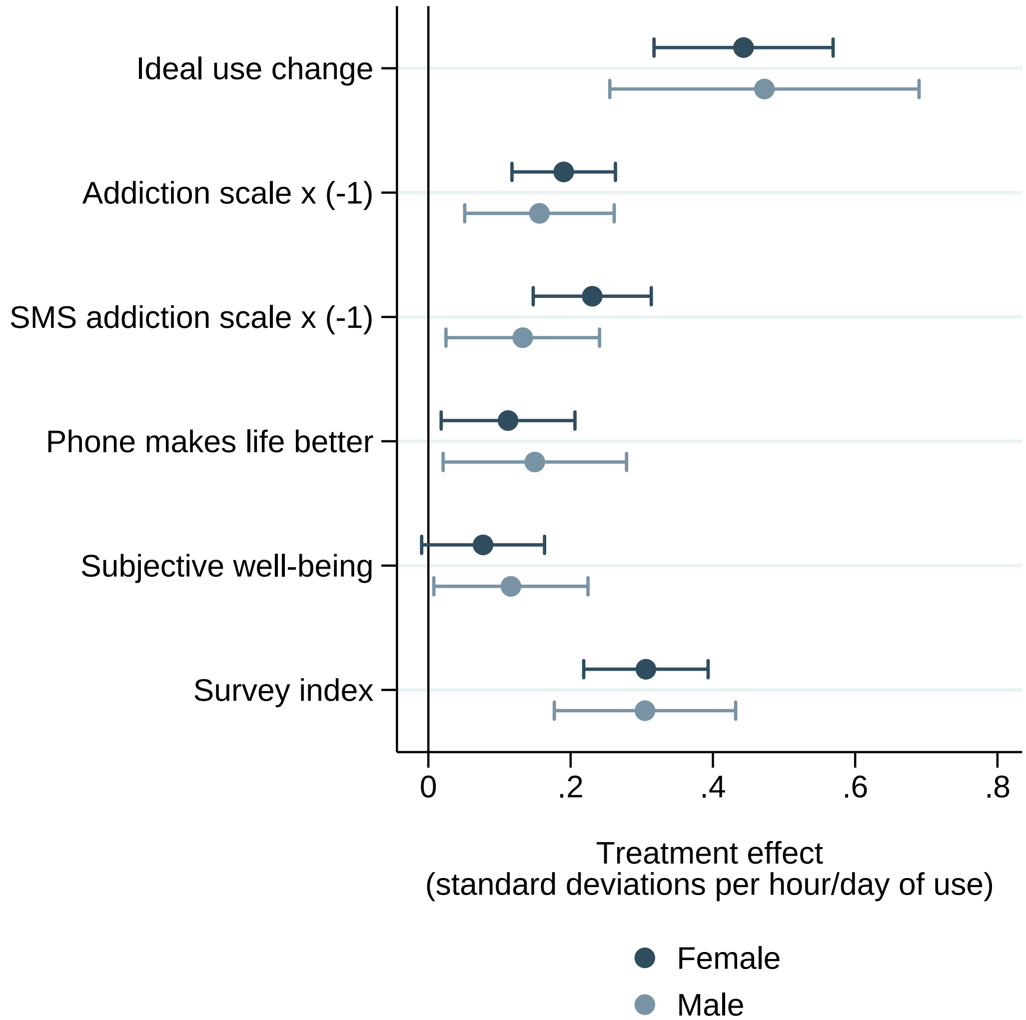
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median education. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A30: **Heterogeneous Effects on Survey Outcome Variables by Age**



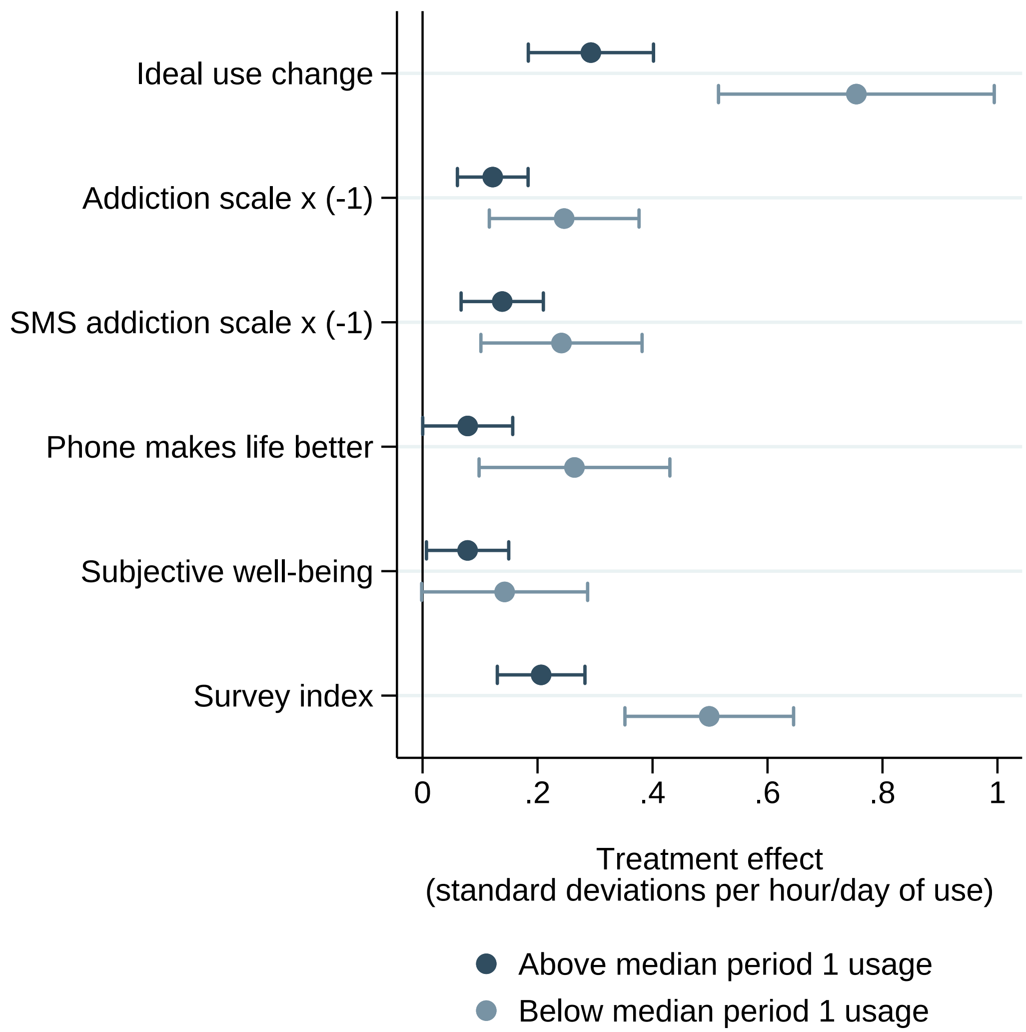
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median age. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A31: **Heterogeneous Effects on Survey Outcome Variables by Gender**



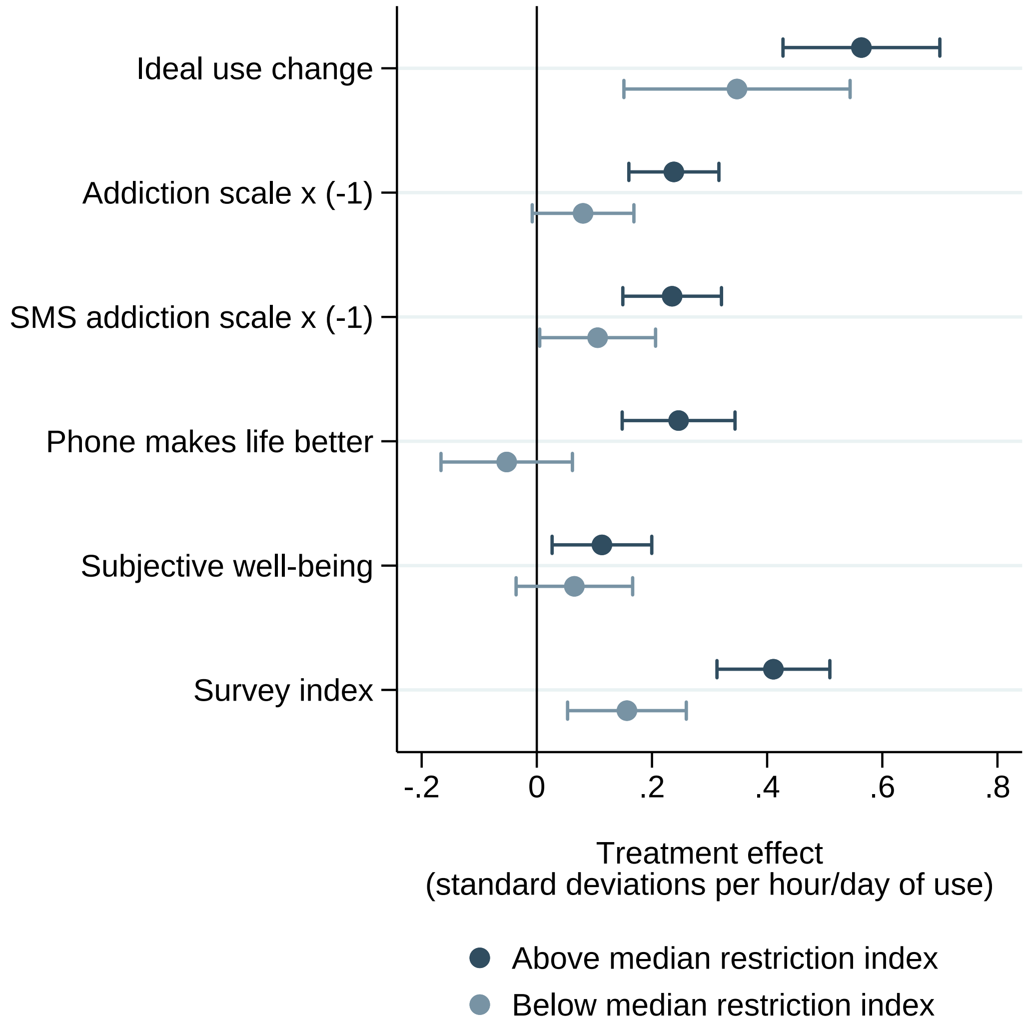
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for men versus women. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A32: **Heterogeneous Effects on Survey Outcome Variables by Baseline FITSBY Use**



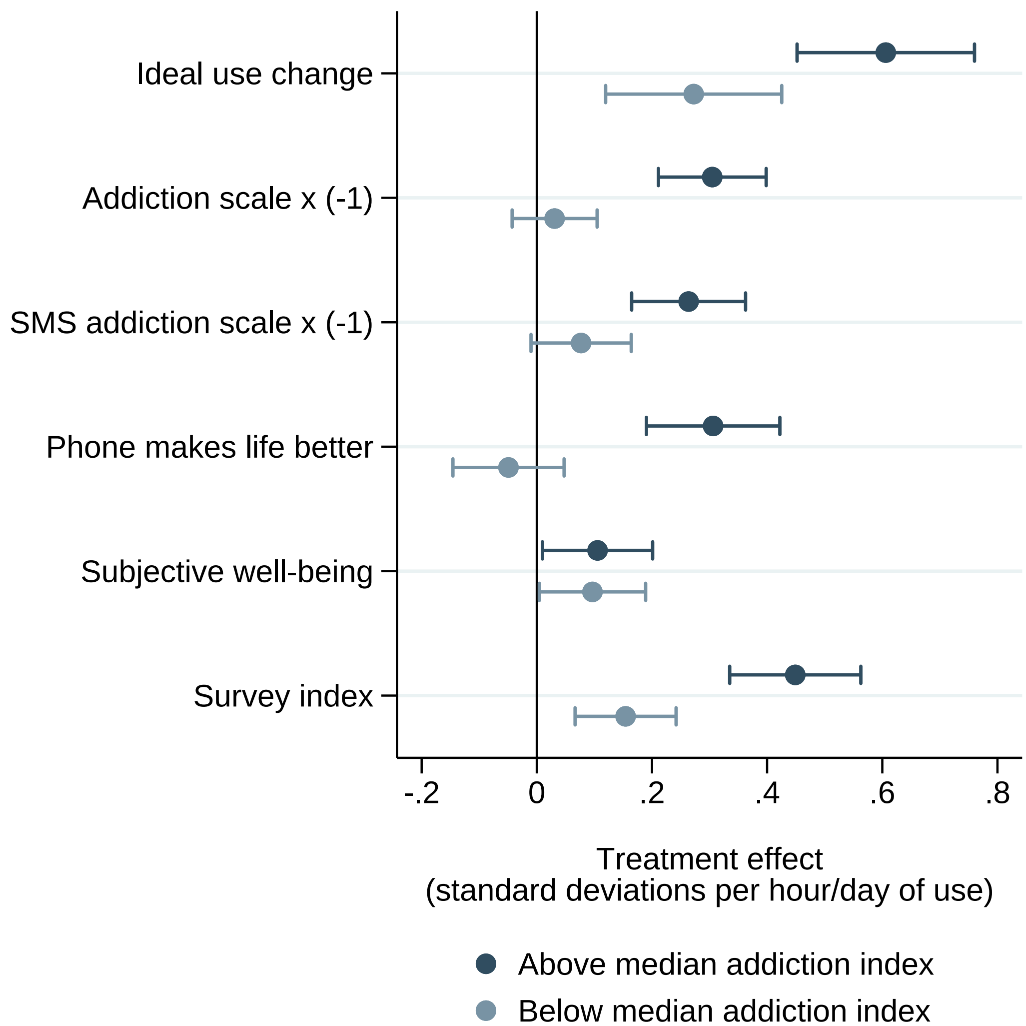
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median baseline FITSBY use. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A33: **Heterogeneous Effects on Survey Outcome Variables by Restriction Index**



Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median values of *restriction index*, a combination of *interest in limits* and *ideal use change*. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A34: **Heterogeneous Effects on Survey Outcome Variables by Addiction Index**



Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median values of *addiction index*, a combination of *addiction scale* and *phone makes life better*. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

E Unrestricted Model and Alternative Temptation Estimates

In this appendix, we estimate the unrestricted model and present alternative estimates of the temptation parameter γ.

E.1 Key Theoretical Results

Three theoretical results are key to our estimation strategy: the Euler equation, linear policy functions, and the steady state.

Euler equation. The first-order conditions of equation (2) for periods t and $t+1$ can be re-arranged into an Euler equation characterizing the equilibrium relationship between consumption in periods t and $t+1$. To simplify notation, define $u_t := u_t(x_t^*; s_t, p_t)$ as current utility, define $\tilde{x}_r := \tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, \boldsymbol{p}_r)$ and $\tilde{u}_r := u_r(\tilde{x}_r; \tilde{s}_r, p_r)$ as predicted consumption and utility for future periods $r > t$, and define $\tilde{\lambda}_r := \frac{\partial \tilde{x}_r}{\partial \tilde{s}_r}$ as the predicted effect of habit stock on consumption.

Proposition 1. *Suppose $u_t(x_t; s_t, p_t)$ is given by equation (3) and $(x_0^*, ..., x_T^*)$ is a perception-perfect strategy profile with differentiable strategies. Then for each $t < T$,*

$$\underbrace{\eta x_t^* + \zeta s_t + \xi_t - p_t}_{\partial u_t/\partial x_t} + \gamma = (1-\alpha)\delta\rho\left[\underbrace{\eta \tilde{x}_{t+1} + \zeta \tilde{s}_{t+1} + \xi_{t+1} - p_{t+1}}_{\partial \tilde{u}_{t+1}/\partial \tilde{x}_{t+1}} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \underbrace{(\zeta \tilde{x}_{t+1} + \phi)}_{\partial \tilde{u}_{t+1}/\partial \tilde{s}_{t+1}}\right] \quad . \tag{18}$$

Proof. See Appendix F.1.

□

With full myopia ($\delta = 0$) or full projection bias ($\alpha = 1$), consumers maximize current-period flow utility, setting the left-hand side of equation (18) to zero. In a "rational" habit formation model with $\alpha = 0$ and $\tilde{\gamma} = \gamma = 0$, the right-hand side adds two effects. First, there is an adjacent complementarity effect where people consume more in period t (driving down marginal utility $\partial u_t/\partial x_t$) if they expect to consume more in $t+1$ (i.e. if future marginal utility $\partial \tilde{u}_{t+1}/\partial x_{t+1}$ is lower). Second, there is a direct habit stock effect where people consume more in period t if the marginal utility from the resulting habit stock $\partial \tilde{u}_{t+1}/\partial s_{t+1}$ is higher.

Temptation adds two forces. First, the balance of the adjacent complementarity effect tilts toward increased consumption, as γ is added to period t marginal utility and $\tilde{\gamma}$ is added to predicted period $t+1$ marginal utility. Second, people reduce current consumption to avoid exacerbating perceived future overconsumption, giving $\tilde{\gamma}\lambda_{t+1}$ on the right-hand side.

Linear policy functions. With quadratic flow utility, equilibrium consumption is linear in habit stock with slope λ_t, and equilibrium predicted consumption is linear in habit stock with slope $\tilde{\lambda}_t$. Furthermore, if the consumer's objective function is concave, λ and $\tilde{\lambda}$ are constant far from the time horizon. This argument follows Gruber and Köszegi (2001).

Proposition 2. *Suppose the conditions for Proposition 1 hold. Then for any t,*

$$x_t^*(s_t,\gamma,\boldsymbol{p}_t) = \lambda_t s_t + \mu_t(\gamma) \tag{19}$$

$$\tilde{x}_t^*(s_t,\tilde{\gamma},\boldsymbol{p}_t) = \tilde{\lambda}_t s_t + \mu_t(\tilde{\gamma}), \tag{20}$$

where λ_t *is a function of only* $\{\eta,\zeta,\delta,\rho,\alpha\}$, $\tilde{\lambda}_t$ *is a function of only* $\{\eta,\zeta,\delta,\rho\}$, *and* μ_t *is linear in* p_t. *Furthermore, if the objective function from equation (2) is concave, then* $\lim_{T\to\infty}\lambda_t = \lambda$ *and* $\lim_{T\to\infty}\tilde{\lambda}_t = \tilde{\lambda}$ *for any fixed* t. *Finally,* $\lim_{T\to\infty}\mu_t = \mu$ *for any fixed* t *if* p_t *and* ξ_t *are constant and* $-\eta > (1-\alpha)\delta\rho\left[(\zeta-\eta)\left(1+\rho\tilde{\lambda}_{t+1}\right)-\rho\zeta\right]$.

Proof. See Appendix F.2. That appendix also provides an explicit condition that guarantees concavity. □

Steady state. Over a period of time when strategies are well approximated by the limiting values λ and μ, consumption converges to a steady state.

Lemma 1. *Suppose that strategies in all periods take the form* $x_t^*(s_t,\gamma,\boldsymbol{p}_t) = \lambda s_t + \mu$, *where* λ *and* μ *are constant. If* $\rho(1+\lambda) < 1$, *both* x_t^* *and* s_t *converge monotonically over time to steady-state values* x_{ss} *and* s_{ss}.

Proof. See Appendix F.3. □

If consumption has reached a steady state, we can use the Euler equation to characterize its level in closed form.

Proposition 3. *Suppose that* p_t *and* ξ_t *are constant and that consumption and habit stock are in steady state with* $s_t = s_{ss}$, $x_t = x_{ss}$, *and* $x_{ss} = \rho(s_{ss}+x_{ss})$. *Then consumption can be written as*

$$x_{ss} = \frac{\kappa - (1-(1-\alpha)\delta\rho)\,p + (1-\alpha)\delta\rho\left[(\zeta-\eta)m_{ss} - \left(1+\tilde{\lambda}\right)\tilde{\gamma}\right] + \gamma}{-\eta - (1-\alpha)\delta\rho(\zeta-\eta) - \zeta\frac{\rho-(1-\alpha)\delta\rho^2}{1-\rho}}, \tag{21}$$

where $\kappa := (1-\alpha)\delta\rho(\phi-\xi)+\xi$ *and* $m_{ss} := \tilde{x}_{t+1} - x_{ss}$ *is steady-state misprediction.*

Proof. See Appendix F.4. □

The parameter restrictions required for Proposition 2 and Lemma 1 (including concavity) essentially amount to requiring that perceived and actual habit formation are not too strong. We have confirmed that these restrictions hold at the parameter estimates presented in Table 4.

E.2 Modeling the Experiment

We need additional notation to map the experiment's treatments and data into the model and estimation. We define x_{it} to be participant i's daily average FITSBY screen time during period t, $\tilde{x}_{it}$ to be participant i's predicted screen time elicited on a survey, and $m_{it} = x_{it} - \tilde{x}_{it}$ to be the difference between the two. The Bonus and Bonus Control groups are denoted $g \in \{B, BC\}$, the Limit and Limit Control groups are $g \in \{L, LC\}$, and the intersection of Bonus Control and Limit Control is $g = C$. We define $\bar{y} := \mathbb{E}_i y_i$ as the expectatation over participants of variable y, and $y^g := \mathbb{E}_{i \in g} y_i$ as the expectation over group g. $\tau_t^g := x_t^g - x_t^{gC}$ and $\tilde{\tau}_t^g := \tilde{x}_t^g - \tilde{x}_t^{gC}$ are the actual and predicted average treatment effects.

We model the Screen Time Bonus as a price $p^B = \$2.50$ per hour in period 3 plus a fixed payment $F_i^B = \$50 \times \text{ceil}(x_{i1} \, \frac{\text{hours}}{\text{day}})$, where ceil$(\cdot)$ rounds up to the nearest integer, giving participant i's Bonus Benchmark. In this appendix, we generalize the primary model from Section 6 by modeling the limit as an intervention that eliminates share ω of temptation.

We define v_i^B as the valuation of the bonus elicited on survey 2, and we define v_i^L as the valuation of access to the limit functionality elicited on survey 3. We assume that on survey t, consumers are aware of period t projection bias when predicting period t consumption and are projection biased when determining their bonus and limit valuations. This assumption means that misprediction of period-t consumption is driven only by naivete about temptation, and that bonus and limit valuations are driven only by perceived temptation, not by an additional desire to offset projection bias. We acknowledge that alternative assumptions could be made.

E.3 Estimating Equations

Using the theoretical results from Appendix E.1, we can now derive equations that characterize how a consumer from our unrestricted model would behave in our experiment. These equations parallel the equation in Section 6.2, with additional terms that account for perceived habit formation. We assume that the discount factor is $\delta = 0.997$ per three-week period, consistent with a five percent annual discount rate. We estimate the remaining parameters in stages, as described below. Appendix G presents formal derivations and additional details.

Habit Formation

We first estimate λ and ρ from the decay of the bonus treatment effects. Even though λ is not a structural parameter, it is easily identified and useful in estimating the other parameters. Using the habit stock evolution formula and the linearity result in equation (19), we can write the period 4 bonus effect as the result of decayed effects from periods 2 and 3: $\tau_4^B = \lambda \left(\rho \tau_3^B + \rho^2 \tau_2^B\right)$. Similarly, the period 5 effect results from the cumulative decayed effects from periods 2–4: $\tau_5^B = \lambda \left(\rho \tau_4^B + \rho^2 \tau_3^B + \rho^3 \tau_2^B\right)$. Rearranging gives a system of two equations for λ and ρ:

$$\lambda = \frac{\tau_4^B}{\rho \tau_3^B + \rho^2 \tau_2^B} \tag{22}$$

$$\rho = \frac{\tau_5^B}{\tau_4^B (1+\lambda)}. \tag{23}$$

This non-linear system has two solutions when $\tau_2^B \neq 0$, but in our data there is only one solution that satisfies the requirement that $\rho \geq 0$.

For estimation, we assume $\tilde{\lambda} = \lambda$. This is reasonable because Figure 7 shows that participants predicted the time path of bonus effects with reasonable accuracy, so calibrating equations (22) and (23) with predicted τ_t^B would not change the estimates much. To the extent that predictions differ from actual behavior, we prefer to err on the side of using actual behavior instead of beliefs to estimate the model.

Perceived Habit Formation, Price Response, and Habit Stock Effect on Marginal Utility

After estimating λ and ρ, we estimate α, η, and ζ from the magnitude and decay of the bonus treatment effects. For each of periods 2, 3, and 4, we difference the Euler equations for the Bonus and Bonus Control groups and rearrange, giving a system of three equations for $(1-\alpha)$, η, and ζ:

$$(1-\alpha) = \frac{\eta \tau_2^B}{\delta \rho \left[-p^B + (\eta - \zeta) \tilde{\tau}_3^B + \zeta \rho \tau_2^B \right]}. \tag{24}$$

$$\eta = \frac{p^B - \zeta \rho \tau_2^B + (1-\alpha)\delta \rho^2 \zeta (1-\tilde{\lambda}) \left(\rho \tau_2^B + \tau_3^B \right)}{\tau_3^B - (1-\alpha)\delta \rho^2 \tilde{\lambda} \left(\rho \tau_2^B + \tau_3^B \right)} \tag{25}$$

$$\zeta = \frac{-\eta \tau_4^B + (1-\alpha)\delta \rho^2 \eta \tilde{\lambda} \left(\rho^2 \tau_2^B + \rho \tau_3^B + \tau_4^B \right)}{\rho \tau_3^B + \rho^2 \tau_2^B - (1-\alpha)\delta \rho^2 \left(1 - \tilde{\lambda} \right) \left(\rho^2 \tau_2^B + \rho \tau_3^B + \tau_4^B \right)}. \tag{26}$$

The first equation shows that as the anticipatory demand response in period 2 grows compared to the predicted demand response in period 3 (making $\tau_2^B / \tilde{\tau}_3^B$ larger), we infer more perceived habit formation (smaller α).

Naivete about Temptation

Next, we estimate naivete about temptation $\gamma - \tilde{\gamma}$ using the Control group's difference between perceived and actual consumption. To solve for $\gamma - \tilde{\gamma}$, we difference the actual versus perceived Euler equations for group C, giving

$$\gamma - \tilde{\gamma} = m_t^C \cdot \left[-\eta + (1-\alpha)\delta \rho^2 \left((\eta - \zeta)\tilde{\lambda} + \zeta \right) \right]. \tag{27}$$

Temptation

We estimate temptation γ using three different strategies: the limit treatment effect and valuations of the bonus and limit. Each strategy delivers an equation that we combine with equation (27) to form a system of two equations for γ and $\tilde{\gamma}$.

Limit effect. Recall that we model the limit as an intervention that eliminates share ω of temptation, starting in period 2. Thus, we can identify γ using an assumed ω plus the effect of the limit on consumption. To solve for γ, we difference the Euler equations for periods 2 versus 3 for the Limit group compared to Limit Control and rearrange, giving

$$\gamma = \eta \tau_2^L/\omega - (1-\alpha)\delta\rho\left[(\eta-\zeta)\tilde{\tau}_3^L/\omega + \zeta\rho\tau_2^L/\omega - \tilde{\gamma} - \tilde{\gamma}\tilde{\lambda}\right]. \tag{28}$$

Our primary estimates in Section 6 use this equation, after setting $\omega = 1$ and $\alpha = 1$.

Bonus valuation. Since the bonus is like a commitment device that reduces future use, people with perceived self-control problems will place higher value on the bonus. We can estimate perceived temptation $\tilde{\gamma}$ from participants' valuations. Our derivation follows Allcott, Kim, Taubinsky, and Zinman (2021), and the approach also follows Acland and Levy (2012), Augenblick and Rabin (2019), Chaloupka, Levy, and White (2019), and Carrera et al. (2021).

Let $V_t(\tilde{s}_t, \cdot)$ be the period t continuation value function conditional on $\tilde{s}_t$, according to predicted consumption and preferences before period t. This reflects preferences of a consumer filling out the multiple price list on a survey before period t. Since utility is quasilinear in money, $V_t(s_t, \cdot)$ is in units of period t dollars.

The effect of a period 3 price increase from 0 to p_3^B on the period 3 continuation value is

$$\Delta V_3\left(p^B\right) := V_3\left(\tilde{s}_3, p_3 = p_3^B\right) - V_3\left(\tilde{s}_3, p_3 = 0\right) = -p_3^B \cdot \frac{1}{2}\left(\tilde{x}_3(p_3^B) + \tilde{x}_3(0)\right) - \tilde{\gamma}\cdot\left(\tilde{x}_3(p_3^B) - \tilde{x}_3(0)\right), \tag{29}$$

where $\tilde{x}_3(p_3) = \tilde{x}_3^*(\tilde{s}_3, \tilde{\gamma}, \boldsymbol{p}_3)$ is shorthand for predicted period 3 consumption as a function of period 3 price. Figure 9 illustrates. The trapezoid $ABCD$ is $p_3^B \cdot \frac{1}{2}\left(\tilde{x}_3(p_3^B) + \tilde{x}_3(0)\right)$: the survey taker's prediction of the consumer surplus loss from the price increase from the period 3 self's perspective. The parallelogram $BCEF$ is $-\tilde{\gamma}\cdot\left(\tilde{x}_3(p_3^B) - \tilde{x}_3(0)\right)$: the predicted additional temptation reduction benefit from the survey taker's perspective.

The Screen Time Bonus combines a price change with a fixed payment of F^B. Thus, the model predicts that people filling out the bonus MPL would be indifferent between the bonus and a fixed payment of $v^B = F^B + \Delta V_3(p^B)$. Taking the expectation over participants to allow mean-zero survey noise, substituting $\tilde{\tau}_3^B := \mathbb{E}_i\left[\tilde{x}_{i3}(p_3^B) - \tilde{x}_{i3}(0)\right]$ and $\bar{\tilde{x}}_3^{B+BC} := \mathbb{E}_i\left[\frac{1}{2}\left(\tilde{x}_{i3}(p_3^B) + \tilde{x}_{i3}(0)\right)\right]$, and rearranging gives perceived temptation:

$$\tilde{\gamma} = \frac{\bar{v}^B - \bar{F}^B + p_3^B \bar{\tilde{x}}_3^{B+BC}}{-\tilde{\tau}_3^B}. \tag{30}$$

The model predicts that if consumers perceive themselves to be time consistent ($\tilde{\gamma} = 0$), the average bonus valuation would equal the average valuation from the period 3 self's perspective, $\bar{F}^B - p_3^B \bar{\bar{x}}_3^{B+BC}$. We refer to the difference between the observed average valuation and the modeled time-consistent valuation (the numerator of equation (30)) as "behavior change premium." We infer more perceived temptation $\tilde{\gamma}$ from a larger behavior change premium.

Limit valuation. People who perceive future temptation value the limit, as they perceive that it eliminates share ω of temptation. We can estimate perceived temptation $\tilde{\gamma}$ using an assumed ω plus the valuation the limit functionality. We solve for the modeled valuation similarly to how we solved for the bonus valuation above.

The effect of a period 3 temptation reduction from $\tilde{\gamma}$ to $(1-\omega)\tilde{\gamma}$ on the period 3 continuation value is

$$v^L = V_3\left(s_3, \tilde{\gamma}_3 = (1-\omega)\tilde{\gamma}\right) - V_3\left(s_3, \tilde{\gamma}_3 = \tilde{\gamma}\right) = \tilde{\gamma} \cdot \left(x_3^*(\tilde{\gamma}) - x_3^*((1-\omega)\tilde{\gamma})\right) \cdot \frac{2-\omega}{2}, \tag{31}$$

where $x_3^*(\tilde{\gamma}_3)$ is now shorthand for predicted period 3 consumption as a function of predicted period 3 temptation. Figure 9 illustrates. With $\omega = 1$, the limit valuation is the deadweight loss reduction *CEG* from the survey taker's perspective from consuming the desired amount ($x_3^*(0)$, point *G*) instead of the predicted amount ($x_3^*(\tilde{\gamma})$, point *C*). The height of this triangle is $\tilde{\gamma}$ and the width is $x_3^*(\tilde{\gamma}) - x_3^*(0)$, and thus the area is $\tilde{\gamma} \cdot (x_3^*(\tilde{\gamma}) - x_3^*(0)) \cdot \frac{1}{2}$. With $\omega < 1$, the valuation v^L equals the deadweight loss reduction trapezoid starting to the right of point *G* and bounded by segment *CE*.

Taking the expectation over participants, substituting $\tilde{\tau}_3^L := \mathbb{E}_i\left[x_3^*((1-\omega)\tilde{\gamma}) - x_3^*(\tilde{\gamma})\right]$, and rearranging gives perceived temptation:

$$\tilde{\gamma} = \frac{\bar{v}^L}{-\tilde{\tau}_3^L(2-\omega)/2}. \tag{32}$$

We infer more perceived temptation $\tilde{\gamma}$ from higher valuation $\bar{v}^L$.

Intercept

Finally, we back out a heterogeneous intercept κ_i that explains observed consumption heterogeneity. Our data do not allow us to separately identify ϕ (the direct effect of habit stock on utility) from ξ (the marginal utility shifter), so κ_i includes both of these structural parameters. We assume that participant i's observed baseline consumption x_{i1} is in a steady state characterized by equation (21). Rearranging that equation gives

$$\begin{aligned}\kappa_i := (1-\alpha)\delta\rho(\phi - \xi_i) + \xi_i = (1-(1-\alpha)\delta\rho)\,p - (1-\alpha)\delta\rho\left[(\zeta - \eta)\,m_{ss} - \left(1+\tilde{\lambda}\right)\tilde{\gamma}\right] \\ -\gamma + x_{i1}\left[-\eta - (1-\alpha)\delta\rho(\zeta - \eta) - \zeta\frac{\rho - (1-\alpha)\delta\rho^2}{1-\rho}\right].\end{aligned} \tag{33}$$

E.4 Empirical Moments and Estimation Details

Appendix Table A7 presents the full set of moments and fixed parameter values that are inputs to our unrestricted model and alternative specifications. In light of the discussion in Section 5.3, we omit the first half of period 2 when we estimate the anticipatory bonus effect τ_2^B.[22] The average of predicted use with and without the bonus $\tilde{\bar{x}}_3^{B,BC}$ and the predicted contemporaneous bonus effect $\tilde{\tau}_3^B$ are the predictions before the bonus MPL on survey 2, as displayed in Figure 7. Because we do not have an explicit elicitation of the predicted limit effect, we use the actual limit effect τ_3^L to proxy for the predicted limit effect $\tilde{\tau}_3^L$.[23] Since Figure 6 shows that the average prediction error for period t consumption is similar when elicited on survey t versus survey $t-1$, we let observed Control group misprediction m^C proxy for steady-state misprediction m_{ss}.

We winsorize the anticipatory bonus effect at $\tau_2^B \leq 0$, which affects 15 percent of draws. We also drop the 0.32 percent of bootstrap draws in which the denominator of steady-state consumption in equation (15) is not positive.

[22] Appendix Table A8 presents parameter estimates when we use all of period 2 to estimate τ_2^B. The estimated ρ is larger, as expected, but the other parameter estimates are very similar.

[23] The average difference in predicted FITSBY use between Limit and Limit Control on survey 3 is $\tilde{\tau}_3^L \approx -10.5$ minutes per day, much smaller than the actual limit effect of $\tau_3^L \approx -22.3$ minutes per day. In the limit effect strategy in equation (28), $\tilde{\tau}_3^L$ makes little difference because it is multiplied by $(1-\alpha)$, which is small. However, in the limit valuation strategy in equation (32), $\tilde{\gamma}$ is inversely proportional to $\tilde{\tau}_3^L$, so a much smaller $\tilde{\tau}_3^L$ would make the estimated $\tilde{\gamma}$ much larger.

Table A7: **Empirical Moments and Additional Parameters**

Parameter	Description	(1) Point estimate	(2) Confidence interval
δ	Three-week discount factor (unitless)	0.997	
τ_2^B	Anticipatory bonus effect (minutes/day)	-1.96	$[-7.40, 0]$
τ_3^B	Contemporaneous bonus effect (minutes/day)	-55.9	[-61.7, -50.3]
τ_4^B	Long-term bonus effect (minutes/day)	-19.2	[-24.7, -13.7]
τ_5^B	Long-term bonus effect (minutes/day)	-12.3	[-18.1, -6.54]
τ_2^L	Limit effect (minutes/day)	-24.3	[-28.1, -20.4]
m^C, m_{ss}	Control group misprediction (minutes/day)	6.13	$[4.52, 7.72]$
$\bar{\tilde{x}}_3^{B+BC}$	Predicted use with/without bonus (minutes/day)	122	$[114, 130]$
$\tilde{\tau}_3^B$	Predicted bonus effect (minutes/day)	-45.0	$[-50.0, -40.1]$
$\tilde{\tau}_3^L$	Predicted limit effect (minutes/day)	-22.3	$[-27.3, -17.3]$
ω	Temptation reduction from limit	1	
$\bar{v}^B$	Average bonus valuation ($/day)	3.20	$[3.12, 3.29]$
$\bar{v}^L$	Average limit valuation ($/day)	0.210	[0.184, 0.237]
p^B	Bonus price ($/hour)	2.5	
$\bar{F}^B$	Average bonus fixed payment ($/day)	7.03	$[6.96, 7.09]$
$\bar{x}_1$	Average baseline use (minutes/day)	153	$[149, 157]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for the empirical moments used for estimation. We winsorize at $\tau_2^B \leq 0$, and we drop the 0.32 percent of draws in which the denominator of steady-state consumption in equation (15) is not positive.

Table A8: **Primary Parameter Estimates Using τ_2^B for All of Period 2**

Parameter	Description (units)	(1) Unrestricted model ($\alpha = \hat{\alpha}$)
λ	Habit stock effect on consumption (unitless)	1.08
		$[0.565, 3.09]$
ρ	Habit formation (unitless)	0.308
		$[0.113, 0.507]$
α	Projection bias (unitless)	0.725
		$[0.427, 0.969]$
η	Price coefficient (\$-day/hour^2)	-2.85
		$[-3.15, -2.61]$
ζ	Habit stock effect on marginal utility (\$-day/hour^2)	2.91
		$[1.49, 8.45]$
$\gamma - \tilde{\gamma}$	Naivete about temptation (\$/hour)	0.283
		$[0.208, 0.359]$
γ	Temptation (\$/hour)	1.16
		$[0.938, 1.40]$
$\bar{\kappa}$	Average intercept (\$/hour)	-1.95
		$[-3.40, -0.574]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals from the estimation strategy described in Section E.3. We winsorize at $\tau_2^B \leq 0$, and we drop the 0.32 percent of draws in which the denominator of steady-state consumption in equation (15) is not positive. Temptation γ is from the limit effect strategy, using equation (28). This parallels column 2 of Table 4, except using all of period 2 (instead of only the second half of period 2) to estimate the anticipatory bonus effect τ_2^B.

E.5 Alternative Temptation Estimates

Appendix Table A9 presents alternative estimates of temptation γ in the restricted and unrestricted models. After repeating the primary limit effect estimate, the table reports the bonus valuation estimate. Before the bonus MPL on survey 2, the average participant predicted that they would use FITSBY 2.5 and 1.6 hours per day without and with the bonus, respectively. Thus, the average survey taker would have predicted that the price increase would cause a consumer surplus loss from their period 3 self's perspective of $p_3^B \bar{\tilde{x}}_3 \approx \$2.50 \times \frac{1}{2}(2.5 + 1.6) \approx \$5.09$ per day of period 3. This is the trapezoid *ABCD* on Figure 9. The average bonus fixed payment was $\bar{F}^B \approx \$7.03$ per day. Thus, if the average participant perceived herself to be time consistent, she would have been indifferent between the bonus and a certain payment of $\$7.03 - \$5.09 \approx \$1.94$ per day.

In reality, the average participant was indifferent between the bonus and a certain payment of \$64, or $\bar{v}^B \approx \$64/20 \approx \$3.20$ per day over the 20-day period. This excess valuation implies a behavior change premium of $\$3.20 - \$1.94 \approx \$1.26$ per day. This is the parallelogram *BCEF* on Figure 9: the additional temptation reduction benefit that the period 2 survey taker perceives from the reduced FITSBY use caused by the bonus. Rearranging this logic into equation (30) gives perceived temptation $\hat{\tilde{\gamma}} \approx 1.34$ \$/hour. Using the estimated naivete of $\widehat{\gamma - \tilde{\gamma}} \approx 0.274$ gives $\hat{\gamma} \approx 1.61$ for the bonus valuation strategy in column 1.

The average Limit group participant was indifferent between access to the limit functionality for period 3 and a certain payment of \$4.20, or $\bar{v}^L \approx \$4.20/20 \approx \$0.210$ per day over the 20-day period. This is the triangle on Figure 9: the perceived deadweight loss reduction from the reduced FITSBY use caused by the limit. Inserting this into equation (32) with $\omega = 1$ gives perceived temptation $\hat{\tilde{\gamma}} = \frac{\bar{v}^L}{-\bar{\tau}_3^L/2} \approx \frac{0.210}{(-(-22.3)/60)/2} \approx$ 1.13 \$/hour. Using $\widehat{\gamma - \tilde{\gamma}} \approx 0.274$ gives $\hat{\gamma} \approx 1.41$ for the limit valuation strategy in column 1.

So far, we have modeled FITSBY screen time on other devices as part of an outside option that is not affected by self-control problems. In Appendix G.5, we generalize the model to include multiple temptation goods. As discussed in Section 5.5, self-reports suggest that the limit increased FITSBY use on other devices by 4.2 minutes per day, while the bonus reduced FITSBY use on other devices by 8.1 minutes per day. We use these additional moments to identify the multiple-good model.

The next three rows in Appendix Table A9 present estimates from the multiple-good model. The limit effect estimate increases to $\hat{\gamma} \approx 1.31$ \$/hour, because in the multiple-good model, more temptation is needed to explain the observed limits when consumers setting the limits think they'll evade the limits through substitution to other devices. The bonus valuation estimate decreases to $\hat{\gamma} \approx 1.44$ \$/hour, because in the multiple-good model, less temptation is needed to explain the observed bonus valuation when consumers think the bonus will also reduce FITSBY use on other devices. The limit valuation estimate increases to to $\hat{\gamma} \approx 2.09$ \$/hour, because in the multiple-good model, more temptation is needed to explain the observed limit valuation when consumers think the limit will also increase FITSBY use on other devices.

Next, we return to the single-good model and consider an alternative specification where we estimate ω from differences in self-reported *ideal use change* between the Limit and Limit Control groups. Intuitively, if the Limit group reports on survey 3 that looking back over period 2, they ideally would not have further reduced their screen time, this suggests that the limit functionality fully eliminated temptation ($\omega = 1$). Extending this intuition, we estimate ω as the share of the Limit Control group's *ideal use change* that is eliminated in the Limit treatment group. If d_2^g is group g's average *ideal use change* reported on survey 3 retrospectively about period 2, this is:

$$\omega = \frac{d_2^L - d_2^{LC}}{-d_2^{LC}}. \tag{34}$$

In the data, the Limit and Limit Control groups report that they ideally would have changed use by -9.5 and -15 percent, respectively. This gives $\hat{\omega} \approx \frac{-0.095-(-0.15)}{-(-0.15)} \approx 0.385$.

If we assume that the limit only eliminates share $\omega < 1$ of temptation, the limit effect strategy will deliver larger γ, because we infer that the true effect of temptation on consumption is larger. By contrast, the limit valuation strategy will deliver smaller γ, because a smaller γ is needed to explain a given valuation $\bar{v}^L$ when temptation has a larger effect on consumption. Appendix Table A9 shows that in the restricted model

($\alpha = 1$), the limit effect $\hat{\gamma}$ increases from 1.09 to 2.82, while the limit valuation strategy $\hat{\gamma}$ decreases from 1.41 to 0.975.

Finally, we extend the limit effect strategy to allow for individual-specific heterogeneity in γ. To do this, we exploit the facts that we observe each participant's period 2 *limit tightness* H_{i2} and that tightness is closely related to the limit treatment effect. We estimate heterogeneous period 2 and 3 limit effects as a function of period 2 *limit tightness* by adding an interaction term $\tau^{HL}H_{i2}L_i$ to the treatment effect estimation in equation (4); see Appendix Table A10.[24] For each participant, we insert the fitted limit effect $\hat{\tau}_{it}^{L} = \hat{\tau}_{t}^{L} + \hat{\tau}^{HL}H_{i2}$ into equation (28) to infer γ_i. The final row of Appendix Table A9 shows that although this allows substantial heterogeneity, the average temptation $\bar{\gamma}$ is essentially the same as the homogeneous γ from the limit effect strategy, as one would expect.

These alternative approaches imply temptation γ is between about \$1 and \$3 per hour. Our primary strategy (the limit effect) is relatively conservative.

[24] H_i is missing for the Limit Control group, so we are not able to include the main effect of H_{i2} in this regression. In theory, this could generate omitted variable bias if period 2 or 3 control group consumption varies with the tightness that they would have set. Appendix Table A10 shows that H_{i2} is associated with the Limit group's consumption in the second half of period 1 (before the limit functionality was turned on). However, the association is small compared to the association in periods 2 and 3, which suggests that the potential omitted variables bias is relatively small.

Table A9: **Alternative Temptation Parameter Estimates**

Parameter	Description (units)	(1) Restricted model ($\tau_2^B = 0$, $\alpha = 1$)	(2) Unrestricted model ($\alpha = \hat{\alpha}$)
γ	Temptation ($/hour)		
	Limit effect (primary)	1.09	1.11
		[0.884, 1.30]	[0.903, 1.33]
	Bonus valuation	1.61	1.62
		[1.29, 1.94]	[1.29, 1.94]
	Limit valuation	1.41	1.41
		[1.19, 1.75]	[1.19, 1.76]
	Limit effect, multiple-good model	1.31	
		[1.01, 1.71]	
	Bonus valuation, multiple-good model	1.44	1.45
		[1.16, 1.73]	[1.17, 1.74]
	Limit valuation, multiple-good model	2.09	2.09
		[1.33, 7.10]	[1.33, 7.10]
	Limit effect, $\omega = \hat{\omega}$	2.82	2.92
		[2.11, 3.92]	[2.22, 4.16]
	Limit valuation, $\omega = \hat{\omega}$	0.975	0.979
		[0.826, 1.19]	[0.833, 1.20]
$\bar{\gamma}$	Average temptation ($/hour)	1.08	1.10
	Heterogeneous limit effect	[0.873, 1.29]	[0.889, 1.31]

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for alternative estimates of temptation γ. Each row reflects estimates from a different specification. γ for the limit effect, bonus valuation, and limit valuation strategies is from equations (28), (30), and (32), respectively, combined with naivete $\gamma - \tilde{\gamma}$ from equation (27). γ for the multiple-good model is from equations (182), (187), and (190) in Appendix G.5; we do not have a limit effect estimate for the unrestricted multiple-good model. $\hat{\omega}$ is from equation (34).

Table A10: **Heterogeneity in Limit Effect by Limit Tightness**

	2nd half of period 1 FITSBY use	Period 2 FITSBY use	Period 3 FITSBY use
	(1)	(2)	(3)
Bonus treatment	−4.702	−3.228	−54.384
	(2.001)	(2.154)	(2.835)
Limit treatment	−5.281	0.447	−1.248
	(2.143)	(2.308)	(3.041)
Limit treatment × period 2 limit tightness	0.114	−0.551	−0.469
	(0.027)	(0.029)	(0.039)
1st half of period 1 FITSBY use	0.845		
	(0.014)		
Period 1 FITSBY use		0.894	0.795
		(0.015)	(0.020)
Observations	1,933	1,930	1,931
R^2	0.849	0.795	0.665

Notes: This table presents the effects of bonus and limit treatments on FITSBY use in periods 1, 2, and 3 using equation (4), including an additional interaction between the Limit group indicator and period 2 *limit tightness*. *Limit tightness* is the amount by which a user's limits would have hypothetically reduced overall screen time if applied to their baseline use without snoozes; see equation (5). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube.

E.6 Model Estimates with Sample Weights

Table A11: **Demographics in Weighted Sample**

	(1) Analysis sample	(2) Balanced sample	(3) U.S. adults
Income ($000s)	40.8	42.1	43.0
College	0.67	0.55	0.30
Male	0.39	0.42	0.49
White	0.72	0.72	0.74
Age	33.7	38.7	47.6
Period 1 phone use (minutes/day)	333.0	339.3	.
Period 1 FITSBY use (minutes/day)	152.8	155.4	.

Notes: Column 1 presents average demographics for our analysis sample, column 2 presents average demographics for our weighted sample, and column 3 presents average demographics of American adults using data from the 2018 American Community Survey. The sample weights are initially calculated to make the sample nationally representative on these five demographics but are then winsorized at $[1/3, 3]$ to reduce precision loss.

Table A12: **Empirical Moments and Additional Parameters in Weighted Sample**

Parameter	Description	(1) Point estimate	(2) Confidence interval
δ	Three-week discount factor (unitless)	0.997	
τ_2^B	Anticipatory bonus effect (minutes/day)	-4.41	$[-12.8, 0]$
τ_3^B	Contemporaneous bonus effect (minutes/day)	-58.5	[-67.3, -50.3]
τ_4^B	Long-term bonus effect (minutes/day)	-25.1	[-34.7, -15.7]
τ_5^B	Long-term bonus effect (minutes/day)	-16.4	[-26.5, -7.93]
τ_2^L	Limit effect (minutes/day)	-23.3	[-29.6, -16.7]
m^C	Control group misprediction (minutes/day)	4.96	$[3.03, 7.11]$
$\tilde{\bar{x}}_3^{B+BC}$	Predicted use with/without bonus (minutes/day)	127	$[114, 140]$
$\tilde{\tau}_3^B$	Predicted bonus effect (minutes/day)	-49.4	$[-56.6, -41.9]$
$\tilde{\tau}_3^L$	Predicted limit effect (minutes/day)	-20.7	$[-27.9, -12.7]$
ω	Temptation reduction from limit	1	
$\bar{v}^B$	Average bonus valuation ($/day)	3.29	$[3.15, 3.44]$
$\bar{v}^L$	Average limit valuation ($/day)	0.271	[0.229, 0.315]
p^B	Bonus price ($/hour)	2.5	
$\bar{F}^B$	Average bonus fixed payment ($/day)	6.84	$[6.72, 6.96]$
$\bar{x}_1$	Average baseline use (minutes/day)	156	$[149, 164]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for the empirical moments used for estimation. We winsorize at $\tau_2^B \leq 0$, and we drop the 0.32 percent of draws in which the denominator of steady-state consumption in equation (15) is not positive. This parallels Table 3, except using the weighted sample. The sample weights are initially calculated to make the sample nationally representative on the five demographics in Appendix Table A11 but are then winsorized at $[1/3, 3]$ to reduce precision loss.

Table A13: **Model Parameter Estimates in Weighted Sample**

Parameter	Description (units)	Restricted model ($\tau_2^B = 0$, $\alpha = 1$)
λ	Habit stock effect on consumption (unitless)	1.93
		$[0.757, 3.89]$
ρ	Habit formation (unitless)	0.223
		$[0.122, 0.469]$
α	Projection bias (unitless)	1
η	Price coefficient (\$-day/hour^2)	-2.57
		$[-2.98, -2.23]$
ζ	Habit stock effect on marginal utility (\$-day/hour^2)	4.95
		$[2.07, 9.96]$
$\gamma - \tilde{\gamma}$	Naivete about temptation (\$/hour)	0.212
		$[0.130, 0.307]$
γ	Temptation (\$/hour)	0.998
		$[0.709, 1.30]$
$\bar{\kappa}$	Average intercept (\$/hour)	-1.99
		$[-3.52, -0.422]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals from the estimation strategy described in Section E.3. We winsorize at $\tau_2^B \leq 0$, and we drop the 0.32 percent of draws in which the denominator of steady-state consumption in equation (15) is not positive. Temptation γ is from the limit effect strategy, using equation (28). This parallels Table 4, except using the weighted sample. The sample weights are initially calculated to make the sample nationally representative on the five demographics in Appendix Table A11 but are then winsorized at $[1/3, 3]$ to reduce precision loss.

F Proofs of Propositions in Appendix E.1

Given naivete about projection bias, the predicted continuation value function given predicted consumption and habit stock is

$$V_{t+1}(\tilde{s}_{t+1}) = \sum_{r=t+1}^{T} \delta^{r-t} u_r\left(\tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, \boldsymbol{p}_r); \tilde{s}_r, p_r\right). \tag{35}$$

The consumer's predicted objective function in future period t can thus be written as

$$\tilde{U}_t(x_t; \tilde{s}_t) = u_t(x_t; \tilde{s}_t, p_t) + \tilde{\gamma} x_t + \delta V_{t+1}(\tilde{s}_{t+1}), \tag{36}$$

and the consumer's actual period t objective function from equation (2) can be written as

$$U_t\left(x_t; s_t\right) = u_t\left(x_t; s_t, p_t\right) + \gamma x_t + \begin{array}{c} \alpha \sum_{r=t+1}^{T} \delta^{r-t} u_r\left(\tilde{x}_r^*\left(s_t, \tilde{\gamma}, \boldsymbol{p}_r\right); s_t, p_r\right) \\ +(1-\alpha)\delta V_{t+1}\left(\tilde{s}_{t+1}\right) \end{array}. \tag{37}$$

Recall that we defined $u_t := u_t\left(x_t^*; s_t, p_t\right)$, $\tilde{x}_r := \tilde{x}_r^*\left(\tilde{s}_r, \tilde{\gamma}, \boldsymbol{p}_r\right)$, and $\tilde{u}_r := u_r\left(\tilde{x}_r; \tilde{s}_r, p_r\right)$.

F.1 Proof of Proposition 1: Euler Equation

In this section, we derive the Euler equation (equation (18)), proving Proposition 1.

Proof. The time t first-order condition from maximizing utility (equation (37)) is

$$\begin{aligned}
\frac{\partial u_t}{\partial x_t} + \gamma &= -(1-\alpha)\delta \frac{d\tilde{s}_{t+1}}{dx_t} \frac{dV_{t+1}\left(\tilde{s}_{t+1}\right)}{d\tilde{s}_{t+1}} && (38) \\
&= -(1-\alpha)\delta \frac{d\tilde{s}_{t+1}}{dx_{t+1}} \left[\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}} \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}} + \frac{\partial \tilde{u}_{t+1}}{\partial \tilde{s}_{t+1}}\right] - (1-\alpha)\delta^2 \frac{d\tilde{s}_{t+2}}{dx_t} \frac{dV_{t+2}\left(\tilde{s}_{t+2}\right)}{d\tilde{s}_{t+2}} && (39) \\
&= -(1-\alpha)\delta\rho \left[\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}} \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}} + \frac{\partial \tilde{u}_{t+1}}{\partial \tilde{s}_{t+1}}\right] - (1-\alpha)\left(\delta\rho\right)^2 \left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right) \frac{dV_{t+2}\left(\tilde{s}_{t+2}\right)}{d\tilde{s}_{t+2}}, && (40)
\end{aligned}$$

where the third line uses the fact that the total derivative of predicted period $t+2$ habit stock with respect to period t consumption is

$$\begin{aligned}
\frac{d\tilde{s}_{t+2}}{dx_t} &= \frac{\partial \tilde{s}_{t+2}}{\partial \tilde{s}_{t+1}} \frac{\partial \tilde{s}_{t+1}}{\partial x_t} + \frac{\partial \tilde{s}_{t+2}}{\partial \tilde{x}_{t+1}} \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}} \frac{\partial \tilde{s}_{t+1}}{\partial x_t} \\
&= \rho^2 \left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right) && (41)
\end{aligned}$$

The time t self predicts that the time $t+1$ self will maximize equation (36), setting x_{t+1} according to the following first-order condition:

$$\begin{aligned}
0 &= \frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}} + \tilde{\gamma} + \delta \frac{d\tilde{s}_{t+2}}{dx_{t+1}} \frac{dV_{t+2}\left(\tilde{s}_{t+2}\right)}{d\tilde{s}_{t+2}} && (42) \\
&= \frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}} + \tilde{\gamma} + \delta\rho \frac{dV_{t+2}\left(\tilde{s}_{t+2}\right)}{d\tilde{s}_{t+2}} && (43)
\end{aligned}$$

Multiplying the predicted time $t+1$ first-order condition by $(1-\alpha)\delta\rho\left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right)$ gives

$$0 = (1-\alpha)\delta\rho \left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right) \left(\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}} + \tilde{\gamma}\right) + (1-\alpha)\left(\delta\rho\right)^2 \left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right) \frac{dV_{t+2}\left(\tilde{s}_{t+2}\right)}{d\tilde{s}_{t+2}} \tag{44}$$

The last term is the same as the last term in the time t first-order condition. Adding this equation to the time t first-order condition yields

$$\frac{\partial u_t}{\partial x_t}+\gamma=(1-\alpha)\delta\rho\left(1+\frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right)\left(\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}}+\tilde{\gamma}\right)-(1-\alpha)\delta\rho\left[\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}}\frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}+\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{s}_{t+1}}\right] \tag{45}$$

$$\frac{\partial u_t}{\partial x_t}+\gamma=(1-\alpha)\delta\rho\left[\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}}+\tilde{\gamma}+\frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\tilde{\gamma}-\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{s}_{t+1}}\right]. \tag{46}$$

We now derive the Euler equation with our quadratic functional form. The partial derivatives are

$$\frac{\partial u_t}{\partial x_t}=\eta x_t^*+\zeta s_t+\xi_t-p_t \tag{47}$$

$$\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{x}_{t+1}}=\eta\tilde{x}_{t+1}+\zeta\tilde{s}_{t+1}+\xi_{t+1}-p_{t+1} \tag{48}$$

$$\tilde{\lambda}_{t+1}:=\frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}} \tag{49}$$

$$\frac{\partial \tilde{u}_{t+1}}{\partial \tilde{s}_{t+1}}=\zeta\tilde{x}_{t+1}+\phi. \tag{50}$$

Substituting these into equation (46) yields equation (18).

□

F.2 Proof of Proposition 2: Linear Policy Functions

In this section, we first show that the policy function is linear in habit stock. We then show that if the objective function is concave, λ converges to a constant far from the time horizon. We then show the conditions under which utility is concave. Finally, we show the condition required for μ to converge to a constant far from the time horizon. Our proof strategy follows Gruber and Köszegi (2001).

Lemma 2. *Suppose $u_t(x_t;s_t,p_t)$ is given by equation (3) and $(x_0^*,...,x_T^*)$ is a perception-perfect strategy profile. Then for any t,*

$$x_t^*(s_t,\gamma,\boldsymbol{p}_t)=\lambda_t s_t+\mu_t(\gamma) \tag{51}$$

$$\tilde{x}_t^*(s_t,\tilde{\gamma},\boldsymbol{p}_t)=\tilde{\lambda}_t s_t+\mu_t(\tilde{\gamma}) \tag{52}$$

where λ_t is a function of only $\{\eta,\zeta,\delta,\rho,\alpha\}$, $\tilde{\lambda}_t$ is a function of only $\{\eta,\zeta,\delta,\rho\}$, and μ_t is linear in p_t.

Proof. We prove by backwards induction. First, we show that the result holds for period T. Given our functional form, the period T first-order condition is

$$\eta x_T^* + \zeta s_T + \xi_T - p_T + \gamma = 0, \tag{53}$$

and thus

$$x_T^* = \frac{\zeta s_T + \xi_T - p_T + \gamma}{-\eta}. \tag{54}$$

Thus, x_T^* can be written as

$$x_T^* = \lambda_T s_T + \mu_T(\gamma), \tag{55}$$

with $\lambda_T = \frac{\zeta}{-\eta}$ and $\mu_T(\gamma) = \frac{\xi_T - p_T + \gamma}{-\eta}$.

Analogously, predicted consumption is

$$\tilde{x}_T = \frac{\zeta \tilde{s}_T + \xi_T - p_T + \tilde{\gamma}}{-\eta}, \tag{56}$$

so $\tilde{x}_T$ can be written as

$$\tilde{x}_T = \lambda_T \tilde{s}_T + \mu_T(\tilde{\gamma}), \tag{57}$$

with $\mu_T(\tilde{\gamma}) = \frac{\xi_T - p_T + \tilde{\gamma}}{-\eta}$. The function μ_T is linear in p_T.

Now, we use the Euler equation to show that if the result holds for $t+1$, it holds for t. The Euler equation is

$$\begin{aligned}
\eta x_t^* + \zeta s_t + \xi_t - p_t + \gamma &= (1-\alpha)\delta\rho \left[\eta \tilde{x}_{t+1} + \zeta \tilde{s}_{t+1} + \xi_{t+1} - p_{t+1} + \left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right)\tilde{\gamma} - \zeta \tilde{x}_{t+1} - \phi \right] \\
&= (1-\alpha)\delta\rho \left[(\eta - \zeta)\tilde{x}_{t+1} + \zeta \tilde{s}_{t+1} + \xi_{t+1} - p_{t+1} + \left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right)\tilde{\gamma} - \phi \right]
\end{aligned}$$

Substituting $\tilde{x}_{t+1} = \tilde{\lambda}_{t+1}\tilde{s}_{t+1} + \mu_{t+1}(\tilde{\gamma})$, $\tilde{s}_{t+1} = \rho\,(s_t + x_t^*)$, and $\tilde{\lambda}_{t+1} = \frac{\partial \tilde{x}_{t+1}^*}{\partial \tilde{s}_{t+1}}$ gives

$$\eta x_t^* + \zeta \tilde{s}_t + \xi_t - p_t + \gamma = (1-\alpha)\delta\rho \left[(\eta - \zeta)\left(\tilde{\lambda}_{t+1}\rho\,(x_t^* + s_t) + \mu_{t+1}(\tilde{\gamma})\right) + \zeta\rho(s_t + x_t^*) + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \phi \right]. \tag{58}$$

Solving for x_t^* gives

$$x_t^* = \frac{s_t\left[\zeta - (1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1} + \zeta\right)\right] + \xi_t - p_t + \gamma - (1-\alpha)\delta\rho\left[(\eta-\zeta)\mu_{t+1}(\tilde{\gamma}) + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \phi\right]}{-\eta + (1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1} + \zeta\right)}. \tag{59}$$

Thus, $x_t^* = \lambda_t s_t + \mu_t(\gamma)$, with

$$\lambda_t = \frac{\zeta - (1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1}+\zeta\right)}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1}+\zeta\right)}, \tag{60}$$

and

$$\mu_t(\gamma) = \frac{\xi_t - p_t + \gamma - (1-\alpha)\delta\rho\left[\xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \phi\right] + (1-\alpha)\delta\rho\left(\zeta-\eta\right)\mu_{t+1}(\tilde{\gamma})}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1}+\zeta\right)}. \tag{61}$$

We can analogously begin with the period t Euler equation as *predicted* before period t, which has $\tilde{\gamma}$ and $\tilde{s}_t$ instead of γ and s_t on the left-hand side, and does not have the $(1-\alpha)$ term. This gives $\tilde{x}_t = \tilde{\lambda}_t\tilde{s}_t + \mu_t(\tilde{\gamma})$, with

$$\tilde{\lambda}_t = \frac{\zeta - \delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1}+\zeta\right)}{-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1}+\zeta\right)}. \tag{62}$$

and $\mu_t(\tilde{\gamma})$ given by equation (61) except that, as as implied by writing $\mu_t(\tilde{\gamma})$ instead of $\mu_t(\gamma)$, the third term in the numerator is $\tilde{\gamma}$ instead of γ.[25] Thus, λ_t is not correctly perceived in advance of period t.

λ_t depends only on $\{\eta,\zeta,\delta,\rho,\alpha\}$, and $\tilde{\lambda}_t$ depends only on $\{\eta,\zeta,\delta,\rho\}$, as long as $\tilde{\lambda}_{t+1}$ depends only on $\{\eta,\zeta,\delta,\rho\}$. Because consumers misperceive γ, μ_r is also misperceived for $r>t$. The function μ_t is linear in p_t. □

We now show that with concave utility, λ_t and $\tilde{\lambda}_t$ are constant in t far from the time horizon.

Lemma 3. *Suppose the conditions for Lemma 2 hold and utility is concave. Then for any fixed t,*

$$\lambda = \lim_{T\to\infty}\lambda_t = \frac{\zeta - (1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)}, \tag{63}$$

and

$$\tilde{\lambda} = \lim_{T\to\infty}\tilde{\lambda}_t = \frac{-\eta-\sqrt{\eta^2-4\frac{\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}\zeta}}{\frac{2\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}}. \tag{64}$$

Proof. To show that λ_t is constant in t far from the time horizon, it suffices to prove the convergence of $\tilde{\lambda}_t$ to the steady state, since λ_t is a function of $\tilde{\lambda}_{t+1}$ and other deterministic parameters. We define the function

[25] Equation (60) is much simpler than equation (25) of Gruber and Köszegi (2001), and our expression for λ_t does not depend on actual or perceived temptation γ or $\tilde{\gamma}$, while theirs depends on present focus β. This is because in their quasi-hyperbolic framework, $1-\beta$ multiplies λ_{t+1} parameters in the Euler equation and doesn't drop out.

$f(\tilde{\lambda})$ according to Equation (62) that describes the recursion $\tilde{\lambda}_t = f(\tilde{\lambda}_{t+1})$. We first find the values of $\tilde{\lambda}$ that could be fixed points. Assuming constant $\tilde{\lambda}$ and rearranging Equation (60) gives

$$-\eta\tilde{\lambda} + \delta\rho^2\left((\eta-\zeta)\tilde{\lambda}^2 + \zeta\tilde{\lambda}\right) = \zeta + \delta\rho^2\left((\zeta-\eta)-\zeta\right). \tag{65}$$

Collecting terms gives

$$\tilde{\lambda}^2\delta\rho^2(\eta-\zeta) + \tilde{\lambda}\eta\left(\delta\rho^2-1\right) + \zeta\left(\delta\rho^2-1\right) = 0 \tag{66}$$

$$\tilde{\lambda}^2\frac{\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)} + \tilde{\lambda}\eta + \zeta = 0. \tag{67}$$

Using the quadratic formula gives

$$\tilde{\lambda} = \frac{-\eta \pm \sqrt{\eta^2 - 4\frac{\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}\zeta}}{\frac{2\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}}. \tag{68}$$

We now prove convergence. The function $f(\lambda)$ has the following properties. First, $f(\lambda)$ is always increasing as

$$f'(\tilde{\lambda}) = \frac{-\delta\rho^2(\eta-\zeta)\left(-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right) - \delta\rho^2(\eta-\zeta)\left(\zeta-\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right)}{\left(-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right)^2} \tag{69}$$

$$= \frac{\delta\rho^2(\zeta-\eta)^2}{\left(-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right)^2} > 0. \tag{70}$$

Second, f is convex on $(-\infty,\bar{\tilde{\lambda}})$,where $\bar{\tilde{\lambda}} = \dfrac{-\eta+\delta\rho^2\zeta}{\delta\rho^2(\zeta-\eta)} > 0$. This comes from the sign of its second derivative

$$f''(\tilde{\lambda}) = \frac{2\delta^2\rho^4(-\eta+\zeta)^3}{\left(-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right)^3}, \tag{71}$$

which is determined by the sign of the denominator.

Third, for $\tilde{\lambda} > \bar{\tilde{\lambda}}$, $f(\tilde{\lambda})$ is always negative due to the denominator in equation (62), hence none of the solutions for a constant $\tilde{\lambda}_t$ are in this region.

Fourth, $f(0) > 0$ since $\delta\rho^2 < 1$ and

$$f(0) = \frac{\zeta(1-\delta\rho^2)}{-\eta+\delta\rho^2\zeta}. \tag{72}$$

Fifth, $f(\tilde{\lambda})$ is continuous on $[0,\bar{\bar{\lambda}})$ and $\lim_{\tilde{\lambda}\to\bar{\bar{\lambda}}} f(\tilde{\lambda}) = \infty$ as the denominator in equation (60) goes to 0.

The properties highlighted above imply that both candidate solutions for a constant $\tilde{\lambda}_t$ in equation (68) are positive. To see this, denote the two candidate solutions as $(\tilde{\lambda}_1,\tilde{\lambda}_2)$, with $\tilde{\lambda}_1 < \tilde{\lambda}_2$. Since $f(0) > 0$, we know that at least one solution for $\tilde{\lambda}$ is positive given $-\eta > 0$. Furthermore, since $f(\tilde{\lambda}) > 0$ on $(-\infty,\bar{\bar{\lambda}}]$, it cannot be true that an increasing, continuous, and convex function that diverges to infinity at $\bar{\bar{\lambda}}$ only crosses the identity function once in $[0,\bar{\bar{\lambda}})$. Hence, both solutions are in $[0,\bar{\bar{\lambda}}]$.

Given this result and the convex shape of this function, it must be true that $\tilde{\lambda}_1$ is a stable constant solution for the recursion while $\tilde{\lambda}_2$ is unstable. For any point in $[0,\tilde{\lambda}_1]$ the recursion implies an increase in $\tilde{\lambda}_t$ ($f(\tilde{\lambda}) > \tilde{\lambda}$), for any point in $[\tilde{\lambda}_1,\tilde{\lambda}_2]$ the recursion implies a decrease in λ_t ($f(\tilde{\lambda}) < \tilde{\lambda}$), and for any point in $[\tilde{\lambda}_2,\bar{\bar{\lambda}}]$ the recursion implies an increase in $\tilde{\lambda}_t$ ($f(\tilde{\lambda}) > \tilde{\lambda}$). Overall, this means that for any starting value of $\tilde{\lambda}_t \in [0,\tilde{\lambda}_2)$ the recursion converges to $\tilde{\lambda}_1$.

To complete the proof, we begin with $\tilde{\lambda}_T$ and then prove that far away from the time horizon, $\tilde{\lambda}_t$ is constant. To do this, we need to show that this initial value, given by $\tilde{\lambda}_T = \dfrac{\zeta}{-\eta}$, is less than $\tilde{\lambda}_2$. To show this, notice that the two solutions $(\tilde{\lambda}_1,\tilde{\lambda}_2)$ are symmetrically placed around $\tilde{\lambda}_s = \dfrac{-\eta(1-\delta\rho^2)}{2\delta\rho^2(\zeta-\eta)}$. Given this value, by the parametric assumption that guarantees the existence of the two constant solutions for the recursion, we know that

$$\eta^2 - 4\frac{\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}\zeta > 0, \tag{73}$$

and since

$$\eta^2 > 2\frac{\delta\rho^2(\zeta-\eta)\zeta}{(1-\delta\tilde{\rho}^2)} \iff \frac{\zeta}{-\eta} < \frac{-\eta(1-\delta\rho^2)}{2\delta\rho^2(\zeta-\eta)}, \tag{74}$$

we have that $\tilde{\lambda}_T < \tilde{\lambda}_s$. Then $\tilde{\lambda}_T < \tilde{\lambda}_s < \lambda_2$, and hence the backward recursion starting from $\tilde{\lambda}_T$ converges far from the time horizon to a stationary value $\tilde{\lambda}^* = \tilde{\lambda}_1$. Moreover, $f(\tilde{\lambda}_T)$ can be written as $\frac{\zeta-X}{-\eta+X}$, and we know that $\frac{\zeta-X}{-\eta+X} > \frac{\zeta}{-\eta}$ whenever $X < 0$. Then, given that

$$X = (1-\alpha)\delta\rho^2\Big((\eta-\zeta)\tilde{\lambda}_T + \zeta\Big) < 0 \iff (\eta-\zeta)\frac{\zeta}{-\eta} + \zeta < 0 \iff \frac{\zeta^2}{\eta} < 0 \iff \eta < 0,$$

we have $X < 0$. Thus we can conclude that $f(\tilde{\lambda}_T) > \tilde{\lambda}_T$ and therefore, $\tilde{\lambda}_T < \tilde{\lambda}_1$. Thus, we have proved that the backward recursion converges to an stationary value of $\tilde{\lambda}^* = \tilde{\lambda}_1$, and it does so as an increasing sequence.

Finally, we demonstrate that λ_t also converges to a steady-state in a decreasing manner. We note that

$$\lambda = g(\tilde{\lambda}) = \frac{\zeta - (1-\alpha)\delta\rho^2\Big((\eta-\zeta)\tilde{\lambda} + \zeta\Big)}{-\eta + (1-\alpha)\delta\rho^2\Big((\eta-\zeta)\tilde{\lambda} + \zeta\Big)} \tag{75}$$

Which we can rewrite as

$$\lambda = g(\tilde{\lambda}) = \frac{\zeta + (1-\alpha)\delta\rho^2\left((\zeta-\eta)\tilde{\lambda}+\zeta\right)}{-\eta-(1-\alpha)\delta\rho^2\left((\zeta-\eta)\tilde{\lambda}+\zeta\right)} \tag{76}$$

Note that $(1-\alpha)\delta\rho^2(\zeta-\eta)$ is positive, so the numerator decreases when $\tilde{\lambda}$ decreases, whereas the denominator increases, since $-(1-\alpha)\delta\rho^2(\zeta-\eta)\tilde{\lambda}$ becomes less negative. Hence, $g(\tilde{\lambda})=\lambda$ also decreases when $\tilde{\lambda}$ decreases. $\square$

We now show that utility is concave in x_t as long as there is not too much habit formation in a specific sense.

Lemma 4. *Suppose the conditions for Lemma 2 hold and U_t is given by equation (37). Then for any t, $\frac{dU_t}{dx_t}$ is continuous in x_t. Furthermore, if $\tilde{\lambda}^b$ is an upper bound on $\tilde{\lambda}_t$ and $\frac{(1-\alpha)\tilde{\lambda}^b}{\left(1+\tilde{\lambda}^b\right)-\delta\rho^2\left(1+\tilde{\lambda}^b\right)^2} < \frac{-\eta}{\zeta}$, then $\frac{\partial^2 U_t}{\partial x_t^2}<0$ for all $t \geq 0$.*

Proof. The period t decisionmaker maximizes equation (37). The derivative of equation (37) can be written as

$$\frac{dU_t\left(x_t;s_t\right)}{dx_t} = \frac{\partial u_t}{\partial x_t} + \gamma + (1-\alpha)\sum_{r=t+1}^{T}\delta^{r-t}\frac{\partial \tilde{s}_r}{\partial x_t}\left[\underbrace{\frac{\partial \tilde{u}_r}{\partial \tilde{x}_r}\frac{\partial \tilde{x}_r}{\partial \tilde{s}_r}+\frac{\partial \tilde{u}_r}{\partial \tilde{s}_r}}_{\text{effect of } \tilde{s}_r \text{ on period } r \text{ utility}} + \underbrace{\delta\rho\frac{\partial V_{r+1}}{\partial \tilde{s}_{r+1}}\frac{\partial \tilde{x}_r}{\partial \tilde{s}_r}}_{\text{partial effect on future utility}}\right]. \tag{77}$$

The summation term in equation (77) is the effect on future utility from the change in habit stock brought into future periods. $\frac{\partial \tilde{s}_r}{\partial x_t}=\rho^{r-t}$ is the predicted direct effect of consumption $\tilde{x}_t$ on stock in period r. The first two terms inside brackets are the effect of that change on period r utility. The final term inside brackets accounts for the fact that the resulting change in $\tilde{x}_r$ will affect utility in later periods.

The period t decisionmaker predicts that her period $r>t$ self will maximize equation (36). The predicted period r first-order condition is

$$\left.\frac{d\tilde{U}_r\left(x_r;\tilde{s}_r\right)}{d\tilde{x}_r}\right|_{\tilde{x}_r} = 0 = \frac{\partial \tilde{u}_r}{\partial \tilde{x}_r} + \tilde{\gamma} + \delta\rho\frac{\partial V_{r+1}}{\partial \tilde{s}_{r+1}}. \tag{78}$$

Multiplying this FOC by $\tilde{\lambda}_r := \frac{\partial \tilde{x}_r}{\partial \tilde{s}_r}$ and subtracting it from the term inside brackets in equation (77) gives

$$\frac{dU_t}{dx_t} = \frac{\partial u_t}{\partial x_t} + \gamma + (1-\alpha)\sum_{r=t+1}^{T}\delta^{r-t}\rho^{r-t}\left[\begin{array}{c} \frac{\partial \tilde{u}_r}{\partial \tilde{x}_r}\tilde{\lambda}_r + \frac{\partial \tilde{u}_r}{\partial \tilde{s}_r} + \delta\rho\frac{\partial V_{r+1}}{\partial \tilde{s}_{r+1}}\tilde{\lambda}_r \\ -\left[\frac{\partial \tilde{u}_r}{\partial \tilde{x}_r}\tilde{\lambda}_r + \tilde{\gamma}\tilde{\lambda}_r + \delta\rho\frac{\partial V_{r+1}}{\partial \tilde{s}_{r+1}}\tilde{\lambda}_r\right] \end{array}\right] \tag{79}$$

$$= \frac{\partial u_t}{\partial x_t} + \gamma + (1-\alpha)\sum_{r=t+1}^{T}(\delta\rho)^{r-t}\left[\frac{\partial \tilde{u}_r}{\partial \tilde{s}_r} - \tilde{\gamma}\tilde{\lambda}_r\right] \tag{80}$$

With the quadratic functional form, this becomes

$$\frac{dU_t}{dx_t} = \eta x_t + \zeta s_t + \xi_t - p_t + \gamma + (1-\alpha)\sum_{r=t+1}^{T}(\delta\rho)^{r-t}\left[\zeta\tilde{x}_r + \phi - \tilde{\gamma}\tilde{\lambda}\right]. \tag{81}$$

In this equation, two terms (x_t and $\tilde{x}_r$) depend on x_t. x_t is by definition continuous in x_t , and $\tilde{x}_r$ is continuous in past consumption x_t due to the evolution of habit stock and Lemma 2. Thus, $\frac{dU_t}{dx_t}$ is continuous in x.

We now turn to concavity. The derivative of equation (81) is

$$\frac{d^2U_t}{dx_t^2} = \eta + (1-\alpha)\sum_{r=t+1}^{\infty}(\delta\rho)^{r-t}\zeta\frac{d\tilde{x}_r}{dx_t}. \tag{82}$$

$$= \eta + (1-\alpha)\sum_{r=t+1}^{\infty}(\delta\rho)^{r-t}\zeta\tilde{\lambda}_r\left[\rho^{r-t}\prod_{j=t+1}^{r-1}\left(1+\tilde{\lambda}_j\right)\right] \tag{83}$$

Intuitively, $\frac{d^2U_t}{dx_t^2} < 0$ requires that the diminishing marginal utility in period t outweighs the incentive to increase current consumption for the purpose of increasing future utility through ζ. This will tend to be true when projection bias α is large and/or habit formation ρ is small. A small ρ has a direct effect by causing the habit stock from dx_t to decay faster. It also has an indirect effect by reducing $\frac{d\tilde{x}_r}{dx_t}$, the perceived effect of current consumption on future consumption.

If we know an upper bound $\tilde{\lambda}^b$ such that $\tilde{\lambda}^b > \tilde{\lambda}_t$ for all t, we can write a simpler necessary condition for concavity: $\frac{d^2U_t}{dx_t^2} < 0$ for all $t \geq 0$ if

$$(1-\alpha)\sum_{r=t+1}^{\infty}(\delta\rho)^{r-t}\tilde{\lambda}_r\left[\rho^{r-t}\prod_{j=t+1}^{r-1}\left(1+\tilde{\lambda}_j\right)\right] < \frac{-\eta}{\zeta} \tag{84}$$

$$(1-\alpha)\sum_{r=t+1}^{\infty}(\delta\rho)^{r-t}\tilde{\lambda}^b\left[\rho^{r-t}\left(1+\tilde{\lambda}^b\right)^{r-t-1}\right] < \frac{-\eta}{\zeta} \tag{85}$$

$$(1-\alpha)\frac{\tilde{\lambda}^b}{1+\tilde{\lambda}^b}\cdot\sum_{r=1}^{\infty}\left(\delta\rho^2\left(1+\tilde{\lambda}^b\right)\right)^{r-1} < \frac{-\eta}{\zeta} \tag{86}$$

$$(1-\alpha)\frac{\tilde{\lambda}^b}{1+\tilde{\lambda}^b}\cdot\left[\frac{1}{1-\left(\delta\rho^2\left(1+\tilde{\lambda}^b\right)\right)}\right] < \frac{-\eta}{\zeta} \tag{87}$$

$$\frac{(1-\alpha)\tilde{\lambda}^b}{\left(1+\tilde{\lambda}^b\right)-\delta\rho^2\left(1+\tilde{\lambda}^b\right)^2} < \frac{-\eta}{\zeta}. \tag{88}$$

□

From the proof of Lemma 3, we know that $\tilde{\lambda}_t$ decreases as $t \to T$.

Finally, we show the conditions under which μ_t converges to a constant far from the time horizon.

Lemma 5. *Suppose the conditions for Lemma 2 hold, and* $-\eta > (1-\alpha)\delta\rho\left[(\zeta-\eta)\left(1+\rho\tilde{\lambda}_{t+1}\right)-\rho\zeta\right]$. *Then* $\lim_{(T-t)\to\infty}\mu_t = \mu$.

Proof. Since $\mu_t(\gamma)$ is a function of only constants, $\tilde{\lambda}_{t+1}$ (which converges per Lemma 3), and $\mu_{t+1}(\tilde{\gamma})$, it is sufficient to show that the sequence $\mu_t(\tilde{\gamma})$ converges. The coefficient on $\mu_{t+1}(\tilde{\gamma})$ in equation (61) is

$$\frac{(1-\alpha)\delta\rho(\zeta-\eta)}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1}+\zeta\right)}. \tag{89}$$

The sequence $\mu_{t+1}(\tilde{\gamma})$ will converge if and only if

$$\frac{(1-\alpha)\delta\rho(\zeta-\eta)}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1}+\zeta\right)} < 1. \tag{90}$$

The denominator is positive at our parameter values, so this inequality requires

$$-\eta > (1-\alpha)\delta\rho\left[(\zeta-\eta)\left(1+\rho\tilde{\lambda}_{t+1}\right)-\rho\zeta\right]. \tag{91}$$

In words, this requires that perceived habit formation $(1-\alpha)\rho$ is small relative to the demand slope parameter η. □

Proposition 2 combines Lemmas 2, 3, 4, and 5.

F.3 Proof of Lemma 1: Steady-State Convergence

Proof. Capital stock evolves according to $s_t = \rho\left(s_{t-1} + x_{t-1}\right)$. Substituting in the stable equilibrium strategy $x_t^* = \lambda s_t + \mu$ gives

$$
\begin{aligned}
s_t &= \rho\left(s_{t-1} + \lambda s_{t-1} + \mu\right) && (92)\\
&= \rho\mu + \rho\left(1+\lambda\right)s_{t-1} && (93)\\
&= \rho\mu + \rho\left(1+\lambda\right)\left(\rho\mu + \rho\left(1+\lambda\right)s_{t-2}\right) && (94)\\
&= \rho\mu + \rho^2\left(1+\lambda\right)\mu + \rho^2\left(1+\lambda\right)^2 s_{t-2} && (95)\\
&= \rho\mu + \rho^2\left(1+\lambda\right)\mu + \rho^3\left(1+\lambda\right)^2\mu + \rho^3\left(1+\lambda\right)^3 s_{t-3}. && (96)
\end{aligned}
$$

Thus

$$
s_t = \frac{\mu}{1+\lambda}\left(\iota + \iota^2 + \ldots + \iota^k\right) + \iota^k s_{t-k}, \tag{97}
$$

where $\iota = \left(1+\lambda\right)\rho$. Thus, provided that $\iota < 1$, in the limit as $k \to \infty$ we have

$$
\begin{aligned}
s_t &= \frac{\mu}{1+\lambda}\cdot\frac{\iota}{1-\iota} && (98)\\
&= \frac{\mu\rho}{1-\left(1+\lambda\right)\rho}. && (99)
\end{aligned}
$$

We can then check that this is indeed a steady state:

$$
\begin{aligned}
s_t &= \rho\left(\frac{\mu\rho}{1-\left(1+\lambda\right)\rho} + \mu + \lambda\left(\frac{\mu\rho}{1-\left(1+\lambda\right)\rho}\right)\right) && (100)\\
&= \rho\left(\frac{\mu\rho + \mu\left(1-\left(1+\lambda\right)\rho\right) + \lambda\mu\rho}{1-\left(1+\lambda\right)\rho}\right) && (101)\\
&= \rho\left(\frac{\mu\rho + \mu - \mu\rho - \mu\lambda\rho + \lambda\mu\rho}{1-\left(1+\lambda\right)\rho}\right) && (102)\\
&= \frac{\mu\rho}{1-\left(1+\lambda\right)\rho} && (103)
\end{aligned}
$$

□

F.4 Proof of Proposition 3: Steady-State Consumption

Proof. We assume steady state implies constant consumption and habit stock, but not necessarily constant *predicted* consumption and habit stock. In steady state, $p_t = p$, $\xi_t = \xi$, $s_t = s_{ss}$, and $x_t = x_{ss}$. By equation (1) governing the evolution of habit stock, $s_{ss} = \rho(s_{ss} + x_{ss})$, and re-arranging this equation gives $s_{ss} = \frac{\rho}{1-\rho} x_{ss}$. Earlier, we defined steady-state misprediction as $m_{ss} := \tilde{x}_{t+1} - x_{ss}$.

We substitute $p_t = p$, $\xi_t = \xi$, $s_t = s_{ss}$, and $x_t = x_{ss}$ into the Euler equation (equation (18)), giving

$$\eta x_{ss} + \zeta s_{ss} + \xi - p + \gamma = (1-\alpha)\delta\rho \left[\eta \tilde{x}_{t+1} + \zeta\rho\left(x_{ss} + s_{ss}\right) + \xi - p + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \zeta\tilde{x}_{t+1} - \phi \right]. \tag{104}$$

Substituting in $s_{ss} = \frac{\rho}{1-\rho} x_{ss}$ and also writing predicted consumption as a deviation from the actual value gives

$$\eta x_{ss} + \xi - p + \frac{\rho\zeta}{1-\rho} x_{ss} + \gamma = (1-\alpha)\delta\rho \left[(\eta-\zeta)\left((\tilde{x}_{t+1} - x_{ss}) + x_{ss}\right) + \zeta\rho\left(\frac{1}{1-\rho} x_{ss}\right) + \xi - p + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \phi \right]. \tag{105}$$

Substituting $m_{ss} := \tilde{x}_{t+1} - x_{ss}$ and collecting terms gives

$$\begin{aligned} x_{ss}\left[\eta + \frac{\rho\zeta}{1-\rho} - (1-\alpha)\delta\rho\left((\eta-\zeta) + \frac{\zeta\rho}{1-\rho}\right)\right] = & \; p - \xi - \gamma + (1-\alpha)\delta\rho\Big[(\eta-\zeta)m_{ss} \\ & + \xi - p + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \phi\Big] \end{aligned} \tag{106}$$

$$\begin{aligned} x_{ss}\left[\eta - (1-\alpha)\delta\rho(\eta-\zeta) + \zeta\frac{\rho - (1-\alpha)\delta\rho^2}{1-\rho}\right] = & \; (1-(1-\alpha)\delta\rho)(p-\xi) + (1-\alpha)\delta\rho\Big[(\eta-\zeta)m_{ss} \\ & + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \phi\Big] - \gamma. \end{aligned} \tag{107}$$

Multiplying both sides by (-1), setting $\kappa := (1-\alpha)\delta\rho(\phi - \xi) + \xi$, and dividing through gives equation (21).

□

G Derivations of Estimating Equations in Appendix E.3

We define $y^g := \mathbb{E}_{i\in g} y_i$ as the expectation over individuals in group g of parameter y. Due to random assignment, $\xi_t^g = \xi_t^{g'}$ and $s_2^g = s_2^{g'}$ for all $\{g, g'\}$, and $\mu_t^B = \mu_t^{BC}$ for $t \in \{2,4,5\}$. The estimating equations for the restricted model in Section 6.2 are the below equations with the additional assumptions that $\tau_2^B = 0$ and $\alpha = 1$.

G.1 Habit Formation

Derivation of equation (22). From equation (19) and the evolution of habit stock, we have

$$
\begin{aligned}
x_4^* &= \lambda s_4 + \mu_4 && (108)\\
&= \lambda \rho \left(s_3 + x_3^*\right) + \mu_4 && (109)\\
&= \lambda \rho \left(\rho \left(s_2 + x_2^*\right) + x_3^*\right) + \mu_4. && (110)
\end{aligned}
$$

Thus, group average consumption is $x_4^g = \lambda \left(\rho^2 \left(s_2^g + x_2^g\right) + \rho x_3^g\right) + \mu_4^g$, and the period 4 bonus effect is

$$
\tau_4^B = \lambda \left(\rho^2 \tau_2^B + \rho \tau_3^B\right). \quad (111)
$$

Re-arranging gives equation (22).

Derivation of equation (23). Similarly, we have

$$
\begin{aligned}
x_5^* &= \lambda s_5 + \mu_5 && (112)\\
&= \lambda \rho \left(s_4 + x_4^*\right) + \mu_5 && (113)\\
&= \lambda \rho \left(\rho \left(s_3 + x_3^*\right) + x_4^*\right) + \mu_5 && (114)\\
&= \lambda \rho \left(\rho \left(\rho \left(s_2 + x_2^*\right) + x_3^*\right) + x_4^*\right) + \mu_5. && (115)
\end{aligned}
$$

Thus, group average consumption is $x_5^g = \lambda \left(\rho^3 \left(s_2^g + x_2^g\right) + \rho^2 x_3^g + \rho x_4^g\right) + \mu_5^g$, and the period 5 bonus effect is

$$
\tau_5^B = \lambda \left(\rho^3 \tau_2^B + \rho^2 \tau_3^B + \rho \tau_4^B\right). \quad (116)
$$

Multiplying equation (111) by ρ and subtracting from equation (116) gives $\tau_5^B - \tau_4^B \rho = \lambda \rho \tau_4^B$, and re-arranging gives equation (23).

System of equations for λ and ρ. Re-arranging equation (23) gives

$$
\lambda = \frac{\tau_5^B}{\tau_4^B \rho} - 1. \quad (117)
$$

Substituting this into equation (22) gives:

$$
\begin{aligned}
\frac{\tau_5^B - \tau_4^B \rho}{\tau_4^B \rho} &= \frac{\tau_4^B}{\rho \tau_3^B + \rho^2 \tau_2^B} && (118)\\
\left(\tau_4^B\right)^2 &= \left(\tau_5^B - \tau_4^B \rho\right)\left(\tau_3^B + \rho \tau_2^B\right) && (119)\\
0 &= \left[\tau_2^B \tau_4^B\right] \rho^2 + \left[\tau_3^B \tau_4^B - \tau_2^B \tau_5^B\right] \rho + \left[\left(\tau_4^B\right)^2 - \tau_3^B \tau_5^B\right]. && (120)
\end{aligned}
$$

The quadratic formula gives

$$\rho = \frac{-\left[\tau_3^B\tau_4^B - \tau_2^B\tau_5^B \pm \sqrt{\left[\tau_3^B\tau_4^B - \tau_2^B\tau_5^B\right]^2 - 4\left[\tau_2^B\tau_4^B\right]\left[\left(\tau_4^B\right)^2 - \tau_3^B\tau_5^B\right]}\right]}{2\left[\tau_2^B\tau_4^B\right]}. \tag{121}$$

In all bootstrap draws in our data, only one of the two solutions satisfies the requirement that $\rho \geq 0$.

Special case with $\tau_2^B = 0$. If there is no anticipatory demand response ($\tau_2^B = 0$), we have $\tau_4^B = \lambda\rho\tau_3^B$ and $\tau_5^B = \lambda\rho^2\tau_3^B + \lambda\rho\tau_4^B$. Dividing the two equations gives

$$\begin{aligned} \frac{\tau_5^B}{\tau_4^B} &= \rho + \frac{\tau_4^B}{\tau_3^B} \\ \rho &= \frac{\tau_5^B}{\tau_4^B} - \frac{\tau_4^B}{\tau_3^B}. \end{aligned} \tag{122}$$

We then solve for λ by inserting this ρ into equation (22) with $\tau_2^B = 0$.

G.2 Perceived Habit Formation, Price Response, and Habit Stock Effect on Marginal Utility

The expectation over i of the Euler equations for group g is

$$\eta x_t^g + \zeta s_t^g + \xi_t^g - p_t + \gamma = (1-\alpha)\delta\rho\left[\eta\tilde{x}_{t+1}^g + \zeta\tilde{s}_{t+1}^g + \xi_{t+1}^g - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \left(\zeta\tilde{x}_{t+1}^g + \phi\right)\right] \quad . \tag{123}$$

Derivation of equation (24). Differencing the Euler equations for periods 2 versus 3 for the Bonus and Bonus Control groups gives

$$\eta\tau_2^B = (1-\alpha)\delta\rho\left[-p^B + (\eta-\zeta)\left(\tilde{x}_3^B - \tilde{x}_3^{BC}\right) + \zeta\left(\tilde{s}_3^B - \tilde{s}_3^{BC}\right)\right]. \tag{124}$$

Substituting $\tilde{x}_3^B - \tilde{x}_3^{BC} = \tilde{\tau}_3^B$ and $\tilde{s}_3^B - \tilde{s}_3^{BC} = \rho\tau_2^B$ gives

$$\eta\tau_2^B = (1-\alpha)\delta\rho\left[-p^B + (\eta-\zeta)\tilde{\tau}_3^B + \zeta\rho\tau_2^B\right]. \tag{125}$$

Rearranging gives equation (24).

If $\tilde{\gamma} \neq \gamma$, then people update their predictions of $\tilde{x}_3$ as they set x_2^*, and thus the predictions of $\tilde{x}_3$ from survey 2 are inconsistent with x_2^*. However, there is only limited misprediction in our data, so this is not very consequential.

Derivation of equation (25). Differencing the Euler equations for periods 3 versus 4 for the Bonus and Bonus Control groups gives

$$\left(-p^B-0\right)+\eta\tau_3^B+\zeta\left(s_3^B-s_3^{BC}\right)=(1-\alpha)\delta\rho\left[(\eta-\zeta)\left(\tilde{x}_4^B-\tilde{x}_4^{BC}\right)+\zeta\left(\tilde{s}_4^B-\tilde{s}_4^{BC}\right)\right]. \tag{126}$$

Habit stock evolution implies $s_3^B-s_3^{BC}=\rho(s_2^B-s_2^{BC}+x_2^B-x_2^{BC})=\rho\tau_2^B$ and $\tilde{s}_4^B-\tilde{s}_4^{BC}=\rho\left(s_3^B-s_3^{BC}+x_3^B-x_3^{BC}\right)=\rho\left(\rho\tau_2^B+\tau_3^B\right)$. Linear policy functions imply $\tilde{x}_4=\tilde{\lambda}\tilde{s}_4+\tilde{\mu}_4$, so $\tilde{x}_4^B-\tilde{x}_4^{BC}=\tilde{\lambda}\left(\tilde{s}_4^B-\tilde{s}_4^{BC}\right)$. Substituting these equations gives

$$\left(-p^B-0\right)+\eta\tau_3^B+\zeta\rho\tau_2^B=(1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\rho\left(\rho\tau_2^B+\tau_3^B\right)\right]. \tag{127}$$

Rearranging gives

$$\eta\left(\tau_3^B-(1-\alpha)\delta\rho^2\tilde{\lambda}\left(\rho\tau_2^B+\tau_3^B\right)\right)=p^B-\zeta\rho\tau_2^B+(1-\alpha)\delta\rho^2\zeta\left(1-\tilde{\lambda}\right)\left(\rho\tau_2^B+\tau_3^B\right). \tag{128}$$

Solving for η gives equation (25).

Derivation of equation (26). Differencing the Euler equations for periods 4 versus 5 for the Bonus and Bonus Control groups gives

$$\eta\left(x_4^B-x_4^{BC}\right)+\zeta\left(s_4^B-s_4^{BC}\right)=(1-\alpha)\delta\rho\left[(\eta-\zeta)\left(\tilde{x}_5^B-\tilde{x}_5^{BC}\right)+\zeta\left(\tilde{s}_5^B-\tilde{s}_5^{BC}\right)\right] \tag{129}$$

Habit stock evolution implies $s_4^B-s_4^{BC}=\rho\left(s_3^B-s_3^{BC}+x_3^B-x_3^{BC}\right)=\rho^2\tau_2^B+\rho\tau_3^B$ and $\tilde{s}_5^B-\tilde{s}_5^{BC}=\rho\left(s_4^B-s_4^{BC}+x_4^B-x_4^{BC}\right)=\rho\left(\rho^2\tau_2^B+\rho\tau_3^B+\tau_4^B\right)$. Linear policy functions imply $\tilde{x}_5=\tilde{\lambda}\tilde{s}_5+\tilde{\mu}_5$, so $\tilde{x}_5^B-\tilde{x}_5^{BC}=\tilde{\lambda}\left(\tilde{s}_5^B-\tilde{s}_5^{BC}\right)$. Substituting these equations gives

$$\eta\tau_4^B+\zeta\left(\rho^2\tau_2^B+\rho\tau_3^B\right)=(1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\rho\left(\rho^2\tau_2^B+\rho\tau_3^B+\tau_4^B\right)\right] \tag{130}$$

$$=(1-\alpha)\delta\rho^2\left[\left(\eta\lambda+\zeta\left(1-\tilde{\lambda}\right)\right)\left(\rho^2\tau_2^B+\rho\tau_3^B+\tau_4^B\right)\right]. \tag{131}$$

Collecting ζ terms gives

$$\zeta\left(\rho\tau_3^B+\rho^2\tau_2^B\right)-(1-\alpha)\delta\rho^2\left[\zeta\left(1-\tilde{\lambda}\right)\left(\rho^2\tau_2^B+\rho\tau_3^B+\tau_4^B\right)\right]=-\eta\tau_4^B+(1-\alpha)\delta\rho^2\eta\tilde{\lambda}\left(\rho^2\tau_2^B+\rho\tau_3^B+\tau_4^B\right). \tag{132}$$

Solving for ζ gives equation (26).

System of equations for $(1-\alpha)$, η, and ζ.

First, we solve explicitly for $(1-\alpha)$ before substituting it back in Equations (25) and (26) to solve for η and ζ.

We define

$$y := \frac{-\tau_4^B + (1-\alpha)\delta\rho^2\lambda\left[\rho^2\tau_2^B + \rho\tau_3^B + \tau_4^B\right]}{\rho\tau_3^B + \rho^2\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left[\rho^2\tau_2^B + \rho\tau_3^B + \tau_4^B\right]}. \tag{133}$$

Observe that

$$\zeta = \eta \cdot y. \tag{134}$$

We can use this observation to rearrange Equation (25):

$$\eta = \frac{p^B - \zeta\rho\tau_2^B + (1-\alpha)\delta\rho^2\zeta(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)}{\tau_3^B - (1-\alpha)\delta\rho^2\lambda\left(\rho\tau_2^B + \tau_3^B\right)} \tag{135}$$

$$\eta\left[\tau_3^B - (1-\alpha)\delta\rho^2\lambda\left(\rho\tau_2^B + \tau_3^B\right)\right] = p^B - \zeta\left(\rho\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)\right) \tag{136}$$

$$= p^B - \eta\cdot y\left(\rho\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)\right) \tag{137}$$

$$p^B = \eta\left[\tau_3^B - (1-\alpha)\delta\rho^2\lambda\left(\rho\tau_2^B + \tau_3^B\right) + y\left(\rho\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)\right)\right]. \tag{138}$$

Then, define

$$x := \tau_3^B - (1-\alpha)\delta\rho^2\lambda\left(\rho\tau_2^B + \tau_3^B\right) + y\left(\rho\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)\right) \tag{139}$$

where we observe that

$$\eta = \frac{p^B}{x}, \tag{140}$$

and

$$\zeta = \frac{p^B y}{x}. \tag{141}$$

Finally, we get that

$$(1-\alpha) = \frac{\eta\tau_2^B}{\delta\rho\left[-p^B + (\eta-\zeta)\tilde{\tau}_3^B + \zeta\rho\tau_2^B\right]} \tag{142}$$

$$= \frac{\frac{p^B}{x}\tau_2^B}{\delta\rho\left[-p^B + (\frac{p^B}{x} - \frac{p^B y}{x})\tilde{\tau}_3^B + \frac{p^B y}{x}\rho\tau_2^B\right]}. \tag{143}$$

Since all scalars are known in the last equation, we can now solve for α. Then, we can estimate η and ζ by substituting α in Equations (25) and (26) respectively.

G.3 Naivete about Temptation

Derivation of equation (27). The Euler equation *predicted* for period t on the survey at the beginning of period t is

$$\eta x_t^*(s_t,\tilde{\gamma},\boldsymbol{p}_t)+\zeta s_t+\xi_t-p_t+\tilde{\gamma}=(1-\alpha)\delta\rho\left[\eta\tilde{x}_{t+1}+\zeta s_{t+1}+\xi_{t+1}-p_{t+1}+\tilde{\gamma}+\tilde{\gamma}\tilde{\lambda}-(\zeta\tilde{x}_{t+1}+\phi)\right]. \tag{144}$$

This equation uses the assumption that consumers are aware of period t projection bias when predicting period t consumption on survey t, so the only reason why the period t survey-taker mispredicts the period t objective function is naivete about period t temptation.

Habit stock evolution implies $\tilde{s}_{t+1}=\rho(s_t+\tilde{x}_t)$. Linear policy functions imply $\tilde{x}_{t+1}=\tilde{\lambda}\tilde{s}_{t+1}+\tilde{\mu}_{t+1}$. Substituting these equations into the predicted Euler equation gives

$$\eta x_t^*(s_t,\tilde{\gamma},\boldsymbol{p}_t)+\zeta s_t+\xi_t-p_t+\tilde{\gamma}=(1-\alpha)\delta\rho\left[(\eta-\zeta)\left(\tilde{\lambda}\tilde{s}_{t+1}+\tilde{\mu}_{t+1}\right)+\zeta\tilde{s}_{t+1}+\xi_{t+1}-p_{t+1}+\tilde{\gamma}+\tilde{\gamma}\tilde{\lambda}-\phi\right]. \tag{145}$$

$$=(1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\tilde{s}_{t+1}+(\eta-\zeta)\tilde{\mu}_{t+1}+\xi_{t+1}-p_{t+1}+\tilde{\gamma}+\tilde{\gamma}\tilde{\lambda}-\phi\right] \tag{146}$$

$$=(1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\rho(s_t+\tilde{x}_t)+(\eta-\zeta)\tilde{\mu}_{t+1}+\xi_{t+1}-p_{t+1}+\tilde{\gamma}+\tilde{\gamma}\tilde{\lambda}-\phi\right]. \tag{147}$$

Analogously, the actual Euler equation for period t can be written as

$$\eta x_t^*(s_t,\gamma,\boldsymbol{p}_t)+\zeta s_t+\xi_t-p_t+\gamma=(1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\rho(s_t+x_t^*)+(\eta-\zeta)\tilde{\mu}_{t+1}+\xi_{t+1}-p_{t+1}+\tilde{\gamma}+\tilde{\gamma}\tilde{\lambda}-\phi\right]. \tag{148}$$

Differencing the actual and predicted Euler equations for period t versus period $t+1$ for the Control group gives

$$\eta\left(x_t^C-\tilde{x}_t^C\right)+\gamma-\tilde{\gamma}=(1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\rho\left(x_t^C-\tilde{x}_t^C\right)\right] \tag{149}$$

Solving for $\gamma-\tilde{\gamma}$ and substituting $m^C=x_t^C-\tilde{x}_t^C$ gives equation (27).

G.4 Temptation

Limit effect: derivation of equation (28). Consider a "zero temptation" intervention that fully eliminates both perceived and actual temptation starting in period 2, generating treatment effects τ_t^0. Differencing the average Euler equations for periods 2 versus 3 for the zero temptation group versus its control group gives

$$\eta\left(x_2^0-x_2^{0C}\right)-\gamma=(1-\alpha)\delta\rho\left[(\eta-\zeta)\left(\tilde{x}_3^0-\tilde{x}_3^{0C}\right)+\zeta\left(\tilde{s}_3^0-\tilde{s}_3^{0C}\right)-\tilde{\gamma}-\tilde{\gamma}\tilde{\lambda}\right] \tag{150}$$

$$\eta\tau_2^0-\gamma=(1-\alpha)\delta\rho\left[(\eta-\zeta)\tilde{\tau}_3^0+\zeta\rho\tau_2^0-\tilde{\gamma}-\tilde{\gamma}\tilde{\lambda}\right] \tag{151}$$

Solving for γ and substituting $\tau^0=\tau^L/\omega$ gives equation (28).

To solve for γ as a function of data and known parameters, we solve equation (27) for $\tilde{\gamma}$, substitute into equation (28), and rearrange, giving

$$\gamma=\frac{\eta\tau_2^L/\omega-(1-\alpha)\delta\rho\left(\left[(\eta-\zeta)\tilde{\tau}_3^L/\omega+\zeta\rho\tau_2^L/\omega\right]+(1+\tilde{\lambda})m_2^C\cdot\left[-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right]\right)}{1-(1-\alpha)\delta\rho(1+\tilde{\lambda})}. \tag{152}$$

Bonus valuation: derivation of equation (29). When we elicited the bonus valuation on survey 2, we had not yet told participants whether the bonus would be in effect for period 2 or 3. The theoretical valuations for a period 2 vs. period 3 bonus are identical if we assume that consumers predict no anticipatory effect of the period 3 bonus. Otherwise, this derivation would need to account for the period 2 survey taker's valuation of the perceived internality reduction from the anticipatory effect. Since the actual bonus was for period 3, we focus the derivation on that case and maintain the assumption of zero predicted anticipatory effect.

From the perspective of the period 2 survey taker, the predicted period 3 continuation value (given naivete about future projection bias) as a function of predicted habit stock and period 3 price is

$$V_3\left(\tilde{s}_3,p_3\right)=u_3\left(\tilde{x}_3^*\left(\tilde{s}_3,\tilde{\gamma},\boldsymbol{p}_3\right);\tilde{s}_3,p_3\right)+\delta V_4\left(\tilde{s}_4,\cdot\right). \tag{153}$$

The change in that predicted continuation value from a marginal change in period 3 price is

$$\frac{dV_3(\tilde{s}_3,p_3)}{dp_3}=\frac{\partial\tilde{u}_3}{\partial p_3}+\frac{\partial\tilde{x}_3}{\partial p_3}\left[\frac{\partial\tilde{u}_3}{\partial\tilde{x}_3}+\delta\frac{dV_4\left(\tilde{s}_4,\cdot\right)}{d\tilde{s}_4}\frac{\partial\tilde{s}_4}{\partial\tilde{x}_3}\right]. \tag{154}$$

People taking survey 2 predict that their period 3 selves will set x_3 to maximize that same function with an additional $\tilde{\gamma}x_3$ in period 3 flow utility:

$$\tilde{x}_3^*(\tilde{s}_3,\tilde{\gamma},\boldsymbol{p}_3)=\arg\max_{x_3}u_3\left(x_3;\tilde{s}_3,\boldsymbol{p}_3\right)+\tilde{\gamma}x_3+\delta V_4\left(\tilde{s}_4,\cdot\right). \tag{155}$$

Thus, people taking survey 2 predict that they will set x_3 such that

$$\frac{\partial \tilde{u}_3}{\partial x_3} + \tilde{\gamma} + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{x}_3} = 0. \tag{156}$$

Substituting equation (156) into equation (154) gives

$$\frac{dV_3(\tilde{s}_3, p_3)}{dp_3} = \frac{\partial \tilde{u}_3}{\partial p_3} - \tilde{\gamma}\frac{\partial \tilde{x}_3}{\partial p_3} \tag{157}$$

$$= -\tilde{x}_3(p_3) - \tilde{\gamma}\frac{\partial \tilde{x}_3}{\partial p_3}. \tag{158}$$

This illustrates a temptation-adjusted envelope theorem: the effect of a marginal price change on the long-run self's utility (given perceived misoptimization from the long-run self's perspective) equals the mechanical effect $\tilde{x}_3(p_3)$ adjusted by the magnitude of the perceived misoptimization $\tilde{\gamma}\frac{\partial \tilde{x}_3}{\partial p_3}$. With zero perceived temptation ($\tilde{\gamma} = 0$), this reduces to the standard envelope theorem. The derivation for a period 2 bonus would be analogous, except with $(1-\alpha)$ multiplying the predicted period 3 continuation value in both the survey taker's objective function and the predicted period 2 objective function.

We integrate over equation (158) to determine the effect of a non-marginal price increase from 0 to p^B:

$$V_3\left(\tilde{s}_3, p_3 = p_3^B\right) - V_3\left(\tilde{s}_3, p_3 = 0\right) = \int_{p_3=0}^{p_3=p_3^B} -\tilde{x}_3(p_3) - \tilde{\gamma}\frac{\partial \tilde{x}_3}{\partial p_3}\, dp_3 \tag{159}$$

$$= -p_3^B \cdot \left(\tilde{x}_3(p_3^B) + \tilde{x}_3(0)\right)/2 - \tilde{\gamma} \cdot \left(\tilde{x}_3(p_3^B) - \tilde{x}_3(0)\right), \tag{160}$$

where the second line follows from the fact that demand is linear in price, which was shown in Proposition 2.

Limit valuation: derivation of equation (31). The period 3 survey-taker's objective function is

$$V_3\left(s_3, \tilde{\gamma}_3\right) = u_3\left(x_3^*\left(s_3, \tilde{\gamma}_3, \boldsymbol{p}_3\right); s_3, p_3\right) + \begin{array}{c} \alpha \sum_{r=4}^{T} \delta^{r-3} u_r\left(\tilde{x}_r^*\left(s_3, \tilde{\gamma}, \boldsymbol{p}_r\right); s_3, p_r\right) \\ +(1-\alpha)\delta V_4\left(\tilde{s}_4, \cdot\right) \end{array}. \tag{161}$$

This equation uses the assumption that the survey taker is projection biased.

The change in that objective function from a marginal change in perceived period 3 temptation is

$$\frac{dV_3\left(s_3, \tilde{\gamma}_3\right)}{d\tilde{\gamma}_3} = \frac{\partial x_3^*\left(s_3, \tilde{\gamma}_3, \boldsymbol{p}_3\right)}{\partial \tilde{\gamma}_3}\left[\frac{\partial u_3}{\partial x_3} + (1-\alpha)\delta\frac{\partial V_4\left(\tilde{s}_4, \cdot\right)}{\partial \tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{x}_3}\right]. \tag{162}$$

People taking survey 3 predict that they will set x_3^* such that

$$\frac{\partial u_3}{\partial x_3} + \tilde{\gamma}_3 + (1-\alpha)\delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4} \frac{\partial \tilde{s}_4}{\partial \tilde{x}_3} = 0. \tag{163}$$

Substituting the period 3 first-order condition from equation (163) into equation (162) gives

$$\frac{dV_3(s_3, \tilde{\gamma}_3)}{d\tilde{\gamma}_3} = -\tilde{\gamma}_3 \frac{\partial x_3^*(s_3, \tilde{\gamma}_3, \boldsymbol{p}_3)}{\partial \tilde{\gamma}_3}. \tag{164}$$

We integrate over equation (164) to determine the effect of the non-marginal temptation reduction from $\tilde{\gamma}$ to $(1-\omega)\tilde{\gamma}$:

$$v^L = V_3(s_3, \tilde{\gamma}_3 = (1-\omega)\tilde{\gamma}) - V_3(s_3, \tilde{\gamma}_3 = \tilde{\gamma}) = \int_{\tilde{\gamma}_3=\tilde{\gamma}}^{\tilde{\gamma}_3=(1-\omega)\tilde{\gamma}} -\tilde{\gamma}_3 \frac{\partial x_3^*(\tilde{\gamma}_3)}{\partial \tilde{\gamma}_3}\, d\tilde{\gamma}_3. \tag{165}$$

$$= (x_3^*(\tilde{\gamma}) - x_3^*(0)) \cdot \tilde{\gamma} \cdot \frac{1}{2} - (1-\omega)^2 \cdot (\tilde{x}_3(\tilde{\gamma}) - \tilde{x}_3(0)) \cdot \tilde{\gamma} \cdot \frac{1}{2} \tag{166}$$

$$= (x_3^*(\tilde{\gamma}) - x_3^*(0)) \cdot \tilde{\gamma} \cdot \left(1 - (1-\omega)^2\right) \cdot \frac{1}{2} \tag{167}$$

$$= (x_3^*(\tilde{\gamma}) - x_3^*(0)) \cdot \tilde{\gamma} \cdot \frac{\omega(2-\omega)}{2} \tag{168}$$

$$= \frac{(x_3^*(\tilde{\gamma}) - x_3^*((1-\omega)\tilde{\gamma}))}{\omega} \cdot \tilde{\gamma} \cdot \frac{\omega(2-\omega)}{2} \tag{169}$$

$$= (x_3^*(\tilde{\gamma}) - x_3^*((1-\omega)\tilde{\gamma})) \cdot \tilde{\gamma} \cdot \frac{(2-\omega)}{2}, \tag{170}$$

where the second line is the area of the long-run self's perceived deadweight loss reduction trapezoid (following from linear demand) and the fifth line follows from the assumption that $\tilde{\tau}^L/\omega = \tilde{\tau}^0$.

G.5 Temptation with Multiple Goods

We now extend our model to include a second temptation good y, which in our experiment is FITSBY use on other devices. Habit stock now evolves according to $s_{t+1} = \rho(s_t + x_t + y_t)$. Before period t, consumers now consider flow utility to be $u_t(x_t, y_t; s_t, p_t)$. In period t, consumers choose as if period t flow utility is $u_t(x_t, y_t; s_t, p_t) + \gamma_x x_t + \gamma_y y_t$. Before period t, consumers predict that they will choose as if period t flow utility is $u_t(x_t, y_t; s_t, p_t) + \tilde{\gamma}_x x_t + \tilde{\gamma}_y y_t$. x is still sold at price p_t, while y_t has zero price. The limit treatment fully eliminates perceived and actual temptation on x.

We derive new equations for γ or $\tilde{\gamma}$ for the limit effect, bonus valuation, and limit valuation strategies. With all three strategies, if y is not a temptation good ($\tilde{\gamma}_y = \gamma_y = 0$) or if y is neither a substitute nor a complement for x, then our original estimating equations are unaffected.

Limit effect. To derive γ using the limit effect strategy, we assume full projection bias ($\alpha = 1$). We assume that the static quadratic flow utility function is now

$$u(x,y;p) = \frac{\eta_x}{2}x^2 + \xi_x x - px + \sigma xy + \frac{\eta_y}{2}y^2 + \xi_y y. \tag{171}$$

Without the limit, consumers maximize $u(x,y;p) + \gamma_x x + \gamma_y y$, giving

$$y^*(x) = \frac{\sigma x + \xi_y + \gamma_y}{-\eta_y} \tag{172}$$

$$x^* = \frac{\xi_x - p + \sigma \frac{\xi_y + \gamma_y}{-\eta_y} + \gamma_x}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{173}$$

Taking the expectation over individuals, the bonus effect on x^* is

$$\tau_x^B = \frac{p^B}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{174}$$

The limit allows consumers to set x_L before period t. When setting the limit, consumers predict that in period t they will set y conditional on x_L to maximize $u(x_L, y; p) + \tilde{\gamma}_x x_L + \tilde{\gamma}_y y$, giving

$$y^*(x_L) = \frac{\sigma x_L + \xi_y + \tilde{\gamma}_y}{-\eta_y}. \tag{175}$$

Consumers thus set x_L to maximize $u(x_L, y^*(x_L); p)$, giving

$$x_L = \frac{\xi_x - p + \xi_y \frac{\sigma}{-\eta_y}}{-\eta_x + \frac{\sigma^2}{\eta_y}}. \tag{176}$$

The effect of the limit on y is $y^*(x_L) - y^*(x^*) = \frac{\sigma x_L + \xi_y + \tilde{\gamma}_y}{-\eta_y} - \frac{\sigma x^* + \xi_y + \gamma_y}{-\eta_y}$. Taking the expectation over individuals, the limit effect on y is

$$\tau_y^L = \frac{\sigma}{-\eta_y}\tau_x^L \tag{177}$$

The effect of the limit on x is

$$x_L - x^* = \frac{\xi_x - p + \xi_y \frac{\sigma}{-\eta_y}}{-\eta_x + \frac{\sigma^2}{\eta_y}} - \frac{\xi_x - p + \sigma \frac{\xi_y + \gamma_y}{-\eta_y} + \gamma_x}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{178}$$

$$= \frac{\frac{\sigma}{-\eta_y}\gamma_y + \gamma_x}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{179}$$

$$= \frac{-\gamma\left(1 + \frac{\sigma}{-\eta_y}\right)}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{180}$$

where the third line assumes $\gamma_x = \gamma_y = \gamma$.

Taking the expectation over individuals and substituting equations (174) and (177) gives

$$\tau_x^L = \frac{-\gamma\left(1 + \frac{\tau_y^L}{\tau_x^L}\right)}{p^B / \tau_x^B}. \tag{181}$$

Rearranging gives

$$\gamma = \frac{\tau_x^L \cdot \left(p^B / \tau_x^B\right)}{1 + \frac{\tau_y^L}{\tau_x^L}}. \tag{182}$$

This exactly parallels equation (28) for the $\alpha = 1$ case, except adjusting the denominator for substitution. If x and y are substitutes, then the estimated γ increases: more temptation is required to explain a given limit when the consumer knows that she can evade the limit through substitution to another temptation good. If x and y are complements, then the estimated γ decreases: less temptation is needed to explain a given limit when the consumer knows that the limit will also cause reductions in another temptation good.

Bonus valuation. The derivation for the bonus valuation with substitute goods is very similar to the one-good case. The change in the period 3 continuation value function from a marginal change in p_3 is

$$\frac{dV_3(\tilde{s}_3, p_3)}{dp_3} = \frac{\partial \tilde{u}_3}{\partial p_3} + \frac{\partial \tilde{x}_3}{\partial p_3}\left[\frac{\partial \tilde{u}_3}{\partial \tilde{x}_3} + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{x}_3}\right] + \frac{\partial \tilde{y}_3}{\partial p_3}\left[\frac{\partial \tilde{u}_3}{\partial \tilde{y}_3} + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{y}_3}\right]. \tag{183}$$

People taking survey 2 predict that they will set x_3 and y_3 according to

$$\frac{\partial \tilde{u}_3}{\partial x_3} + \tilde{\gamma}_x + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{x}_3} = 0 \tag{184}$$

$$\frac{\partial \tilde{u}_3}{\partial y_3} + \tilde{\gamma}_y + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{y}_3} = 0. \tag{185}$$

Substituting equations (184) and (185) as well as $\frac{\partial \tilde{u}_3}{\partial p_3} = -\tilde{x}_3(p_3)$ into equation (183) gives

$$\frac{dV_3(\tilde{s}_3, p_3)}{dp_3} = -\tilde{x}_3(p_3) - \tilde{\gamma}_x \frac{\partial \tilde{x}_3}{\partial p_3} - \tilde{\gamma}_y \frac{\partial \tilde{y}_3}{\partial p_3}. \tag{186}$$

Integrating over a non-marginal price increase from 0 to p^B assuming linear demand, also assuming $\tilde{\gamma}_x = \tilde{\gamma}_y = \tilde{\gamma}$, taking the expectation over participants, and rearranging gives

$$\tilde{\gamma} = \frac{\bar{v}^B - \bar{F}^B + p_3^B \bar{\tilde{x}}_3^{B+BC}}{-\left(\bar{\tau}_{x3}^B + \bar{\tau}_{y3}^B\right)} \tag{187}$$

This exactly parallels equation (30), except adjusting the denominator for substitution. The survey taker values the total temptation reduction $-\left(\bar{\tau}_{x3}^B + \bar{\tau}_{y3}^B\right)$ induced by the bonus. If x and y are substitutes, the total temptation reduction is lower, and more temptation is needed to justify a given valuation. If x and y are complements, the total temptation reduction is higher, and less temptation is needed to justify a given valuation.

Limit valuation. The derivation for the limit valuation with substitute goods is also similar to the one-good case. The change in the period 3 survey-taker's objective function from a marginal change in perceived period 3 temptation for good x only is

$$\frac{dV_3(s_3, \tilde{\gamma}_{x3})}{d\tilde{\gamma}_{x3}} = \frac{\partial x_3^*}{\partial \tilde{\gamma}_{x3}} \left[\frac{\partial u_3}{\partial x_3} + (1-\alpha)\delta \frac{\partial V_4(\tilde{s}_4, \cdot)}{\partial \tilde{s}_4} \frac{\partial \tilde{s}_4}{\partial \tilde{x}_3}\right] + \frac{\partial y_3^*}{\partial \tilde{\gamma}_{x3}} \left[\frac{\partial u_3}{\partial y_3} + (1-\alpha)\delta \frac{\partial V_4(\tilde{s}_4, \cdot)}{\partial \tilde{s}_4} \frac{\partial \tilde{s}_4}{\partial \tilde{y}_3}\right]. \tag{188}$$

Substituting the predicted period 3 first-order conditions for x and y gives

$$\frac{dV_3(s_3, \tilde{\gamma}_{x3})}{d\tilde{\gamma}_{x3}} = -\tilde{\gamma}_{x3} \frac{\partial x_3^*}{\partial \tilde{\gamma}_{x3}} - \tilde{\gamma}_y \frac{\partial y_3^*}{\partial \tilde{\gamma}_{x3}}. \tag{189}$$

Integrating over this from $\tilde{\gamma}_x$ to $(1-\omega)\tilde{\gamma}_x$ assuming linear demand, also assuming $\tilde{\gamma}_x = \tilde{\gamma}_y = \tilde{\gamma}$, taking the expectation over participants, and rearranging gives

$$\tilde{\gamma} = \frac{\bar{v}^L}{-\left(\bar{\tau}_3^L(2-\omega)/2 + \bar{\tau}_{y3}^L\right)}. \tag{190}$$

As with the bonus valuation, the survey taker values the total temptation deadweight loss reduction induced by the limit. If x and y are substitutes, the total temptation reduction is lower, and more temptation is needed to justify a given valuation. If x and y are complements, the total temptation reduction is higher, and less temptation is needed to justify a given valuation.

G.6 Intercept

Derivation of equation (33).

Re-arranging steady state consumption from equation (21) gives

$$
\begin{aligned}
(1-\alpha)\delta\rho(\phi-\xi)+\xi-(1-(1-\alpha)\delta\rho)\,p+(1-\alpha)\delta\rho\left[(\zeta-\eta)m_{ss}-\left(1+\tilde{\lambda}\right)\tilde{\gamma}\right]+\gamma= \\
x_{ss}\left[-\eta-(1-\alpha)\delta\rho(\zeta-\eta)-\zeta\frac{\rho-(1-\alpha)\delta\rho^2}{1-\rho}\right].
\end{aligned}
\tag{191}
$$

Solving for the intercept and substituting $x_{i1}=x_{ss}$ gives equation (33).

H Counterfactual Simulations Appendix

Table A14: **Effects of Temptation and Habit Formation on FITSBY Use**

FITSBY use (minutes/day)	(1) Restricted model ($\tau_2^B=0$, $\alpha=1$)	(2) Unrestricted model ($\alpha=\hat{\alpha}$)
Baseline	153	153
	[149, 157]	[149, 157]
No naivete	153	151
	[149, 157]	[140, 156]
No temptation	105	103
	[76.9, 120]	[67.0, 119]
No habit formation	78.1	73.3
	[50.2, 102]	[43.1, 99.4]
No temptation or habit formation	53.8	49.0
	[25.6, 77.9]	[17.2, 75.7]

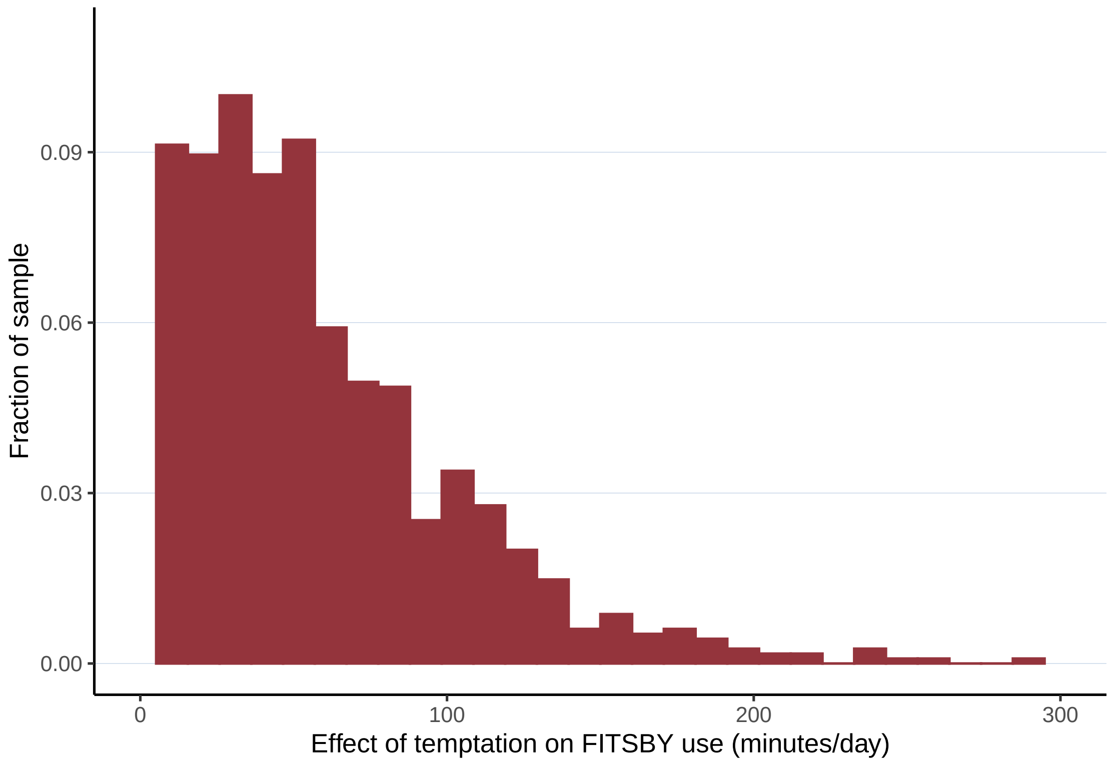
Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for predicted steady-state FITSBY use with different parameter assumptions, using equation (15). The numbers are as plotted in Figure 10.

Table A15: **Effects of Temptation on FITSBY Use Under Alternative Assumptions**

Effect of temptation on FITSBY use (minutes/day)	(1) Restricted model ($\tau_2^B = 0$, $\alpha = 1$)	(2) Unrestricted model ($\alpha = \hat{\alpha}$)
Limit effect	47.5	49.5
	$[34.3, 75.0]$	$[34.9, 86.4]$
Bonus valuation	70.5	71.2
	$[49.2, 116]$	$[49.5, 118]$
Limit valuation	61.5	62.3
	$[42.8, 103]$	$[43.3, 106]$
Limit effect, multiple-good model	57.4	
	$[40.0, 97.0]$	
Bonus valuation, multiple-good model	63.2	63.9
	$[44.5, 103]$	$[44.7, 107]$
Limit valuation, multiple-good model	91.3	91.6
	$[50.1, 155]$	$[51.0, 155]$
Limit effect, $\omega = \hat{\omega}$	123	127
	$[85.2, 155]$	$[87.3, 156]$
Limit valuation, $\omega = \hat{\omega}$	42.7	43.8
	$[29.7, 71.1]$	$[30.2, 76.6]$
Heterogeneous limit effect	47.1	48.6
	$[34.2, 71.9]$	$[34.6, 76.5]$
Limit effect, weighted sample	52.2	57.8
	$[32.3, 112]$	$[33.9, 144]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for the effects of temptation on average steady-state FITSBY use, using equation (15). The first nine estimates are for the nine temptation estimation strategies presented in Table A9. The tenth estimate is for the limit effect strategy after reweighting the sample to be more representative of U.S. adults. Appendix Tables A11–A13 present the demographics, moments, and parameter estimates in the weighted sample. Average baseline FITSBY use is 153 and 156 minutes per day for the unweighted and weighted samples, respectively. We do not have a limit effect estimate for the unrestricted multiple-good model.

Figure A35: **Distribution of Effects of Temptation on FITSBY Use**



Notes: Using the heterogeneous limit effect strategy, we estimate temptation $\hat{\gamma}_i$ for each Limit group participant, which we then insert into equation (15) to predict the individual-specific effect of temptation on steady-state FITSBY use. This figure presents the distribution of effects across participants, winsorized at 300 minutes per day.