Digital Addiction

Hunt Allcott, Matthew Gentzkow, and Lena Song*

January 29, 2022

Abstract

Many have argued that digital technologies such as smartphones and social media are addictive. We develop an economic model of digital addiction and estimate it using a randomized experiment. Temporary incentives to reduce social media use have persistent effects, suggesting social media are habit forming. Allowing people to set limits on their future screen time substantially reduces use, suggesting self-control problems. Additional evidence suggests people are inattentive to habit formation and partially unaware of self-control problems. Looking at these facts through the lens of our model suggests that self-control problems cause 31 percent of social media use.

JEL Codes: D12, D61, D90, D91, I31, L86, O33.

Keywords: Habit formation, projection bias, self-control, temptation, naivete, commitment devices, randomized experiments, social media.

*Allcott: Microsoft Research and NBER. hunt.allcott@microsoft.com. Gentzkow: Stanford University and NBER. gentzkow@stanford.edu. Song: New York University. lena.song@nyu.edu. We thank Dan Acland, Matthew Levy, Peter Maxted, Matthew Rabin, Dmitry Taubinsky, and seminar participants at the Behavioral Economics Annual Meeting, the Berkeley-Chicago Behavioral Economics Workshop, Bocconi, Boston University, Chicago Harris, Columbia Business School, Cornell, Di Tella University, the Federal Trade Commission Microeconomics Conference, Harvard, HBS, London Business School, London School of Economics, the Marketplace Innovation Workshop, Microsoft Research, MIT, the National Association for Business Economics Tech Economics Conference, the New York City Media Seminar, the New York Fed, NYU, Paris School of Economics, Princeton, Stanford Institute for Theoretical Economics, Trinity College Dublin, University of British Columbia, University College London, USC, Wharton, and Yale for helpful comments. We thank Michael Butler, Zong Huang, Zane Kashner, Uyseok Lee, Ana Carolina Paixao de Queiroz, Houda Nait El Barj, Bora Ozaltun, Ahmad Rahman, Andres Rodriguez, Eric Tang, and Sherry Yan for exceptional research assistance. We thank Chris Karr and Audacious Software for dedicated work on the Phone Dashboard app. We are grateful to the Sloan Foundation for generous support. Research was also supported by the Army Research Office under Grant Number W911NF-20-1-0252. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. The study was approved by Institutional Review Boards at Stanford (eProtocol #50759) and NYU (IRB-FY2020-3618). This experiment was registered in the American Economic Association Registry for randomized control trials; the pre-analysis plan is available from https://www.socialscienceregistry.org/trials/5796. Replication files and survey instruments are available from https://sites.google.com/site/allcott/research. Disclosures: Gentzkow does paid consulting work for Amazon, has done litigation consulting for clients including Facebook, and has been a member of the Toulouse Network for Information Technology, a research group funded by Microsoft. Both Allcott and Gentzkow are unpaid members of Facebook's 2020 Election Research Project.

1 Introduction

Digital technologies occupy a large and growing share of leisure time for people around the world. The average person with internet access spends 2.5 hours each day on social media, and there are now 3.8 billion social media users (Kemp 2020). In a 57-country survey, people now say they spend more time consuming online media than they do watching television (Zenith Media 2019). Americans check their smartphones 50 to 80 times each day (Deloitte 2018; Vox 2020; New York Post 2017).

A natural interpretation of these facts is that digital technologies provide tremendous consumer surplus. However, an increasingly popular alternative view is that habit formation and self-control problems—what we call "digital addiction"—play a substantial role. Many argue that smartphones, video games, and social media apps may be harmful and addictive in the same ways as cigarettes, drugs, or gambling (Alter 2018; Newport 2019; Eyal 2020). The World Health Organization (2018) has listed digital gaming disorder as an official medical condition. Recent experimental studies find that social media use can decrease subjective well-being (e.g. Mosquera et al. 2019; Allcott, Braghieri, Eichmeyer, and Gentzkow 2020). Figure 1 shows that social media and smartphone use are two of the top five activities that a sample of Americans think they do "too little" or "too much." Compared to the other three top activities ordered at left (exercise, retirement savings, and healthy eating), digital self-control problems have received much less attention from economists.[1]

The nature and magnitude of digital addiction matter for a number of important questions. Should people take steps to limit the amount of time they and their children spend on their smartphones and social media? What is the best way to design digital self-control tools? How can companies that make video games, social media, and smartphones best align their products with consumer welfare? Are proposed regulations such as the Social Media Addiction Reduction Technology (SMART) Act a good idea?[2]

In this paper, we formalize an economic model of digital addiction, use a randomized experiment to provide model-free evidence and estimate model parameters, and use the model to simulate the effects of habit formation and self-control problems on smartphone use. We focus on six apps that account for much of smartphone screen time and that participants report to be especially tempting: Facebook, Instagram, Twitter, Snapchat, web browsers, and YouTube. We refer to these apps as "FITSBY."

Our model follows Gruber and Köszegi (2001), Gul and Pesendorfer (2007), Bernheim and Rangel (2004), and others in defining addiction as the combination of two key forces: habit formation and self-control problems. As in Becker and Murphy (1988), habit formation means that today's consumption increases tomorrow's demand. As in Laibson (1997) and others, self-control problems mean that people consume more today than they would have chosen for themselves in advance. These two forces are central to classic addictive goods such as cigarettes, drugs, and alcohol.

[1] Among many important examples, see Charness and Gneezy (2009) and Carrera et al. (2021) on exercise, Madrian and Shea (2001) and Carroll et al. (2009) on retirement savings, and Sadoff, Samek, and Sprenger (2020) on healthy eating.

[2] This bill, introduced in 2019 by Republican Senator Josh Hawley, proposed to prohibit the use of design features such as infinite scroll and autoplay believed to make social media more addictive, and to require companies to default users into a limit of 30 minutes per day of social media use. See Hawley (2019).

Our model allows for projection bias (Loewenstein, O'Donoghue, and Rabin 2003), where people choose as if they are inattentive to habit formation, as well as naivete about self-control problems. As in Becker and Murphy (1988), people who perceive at least some habit formation would reduce consumption if they know the price will increase in the future, while projection bias would dampen that effect. As in many other models (see Ericson and Laibson 2019), people who are at least partially aware of self-control problems might want commitment devices to restrict future consumption, and people who are at least partially unaware will underestimate future consumption.

For our experiment, we used Facebook and Instagram ads to recruit about 2,000 American adults with Android smartphones and asked them to install Phone Dashboard, an app designed for our experiment that records smartphone screen time and allows participants to set screen time limits. Participants completed four surveys at three-week intervals—a baseline (survey 1) and three follow-ups (surveys 2, 3, and 4)—that included survey measures of smartphone addiction and subjective well-being as well as predictions of future FITSBY use. Participants answered three text message survey questions per week and kept Phone Dashboard installed for six weeks after survey 4.

We independently randomized two treatments. The *bonus treatment* was a temporary subsidy of $2.50 per hour for reducing FITSBY use during the three weeks between surveys 3 and 4. We informed people whether or not they were assigned to the bonus treatment in advance, on survey 2. The *limit treatment* made available screen time limit functionality in Phone Dashboard. Participants in this group could set personalized daily time limits for each app on their phone, with changes effective the next day. These limits forced participants to stop using the relevant app and in most cases could not be immediately overridden, unlike the flexible limits in existing tools such as Android's Digital Wellbeing and iOS's Screen Time. The surveys encouraged participants to set limits in line with their self-reported ideal screen time, but doing so was entirely optional. We used multiple price lists (MPLs) to elicit participants' valuations of the bonus treatment and the limit functionality.

The bonus treatment had persistent effects that are consistent with habit formation. The bonus reduced FITSBY use by 56 minutes per day during the three weeks when the incentives were in effect, a 39 percent reduction from the control group average. In the three weeks after the incentive had ended, the bonus treatment group still used 19 minutes less per day. In the three weeks after that, they used 12 minutes less per day.

Participants correctly predict habit formation: the effects of the bonus on predicted post-incentive FITSBY use line up closely with the effects on actual use. However, in the three weeks between when the bonus was announced and when it took effect, there was only a modest (and possibly zero) anticipatory response, which is only 12 percent of what our model would predict for forward-looking habit formation without projection bias. These results are consistent with a form of projection bias in which consumers are aware of habit formation while consuming as if they are inattentive to it.[3]

[3]This distinction between awareness and attention raises interesting questions about other evidence of projection bias. For example, Busse et al. (2015) find that people are more likely to buy a convertible on sunny days. On sunny days, do people have

We also find clear evidence that people have self-control problems and are at least partly aware of them. The limit treatment reduced FITSBY screen time by 22 minutes per day (16 percent) over 12 weeks. The effects decline slightly over the course of the experiment; this decline is consistent with some loss of motivation, but the fact that the decline is slight means that the effects are unlikely to be driven by confusion or temporary novelty. Although the experiment offered no incentive to set limits, 78 percent of participants set binding limits and continued using them through the final weeks of the experiment. This far exceeds takeup of almost all commitment devices studied in the literature reviewed by Schilbach (2019, Table 1). On average, participants were willing to give up $4.20 for three weeks of access to the limit functionality, and when trading off the bonus versus a fixed payment, 24 percent said they valued the bonus more highly because they wanted to give themselves an incentive to reduce consumption. These distinct measures of commitment demand are correlated with each other and with survey measures of addiction and desire to reduce screen time.

Notwithstanding their demand for commitment, participants seem to slightly underestimate their self-control problems. The control group modestly but repeatedly underestimated their future FITSBY use in all of our surveys, even though use is fairly steady over time and we reminded them of recent past use before asking them to predict. On average, the control group underestimated next-period FITSBY use by 6.1 minutes per day, or about 4 percent.

To further evaluate whether our interventions reduced addiction in a way that participants perceive to be beneficial, we examine effects on a variety of survey outcomes. On both the main surveys and text messages, the bonus and limit treatments significantly reduced an index of smartphone addiction adapted from the psychology literature. For example, both treatment groups reported being less likely to use their phone longer than intended, use their phone to distract from anxiety or fall asleep, have difficulty putting down their phone, lose sleep from phone use, procrastinate by using their phone, and use their phone mindlessly. Both treatment groups reported improved alignment between ideal and actual screen time. The bonus treatment group also scored higher on an index of subjective well-being, with statistically significant increases in components related to concentration and avoiding distraction and statistically insignificant changes in measures of happiness, life satisfaction, anxiety, and depression. Finally, both treatments are well-targeted in the sense that effects were more positive for people who report more interest in reducing their use and who score higher on our addiction measures at baseline.

In the final section of the paper, we look at these results through the lens of our structural model. The model allows us to translate our short-run experimental estimates into effects on long-run steady state behavior, to quantify the magnitude of the effects we observe in terms of economically meaningful parameters, and to decompose the role of different behavioral forces through counterfactuals. We first estimate the model parameters by matching key moments from the experiment. We model the limit treatment as eliminating share ω of self-control problems, and for our primary estimates we conservatively assume $\omega = 1$. The estimates reflect our experimental results: substantial habit formation and self-control problems, substantial

different beliefs about future weather or how much they would drive a convertible?

projection bias, and slight naivete about self-control problems. We then evaluate how steady-state consumption would change in counterfactuals where we eliminate self-control problems. Without habit formation, a conservative estimate of the effect of self-control problems is the effect of giving people screen time limit functionality: 22 minutes per day. But habit formation amplifies the effect of self-control problems, as the increase in current consumption also increases future marginal utility. In the presence of habit formation, our primary model prediction is that eliminating self-control problems would reduce FITSBY use by 48 minutes per day, or 31 percent of baseline use. Alternative assumptions mostly imply more self-control problems, more attention to habit formation, and larger effects on use.

Our results should be interpreted with caution for several reasons. First, our experiment took place during the beginning of the coronavirus pandemic. Our survey evidence suggests that this increased screen time but did not have clear effects on the magnitude of self-control problems. Furthermore, even as the pandemic evolved over the three-month experiment, average screen time and the treatment effects of the limit were fairly stable. Second, our estimates apply to the 2,000 people who selected into our experiment, and these people are not representative of U.S. adults. When we reweight our estimates to more closely approximate national average demographic characteristics, the modeled effect of self-control problems increases. Third, our model's predictions of FITSBY use without self-control problems depend on assumptions such as linear demand and geometric decay of habit stock. Fourth, our analysis is partial equilibrium in the sense that we do not model network effects and other externalities across users. If one person's social media use increases others' use, such positive network externalities would magnify the effects of self-control problems on population-wide social media use. Finally, our surveys walked participants through a process of setting optional screen time limits that implemented their self-reported ideal screen time, and we hypothesize that simply offering time limit functionality without walking through that process would have had smaller effects.[4]

Our work builds on several existing literatures. We extend a distinguished literature documenting present focus in diverse settings including exercise, healthy eating, consumption-savings decisions, and laboratory tasks (Ericson and Laibson 2019).[5] Ours is one of a small handful of papers that estimate the parameters of a present focus model with partial naivete using field (instead of laboratory) behavior.[6] The digital self-control

[4]While Carrera et al. (2021) show that takeup of commitment devices can be driven by experimenter demand effects or decision-making noise instead of perceived self-control problems, there are three reasons why their concerns are less likely to apply to our experiment. First, while Carrera et al. (2021) studied one-time takeup of an unfamiliar commitment contract, our participants repeatedly set and continually kept screen time limits over a 12-week period. Second, we estimate even larger perceived self-control problems using participants' valuations of the bonus treatment, which leverages an alternative methodology favored by Carrera et al. (2021) as well as Acland and Levy (2012), Augenblick and Rabin (2019), Chaloupka, Levy, and White (2019), Allcott, Kim, Taubinsky, and Zinman (2021), and Strack and Taubinsky (2021). Third, unlike Carrera et al. (2021), we find strong correlations between use of screen time limits and other measures of perceived self-control problems.

[5]This includes Read and Van Leeuwen (1998), Fang and Silverman (2004), Shapiro (2005), Shui and Ausubel (2005), Ashraf, Karlan, and Yin (2006), DellaVigna and Malmendier (2006), Paserman (2008), Gine, Karlan, and Zinman (2010), Duflo, Kremer, and Robinson (2011), Acland and Levy (2012), Andreoni and Sprenger (2012a; 2012b), Augenblick, Niederle, and Sprenger (2015), Beshears et al. (2015), Goda et al. (2015), Kaur, Kremer, and Mullainathan (2015), Laibson et al. (2015), Royer, Stehr, and Sydnor (2015), Exley and Naecker (2017), Augenblick (2018), Kuchler and Pagel (2018), Toussaert (2018), Augenblick and Rabin (2019), Casaburi and Macchiavello (2019), Schilbach (2019), John (2019), Toussaert (2018), and Sadoff, Samek, and Sprenger (2020).

[6]To our knowledge, these are Allcott, Kim, Taubinsky, and Zinman (2021), Bai et al. (2018), Carrera et al. (2021), Chaloupka,

problems we study are particularly interesting because this is one of the few domains where market forces have created commitment devices, such as blockers for smartphone apps, email, and websites (Laibson 2018). Our results suggest additional unmet demand for these commitment devices.

We also extend a distinguished literature on habit formation. One set of papers documents persistent impacts of temporary interventions in settings such as academic performance, energy use, exercise, hand washing, political protest, smoking, recycling, voting, water use, and weight loss.[7] We provide evidence in an important new domain. A second set of papers tests for forward-looking habit formation using belief elicitation or advance responses to future price changes, sometimes interpreting such forward-looking behavior as support for "rational" models of addiction.[8] We estimate anticipatory responses using an experimental approach that, like the one in Hussam et al. (2019), addresses many confounds that arise in observational data (Chaloupka and Warner 1999; Gruber and Köszegi 2001; Auld and Grootendorst 2004; Rees-Jones and Rozema 2020). Furthermore, we use our model to actually estimate the magnitude of projection bias, which is important because earlier studies that reject a null hypothesis of fully myopic habit formation could still be consistent with substantial projection bias.

Finally, we extend three literatures that speak directly to digital addiction. The first literature includes theoretical papers modeling temptation in digital networks (Makarov 2011; Liu, Sockin, and Xiong 2020). The second includes experimental papers studying the effects of social media use on outcomes such as subjective well-being and academic performance.[9] The third studies the effects of digital self-control tools.[10] Hoong (2021) is particularly related, and is an important antecedent to our study. In a smaller-scale experiment, she pioneers the use of encouragement to adopt self-control tools, compares predicted and ideal use to actual use, and shows results consistent with significant self-control problems. Our paper helps to unify the previous empirical literature with a formal model of digital addiction, relatively large sample, multiple treatment arms that convincingly identify habit formation and self-control problems using several different strategies, and robust measurement of screen time and survey outcomes.

Section 2 sets up the model. Sections 3–5 detail the experimental design, data, and model-free results. Section 6 presents the model estimation strategy and parameter estimates, and Section 7 presents the modeled effects of temptation on time use.

Levy, and White (2019), and Skiba and Tobacman (2018).

[7] This includes Gerber, Green, and Shachar (2003), Charness and Gneezy (2009), Gine, Karlan, and Zinman (2010), Ferraro, Miranda, and Price (2011), John et al. (2011), Allcott and Rogers (2014), Bernedo, Ferraro, and Price (2014), Acland and Levy (2015), Royer, Stehr, and Sydnor (2015), Fujiwara, Meng, and Vogl (2016), Levitt, List, and Sadoff (2016), Beshears and Milkman (2017), Brandon et al. (2017), Carrera et al. (2018), Allcott, Braghieri, Eichmeyer, and Gentzkow (2020), Bursztyn et al. (2020), Gosnell, List, and Metcalfe (2020), and Van Soest and Vollaard (2019).

[8] This includes Chaloupka (1991), Becker, Grossman, and Murphy (1994), Gruber and Köszegi (2001), Acland and Levy (2015), Hussam et al. (2019), and Do and Jacoby (2020).

[9] This includes Sagioglu and Greitemeyer (2014), Tromholt (2016), Hunt et al. (2018), Vanman, Baker, and Tobin (2018), Mosquera et al. (2019), Allcott, Braghieri, Eichmeyer, and Gentzkow (2020), and Collis and Eggers (2019).

[10] This includes Marotta and Acquisti (2017) and Acland and Chow (2018).

2 Model

The goal of the model is to formalize the meaning of "digital addiction" and foreshadow how we identify the model parameters using our experiment.

In each period $t \leq T$, consumers choose consumption of a good x_t sold at price p_t that delivers flow utility $u_t(x_t; s_t, p_t)$. To model habit formation, utility depends on a stock s_t of past consumption that evolves according to

$$s_{t+1} = \rho(s_t + x_t), \tag{1}$$

where $\rho \in [0,1)$ captures the strength of habit formation. Habit formation captures why temporary price changes generate persistent effects in our experiment.

To model self-control problems, we follow Banerjee and Mullainathan (2010) in modeling x as a temptation good. Before period t, consumers consider period t flow utility to be $u_t(x_t; s_t, p_t)$. In period t, however, consumers choose as if period t flow utility is $u_t(x_t; s_t, p_t) + \gamma x_t$, where $\gamma \geq 0$ reflects the amount of temptation. If $\gamma > 0$, consumers choose more x_t in period t than they would choose in advance. This temptation good framework generates similar predictions to the quasi-hyperbolic model from Laibson (1997) and Gruber and Köszegi (2001), but it naturally matches our application to a single addictive good and yields simpler estimating equations where temptation is additively separable.

Consumers may misperceive temptation: before period t, consumers predict that in period t, they will consider flow utility to be $u_t(x_t; s_t, p_t) + \tilde{\gamma} x_t$. We say that consumers are fully naive if $\tilde{\gamma} = 0$, and fully sophisticated if $\tilde{\gamma} = \gamma$. Partial naivete captures why our experiment participants underestimate x_t when asked to predict in advance. Partial sophistication captures why our participants want commitment devices to change their future behavior.

Following Loewenstein, O'Donoghue, and Rabin (2003), we allow the possibility of projection bias, in which consumers choose as if to maximize a weighted average of utility given the current habit stock s_t and utility given the predicted habit stock $\tilde{s}_r$ in future period $r > t$. We let α denote the weight on the current habit stock, and thus the magnitude of projection bias. Projection bias captures why consumers in our experiment might not reduce consumption in anticipation of a known future price change. We assume that consumers are fully naive about projection bias; sophistication would introduce strategic incentives to adjust current consumption to offset future bias.[11]

Following O'Donoghue and Rabin (1999) and others, we solve for perception-perfect equilibrium strategies, where consumers maximize current utility given predictions of future behavior. Let $x_t(s_t, \gamma, \boldsymbol{p}_t)$ denote a strategy of the period-t self, which depends on habit stock, temptation, and the vector of future prices $\boldsymbol{p}_t = \{p_t, p_{t+1}, ..., p_T\}$. Let $\tilde{x}_r(s_r, \tilde{\gamma}, \boldsymbol{p}_r)$ be a consumer's *prediction,* as of period $t < r$, of her period-r

[11] Loewenstein, O'Donoghue, and Rabin (2003, page 1219) also assume naivete about projection bias, writing that "because this time inconsistency derives solely from misprediction of future utilities, it would make little sense to assume that the person is fully aware of it." We note that our formulation of projection bias is slightly different than in Loewenstein, O'Donoghue, and Rabin (2003): while their consumers' predictions of future consumption are biased due to projection bias, our consumers predict consumption accounting for habit formation, but choose as if they are inattentive to it. This matches our empirical results.

strategy. A strategy profile $(x_0^*, ..., x_T^*)$ is perception perfect if in each period t

$$x_t^*(s_t, \gamma, \boldsymbol{p}_t) = \arg\max_{x_t} u_t(x_t; s_t, p_t) + \gamma x_t + \left[\begin{array}{c} \alpha \sum_{r=t+1}^{T} \delta^{r-t} u_r(\tilde{x}_r^*(s_t, \tilde{\gamma}, \boldsymbol{p}_r); s_t, p_r) \\ +(1-\alpha) \sum_{r=t+1}^{T} \delta^{r-t} u_r(\tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, \boldsymbol{p}_r); \tilde{s}_r, p_r) \end{array} \right], \quad (2)$$

where $\delta \leq 1$ is the discount factor.

Predicted and actual consumption differ due to naivete about temptation and projection bias and the resulting misprediction of habit stock. We assume that the equilibrium prediction $\tilde{x}_r^*(s_r, \tilde{\gamma}, \boldsymbol{p}_r)$ is the solution to equation (2) with $\alpha = 0$ and $\gamma = \tilde{\gamma}$. Predicted habit stock $\tilde{s}_r$ evolves according to $\tilde{s}_{r+1} = \rho(\tilde{s}_r + \tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, \boldsymbol{p}_r))$.[12] The "rational" habit formation model of Becker and Murphy (1988) is the special case with $\alpha = 0$ and $\tilde{\gamma} = \gamma = 0$.

To estimate the model, we follow Becker and Murphy (1988) and Gruber and Köszegi (2001) in specializing to the case of quadratic flow utility:

$$u_t(x_t; s_t, p_t) = \frac{\eta}{2} x_t^2 + \zeta x_t s_t + \phi s_t + (\xi_t - p_t) x_t \quad (3)$$

where $\eta < 0$ measures the demand slope, ζ regulates the extent of habit formation, ϕ is the direct effect of habit stock on utility (which could be positive or negative), and ξ_t is a deterministic period-specific demand shifter. This can be microfounded by assuming that consumers have income w that they must spend in each period, and income not spent on x_t is spent on a numeraire $c_t = w - p_t x_t$ that is additively separable in u_t. In this specification, u_t is in units of dollars per period.

3 Experimental Design

3.1 Overview

Our experiment is designed to provide direct evidence on the magnitude of habit formation, perceived habit formation, temptation, and perceived temptation, as well as to identify the remaining key parameters of the quadratic model. The experiment ran from March 22 to July 26, 2020, with participants completing an intake questionnaire and four surveys. Figure 2 summarizes the experimental design, and Table 1 presents sample sizes at each step.

Between March 22 and April 8, we recruited participants using Facebook and Instagram ads. Appendix Figure A1 presents the ads. To minimize sample selection bias, the ads did not hint at our research questions or suggest that the study was related to smartphone use or social media. 3,271,165 unique users were shown one of the ads, of whom 26,101 clicked on it. This 0.8 percent click-through rate is close to the average click-through rate on Facebook ads (Irvine 2018).

[12] Since the predicted equilibrium strategy $\tilde{x}_r^*(s_r, \tilde{\gamma}, \boldsymbol{p}_r)$ conditions on the state s_r inherited at time r, it will be the same when evaluated in all periods $t < r$. However, the predicted *action* in period r is not generally the same when evaluated in all periods $t < r$, as the predicted $\tilde{s}_r$ will differ by t.

Clicking on the ad took the participant to a brief screening survey, which included several background questions, the consent form, and instructions on how to install Phone Dashboard. To be eligible, participants had to be a U.S. resident between 18 and 64 years old, use an Android as their primary phone, and use only one smartphone regularly. 18,589 people satisfied these criteria, of whom 8,514 consented to participate in the study. Of these, 5,320 successfully installed Phone Dashboard and finished the intake survey.

Surveys 1–4 were administered on Sundays at three week intervals between April 12th and June 14th. We define $t = 1, 2, 3, ...$ to be the three-week periods beginning Monday April 13th, so period t is the three weeks immediately after survey t. For our data analysis and interventions, we want to exclude survey days, so all periods are 20 days long, from a Monday to a Saturday. Survey 1 recorded participant demographics. We describe the other survey content below.

As illustrated in Figure 2, we randomized participants into bonus and limit treatment conditions (detailed below) using a factorial design. We randomized participants to the Bonus, Bonus Control, or the Multiple Price List (MPL) group with 25, 75, and 0.2 percent probability, respectively. We independently randomized participants to the Limit or Limit Control groups with 60 and 40 percent probability, respectively. We refer to the intersection of the Bonus Control and Limit Control groups as the Control group. We balanced the randomization within eight strata defined by above- versus below-median baseline FITSBY use, *restriction index*, and *addiction index* (described below). The treatments began on survey 2.

All participants received $5 for completing the baseline survey and $25 if they completed the remaining surveys and kept Phone Dashboard installed through July 26th. Participants were also entered in a drawing for a $500 gift card, in which two winners were drawn.

As shown in Table 1, 4,038 participants completed survey 1. We dropped 1,912 of these participants from the experiment after survey 1 because they reported that they already used another app to limit their phone use (5 percent of the sample) or failed data quality checks.[13] The remaining 2,126 participants were invited to take survey 2, of whom 2,053 opened the survey and reached the point where the treatments began. Of those, 1,938 completed the study—remarkably low attrition for a 12-week study with multiple surveys.

In addition to back-loading the survey payments, several other factors contributed to our limited attrition. There were two surveys (the intake and survey 1) before the treatments began, inducing likely attriters to attrit beforehand. At the beginning of survey 2, just before the treatments began, we informed people that "anyone who drops out after this page can really damage the entire study," and offered them a choice to drop out at that moment or commit to finishing the whole study. For participants who had not yet completed each of surveys 2–4, we sent daily reminders for six days after the survey had been fielded, and after four days we began offering an additional payment for completing all remaining surveys. We also sent reminder emails to people who had failed to respond to two consecutive text messages.

[13] Participants failed data quality checks if they (i) did not to promise to "provide my best answers" on our surveys; (ii) reported having idiosyncratic bugs with Phone Dashboard; (iii) failed to answer more than two of our text message questions between survey 1 and survey 2; (iv) had a device that was incompatible with Phone Dashboard; or (v) were missing screen time data during the baseline period.

3.2 Phone Dashboard

Phone Dashboard is an Android app that was developed by a company called Audacious Software for our experiment. Appendix Figure A2 presents screenshots. Our experiment includes only Android users because a similar functionality cannot be implemented by third-party apps on iOS.

Phone Dashboard records the app that is in the foreground of a smartphone every five seconds when the screen is on; we use these data to construct our measure of consumption. It does not record the content that the user is viewing within the app. Users can see their cumulative screen time by day and by week on the Phone Dashboard home screen. This usage information was designed to be particularly useful for participants in the Bonus and Limit groups who might want to track their usage, but the Control group also used the app: the Bonus, Limit, and Control groups used Phone Dashboard for an average of 1.4, 1.5 and 1.0 minutes per day during periods 2–5.

3.3 Bonus Treatment

The bonus treatment was designed to identify projection bias (the parameter α), actual habit formation (ρ and ζ), and the curvature of utility (η). To facilitate the multiple price list (MPL) described below, survey 2 explained the bonus to all participants before telling them whether they were selected to receive it and when it would be in force. Participants were told,

If you're selected for the Screen Time Bonus, you would receive $50 for every hour you reduce your average daily FITSBY screen time below a Bonus Benchmark of [X] hours per day over the 3-week period, up to $150.

The survey then gave several examples, including:

- *If you reduce your FITSBY screen time to $[X-1] hours and 30 minutes per day over the next 3 weeks, you would receive $25.*
- *If your FITSBY screen time is above $X hours per day, you would receive $0.*

We set the Bonus Benchmark [X] as the participant's average FITSBY hours per day from period 1, rounded up to the nearest integer.

After the MPL described below, the Bonus group was informed that they had been randomly selected to receive the bonus for screen time reductions during period 3—i.e., starting in three weeks. The Bonus Control group was informed that they would not receive the bonus. To ensure that participants understood, each participant had to answer a question by correctly indicating their bonus treatment condition before advancing. We also sent three text messages reminders to the Bonus group during period 2, which read "Don't forget, we'll pay you $50 for every hour you reduce your average daily screen time between May 24 and June 14. There is no bonus for changing your screen time before then." People were asked to respond to the text message to confirm that they had read it. Survey 3 included an additional reminder for the Bonus treatment group. While we received substantial feedback on the surveys and many emails from our 2,000

participants during the study and our earlier pilots, none of these interactions suggested confusion about the timing of the bonus.

The Bonus group's anticipatory response to the bonus in period 2 (before the incentive was in effect) provides information about the magnitude of projection bias α. The contemporaneous response in period 3 (while the incentive was in effect) provides information about the price response parameter η. The long-term effects in periods 4 and 5 (after the incentive had ended) provides information about the magnitude and decay of habit (ζ and ρ).

3.4 Limit Treatment

The limit treatment was designed to understand self-control problems and help identify the temptation parameter γ. The Limit treatment group was given access to functionality in Phone Dashboard that allows users to set daily time limits for each app on their phone; see Appendix Figure A2 for screenshots. Any changes to the limits take effect the next day. Phone Dashboard serves five-minute and one-minute push notifications as an app's daily time limit approaches. When the limit arrives, users can "snooze" their limit and get an additional amount of time that they specify—but starting only after a delay. Within the Limit group, we randomly assigned participants with equal probability to delays of 0, 2, 5, or 20 minutes or a condition where the ability to snooze was disabled. To keep the scope of this paper manageable, we focus only on the comparison between the Limit and Limit Control groups; we plan to study the variation in snooze delays in a separate paper. To reduce attrition and uninstallation, Phone Dashboard also allows people to permanently opt out of the limits; about 4 percent of the Limit group did so.

The Limit group was first given access to the Phone Dashboard limit functionality on survey 2, after the Screen Time Bonus multiple price list described below, and they retained access to the feature for the duration of the experiment. To introduce the limits, we first gave participants instructions on how to set daily app usage limits for themselves. The survey then asked participants what time limits they would like to set for themselves on each FITSBY app over the next three weeks. We then asked participants to update their Phone Dashboard app, which activated the limit functionality, and encouraged them to set the limits they had reported a moment earlier. The Limit Control group was never told about limits and continued to have a version of Phone Dashboard that did not have the limit functionality.

In the analysis below, we interpret use of the limits as evidence of perceived self-control problems ($\tilde{\gamma} > 0$).

3.5 Bonus and Limit Valuations

We used incentive-compatible multiple price list mechanisms to elicit valuations of the Screen Time Bonus and the limit functionality. Because both the bonus and the limit functionality reduce future social media use, these valuations help identify perceived temptation $\tilde{\gamma}$.

All multiple price lists included a table with a series of choices between "Option A" and "Option B"

in separate rows. Option B was the same in each row, while Option A included an amount of money that decreased monotonically from top to bottom. Participants would typically choose Option A at the top and Option B at the bottom, and we infer their valuation of Option B from the row where they switch. To encourage valid answers, participants who did not switch between Option A and Option B exactly once were alerted to this fact and given a chance to change their answers. All MPLs were incentivized, as described below. To help participants become familiar with MPLs, survey 1 included an incentivized practice MPL that asked participants to choose between receiving different survey completion payments at different times.

Our approach to valuing the Screen Time Bonus builds on Allcott, Kim, Taubinsky, and Zinman (2021) and Carrera et al. (2021). Survey 2 informed participants of their average daily FITSBY screen time over the past three weeks and asked them to predict their screen time over the next three weeks. The survey then introduced the Screen Time Bonus and asked participants to predict how much they would reduce their FITSBY screen time relative to their original prediction if they were selected for the bonus.

After these two predictions, we asked participants to make a hypothetical choice between the Screen Time Bonus and a payment equal to their expected earnings from the bonus. The survey described potential considerations as follows:

- *You might prefer $[expected earnings] instead of the Screen Time Bonus if you don't want any pressure to reduce your screen time.*
- *You might prefer the Screen Time Bonus instead of $[expected earnings] if you want to give yourself extra incentive to use your phone less.*

Participants then completed an MPL where Option B was receiving the Screen Time Bonus, and Option A was receiving a payment ranging from $150 to $0.

To make the MPL incentive compatible, participants were told, "Last week, the computer randomly selected some participants to receive what they choose on the multiple price list below, and also randomly selected one of the rows to be 'the question that counts.' If you were randomly selected to participate, you will be paid based on what you choose in that row." 0.2 percent of participants were randomly assigned to the MPL group that received what they chose on a randomly selected row.

On survey 3, the Limit group completed an MPL that elicited valuations of the Phone Dashboard limit functions. Option B was retaining access to the Phone Dashboard limit functions, and Option A was having those functions disabled for the following three weeks in exchange for a dollar payment that ranged from $20 to -$1. The MPL group received what they chose on a randomly selected row.

3.6 Predicted Use

At the end of surveys 2, 3 and 4, we elicited predictions of future FITSBY use. These predictions help identify the degree of naivete or sophistication about temptation—the difference between γ and $\tilde{\gamma}$.

Before each elicitation, we told each participant their average FITSBY screen time over the previous three weeks. Surveys 2 and 3 also reminded the Bonus and Limit groups about the bonus and limits. Survey

2 then elicited predictions of FITSBY screen time for the next three weeks (period 2), the three weeks after that (period 3), and the three weeks after that (period 4). Survey 3 elicited separate predictions for periods 3, 4, and 5. Survey 4 elicited separate predictions for periods 4 and 5.

Predictions were incentivized. Survey 2 told participants, "Answer carefully, because you might earn a Prediction Reward. After the study ends, we will pick a prediction question at random and check how close your prediction is. If your predicted daily screen time is within 15 minutes of your actual screen time, we will pay you an additional $X." We randomized the prediction reward X to be $1 or $5, each with 50 percent probability.

3.7 Survey Outcome Variables

Surveys 1, 3, and 4 asked questions designed to measure participants' perceptions of their addiction and subjective well-being (SWB). For the nine weeks between survey 1 and survey 4, we also sent three text messages per week with a subset of questions that we thought were important to ask in real time instead of retrospectively. Using these questions, we construct five pre-specified outcome variables. Appendix A.1 presents details on the survey questions.

Ideal use change. The survey said,

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 smartphone use over the past 3 weeks?

- *I use my smartphone too much.*
- *I use my smartphone the right amount.*
- *I use my smartphone too little.*

For people who said they used their smartphone "too much" or "too little," we then asked, *Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your smartphone use?* The *ideal use change* variable is the answer to this question, in percent.

Addiction scale. Our addiction scale is a battery of 16 questions modified from two well-established survey scales, the Mobile Phone Problem Use Scale (Bianchi and Phillips 2005) and the Bergen Facebook Addiction Scale (Andreassen et al. 2012). The questions attempt to measure the six core components of addition identified in the addiction literature: salience, tolerance, mood modification, relapse, withdrawal, and conflict (Griffiths 2005).

The survey asked, *In the past three weeks, how often have you ...*, with a matrix of 16 questions, such as

- *used your phone longer than intended?*
- *felt anxious when you don't have your phone?*

- *lost sleep due to using your phone late at night?*

Possible answers were Never, Rarely, Sometimes, Often, and Always, which we coded as 0, 0.25, 0.5, 0.75, and 1, respectively. *Addiction scale* is the sum of these numerical scores for the 16 questions.

SMS addiction scale. The SMS addiction scale includes shortened versions of nine questions from the addiction scale. Examples include:

- *In the past day, did you feel like you had an easy time controlling your screen time?*
- *In the past day, did you use your phone mindlessly?*
- *When you woke up today, did you immediately check social media, text messages, or email?*

People were instructed to text back their answers on a scale from 1 (not at all) to 10 (definitely). *SMS addiction scale* is the sum of these scores for the nine questions.

Phone makes life better. The survey asked, *To what extent do you think your smartphone use makes your life better or worse?* Responses were on a scale from -5 ("Makes my life worse") through 0 ("Neutral") to +5 ("Makes my life better").

Subjective well-being. We use standard measures from the subjective well-being literature, mostly following the measures from our own earlier work (Allcott, Braghieri, Eichmeyer, and Gentzkow 2020). The survey asked,

Please tell us the extent to which you agree or disagree with each of the following statements. ***Over the last three weeks,*** with a matrix of seven questions:

- … *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*

Possible answers were on a seven-point scale from "strongly disagree" through "neutral" to "strongly agree," which were coded as -1, -2/3, -1/3, 0, 1/3, 2/3, and 1, respectively. The variable *subjective well-being* is the sum of these numerical scores for the seven questions, after reversing *anxious*, *depressed*, and *easily distracted* so that more positive reflects better subjective well-being.

Indices. We define the *survey index* to be the sum of the five survey outcome variables described above, weighted by the baseline inverse covariance matrix as described by Anderson (2008). When presenting results and constructing this index, we orient the variables so that more positive values imply normatively better outcomes. Thus, we multiply *addiction scale* and *SMS addiction scale* by (-1).

We define the *restriction index* to be the sum of *interest in limits* (with the four categorical answers coded as 0, 1, 2, and 3) and *ideal use change*, after normalizing each into standard deviation units. We define the *addiction index* to be the sum of *addiction scale* and *phone makes life better* after normalizing each into standard deviation units. We use these two indices for stratified randomization and as moderators when testing for heterogeneous treatment effects.

3.8 Pre-Analysis Plan

We submitted our pre-analysis plan (PAP) on May 4th, the day that post-treatment data collection began. The PAP specified (i) the equation for treatment effect estimation (equation 4 below); (ii) the construction of the survey outcome variables and indices described in Section 3.7, the *limit tightness* variable, and the winsorization of predicted FITSBY use; and (iii) the analysis of heterogeneous treatment effects by splitting the sample on above- versus below-median values of six moderators: education, age, gender, baseline FITSBY use, *restriction index*, and *addiction index*. The PAP also included shells of Tables 1, 2, and A1–A3, as well as Figures 1–6, A1–A4, A8, and A28–A34.

We deviate from the PAP in five ways. First, the bottom left panel of Figure 3 includes results from each addiction scale question, whereas the PAP figure shell presented the sum across all questions. Second, we clarify that our analysis sample includes only the balanced panel of people who completed the study. Results are essentially identical if we use an unbalanced panel that includes data from attriters before they attritted, but the balanced panel is helpful in ensuring that our habit formation results are not spuriously driven by attrition. Third, three figures from the PAP are not included here, as we plan to study them in a separate paper. Fourth, Figure 6 includes predicted FITSBY use from all surveys before period t, whereas the PAP figure shell presented predictions from only the survey immediately before period t. Fifth, we use equation (4) for subgroup analysis, whereas the PAP specified that we would use an instrumental variables regression. We present the pre-specified instrumental variables estimates in Appendix D.4. The results are similar, and we decided that equation (4) was simpler.

4 Data

The analysis sample for all results reported below is the balanced panel of $1,933$ participants who were assigned to either Bonus or Bonus Control (not the MPL group), completed all four surveys, and kept Phone Dashboard installed until the end of the study on July 26. This group's attrition rate after being informed of treatment was $(1-1,933/2,048)\times 100\% \approx 5.6$ percent. Attrition rates and observable characteristics are balanced across the bonus and limit treatment conditions; see Appendix Tables A1 and A2.

Table 2 quantifies the representativeness of our analysis sample on observables, by comparing their demographics to the U.S. adult population. Our sample is more educated, more heavily female, younger, and slightly lower-income than the U.S. population. We estimate an alternative specification of our structural model with sample weights to adjust for these observable differences.

Table 2 also shows that the average participant had 333 minutes per day of screen time during the baseline period, of which 153 minutes (46 percent) was on FITSBY apps. Different sources report very different estimates of average social media use and smartphone screen time for U.S. adults, so we do not report nationwide averages in the table. Kemp (2020) reports that internet users in the U.S. and worldwide, respectively, spend an average of 123 and 144 minutes per day on social media, mostly on mobile devices. Wurmser (2020) and Brown (2019) report national averages of 186 and 324 minutes of total smartphone screen time per day, respectively. The comparisons suggest that the heavy use in our sample may not be far from the national average.

During the baseline period, the average participant used Facebook, browsers, YouTube, Instagram, Snapchat, and Twitter for 69, 44, 23, 24, 15, and 15 minutes per day, respectively; see Appendix Figure A3. Appendix Figure A4 presents the distribution of baseline FITSBY use. Appendix Table A3 presents descriptive statistics for the survey outcome variables.

5 Model-Free Results

5.1 Treatment Effect Estimating Equation

To estimate treatment effects, define Y_{it} as an outcome for participant i for period t. Y_{it} could represent a survey outcome variable measured on survey $t \in \{3,4\}$, or period t FITSBY use. Define L_i and B_i as Limit and Bonus group indicators. Define $\boldsymbol{X}_{i1}$ as a vector of baseline covariates: baseline FITSBY use and, if and only if Y is a survey outcome variable, the baseline value Y_{i1} and the baseline value of *survey index*. Define ν_{it} as a vector of the eight randomization stratum indicators, allowing separate coefficients for each period t. We estimate the effects of the limit and bonus treatments using the following regression:

$$Y_{it} = \tau_t^B B_i + \tau_t^L L_i + \beta_t \boldsymbol{X}_{i1} + \nu_{it} + \varepsilon_{it}. \tag{4}$$

When combining data across multiple periods, we cluster standard errors by participant.

5.2 Baseline Qualitative Evidence

Figure 3 presents qualitative evidence on digital addiction from the baseline survey. The top two panels present the variables in the *restriction index*. The top left panel shows that 23 percent of people reported being "moderately" or "very" interested in setting time use limits on their smartphone apps, while 34 percent reported being "not at all" interested. The top right panel presents the distribution of responses to the *ideal use change* question. 42 percent of people said that they used their smartphone the right amount over the

past three weeks, and only 0.5 percent said that they used it too little. Among people who said they used their smartphone too much, the average ideal reduction was 34 percent.

Survey 1 also asked people to report their ideal use change for specific apps or categories. FITSBY, games, video streaming, and messaging are the nine apps on which people want to reduce screen time the most; see Appendix Figure A8. Facebook is by far the most tempting app: the average participant would ideally reduce Facebook use by 22 percent. The average participant did not want to change their use of email, news, and maps and wanted to slightly increase use of phone, music, and podcast apps.

The bottom two panels present the variables in the *addiction index*. The bottom left panel presents the share of participants who responded "often" or "always" on each question in the *addiction scale*. The top seven questions capture three components of moderate addictions (salience, tolerance, and mood modification); 33 percent of participants often or always experience each of these, and 84 percent often or always experience at least one. The bottom nine questions capture three components of more severe addictions (relapse, withdrawal, or conflict); 11 percent of participants often or always experience each of these, and 41 percent often or always experience at least one. The bottom right panel shows that while most people think that their smartphone use makes their life better, 19 percent think that it makes their life worse. Taken together, these results suggest substantial heterogeneity: many people report experiences consistent with addiction, while many others do not.

Our experiment took place during the coronavirus pandemic, which significantly disrupted people's daily routines. To understand how this might affect our results, we included several baseline survey questions, which we report in Appendix C. 78 percent of people reported having more free time as a result of the pandemic, and 88 percent of people reported that the pandemic had increased their phone use. However, it is not clear that the pandemic affected the extent of self-control problems: the means and distributions of key qualitative measures of addiction that we also asked for 2019, *ideal use change* and *phone makes life better*, were statistically different but economically similar. *Ideal use change* is closer to zero in 2020 compared to in 2019, suggesting less perceived self-control problems, but *phone makes life better* is also less positive, suggesting more perceived self-control problems.

5.3 Bonus Treatment and Habit Formation

The darker coefficients in Figure 4 present the effect of the bonus on FITSBY use, estimated using equation (4). Recall that the bonus provides an incentive to reduce FITSBY use in period 3, but we informed participants about whether or not they were offered the bonus at the beginning of period 2. The incentive is \$50 per *average* hour measured over the 20-day period, or \$2.50 per hour of consumption.

In period 3 (while the incentive was in effect), the Bonus group reduced FITSBY use by 56 minutes per day, or 39 percent relative to the Control group. This is a striking price response: it implies that participants value a substantial share of smartphone FITSBY use at less than \$2.50 per hour.

In periods 4 and 5 (after the incentive had ended), the Bonus group still reduced FITSBY use by 19 and 12 minutes per day, respectively. This persistent effect suggests substantial habit formation, implying $\zeta > 0$

in our model. The decay of the effect in period 5 relative to period 4 provides information about the habit stock decay parameter ρ in our model.

In period 2 (before the incentive was in effect), the Bonus group reduced FITSBY use by 5.1 minutes per day, which is marginally statistically significant. This is consistent with the model's prediction that a consumer who perceives habit formation should reduce period 2 consumption in order to reduce period 3 marginal utility, which makes it easier to reduce period 3 consumption in response to the financial incentive. However, additional evidence suggests some caution about interpreting the period 2 effect as forward-looking habit formation. Appendix Figures A9 and A10 break out the period 2 effect separately by day and week, showing that it loads mostly on the first half of the period. If anything, forward-looking habit formation would predict the opposite pattern, with larger anticipatory effects closer to the beginning of the incentive period. Possible explanations include intertemporal substitution, a temporary idiosyncratic effect, and the salience of the bonus after its introduction on survey 2.[14]

5.4 Limit Treatment and Temptation

The Limit group made extensive use of the limit functionality. To summarize the stringency of time limits, we define the variable *limit tightness* to be the amount by which a user's limits would have hypothetically reduced screen time if applied to their baseline use.[15] *Limit tightness* equals zero (instead of missing) for an app if the participant doesn't have the app or doesn't set a limit, so this variable speaks to what apps contribute the most to aggregate temptation. About 89 percent of the Limit group had positive *limit tightness* at some point during the experiment, suggesting that they set binding screen time limits, and 78 had positive *limit tightness* in period 5, meaning that they kept those limits for more than three weeks after the final survey. Participants most wanted to restrict Facebook, web browsers, YouTube, and Instagram: *limit tightness* averaged 20, 10, 8, and 6 minutes per day on those apps, respectively, across periods 2–5. Across all apps, the Limit group's average *limit tightness* was 53 minutes per day. See Appendix Figures A11 and A12 for details.

The lighter coefficients on Figure 4 present the effect of the limit on FITSBY use. These actual effects are smaller than the *limit tightness* values in the previous paragraph primarily because users snooze the

[14] Although we stratified randomization on period 1 FITSBY use and also control for period 1 use when estimating equation (4), some idiosyncratic factor could temporarily affect consumption in Bonus versus Bonus Control at the beginning of period 2. Some evidence supports this possibility: Appendix Figure A9 shows that consumption is slightly lower in the Bonus group compared to Bonus Control in the final 11 days of period 1. Salience could also play a role, although as described in Section 3.3, we took many steps to eliminate confusion about the timing of the bonus incentive period, and participants likely would have emailed our team if they were confused.

[15] Specifically, define x_{iadt} as the screen time of person i on app a on day d in period t. Define h_{iat} as the average screen time limit in place in period t, and define $N_{d\in t=1}$ as the number of days in the baseline period. *Limit tightness* is

$$H_{iat} = \frac{1}{N_{d\in t=1}} \sum_{d\in t=1} \max\{0, x_{iad1} - h_{iat}\}. \tag{5}$$

If the daily limit h_{iat} would not have been binding in baseline day d, the max function returns 0. If h_{iat} would have been binding in day d, then the max function returns the excess screen time on that day. We aggregate over apps to construct user-level limit tightness $H_{it} = \sum_a H_{iat}$.

limits. Access to the limit functionality reduced use in periods 2–5 by an average of 22 minutes per day, or 16 percent relative to the Control group. The effects attenuate only slightly as the experiment continues, and the effect is still 19 minutes per day in the last week of period 5. This is notable because while surveys 2 and 3 walked people through a limit-setting process and survey 4 included an optional review of the limits, the end of period 5 is nine weeks after survey 3 and six weeks after survey 4. These continuing effects suggest that while motivation might decrease over time, use of the limits is not primarily driven by confusion or temporary novelty. Furthermore, Appendix D.1 shows that *limit tightness* is correlated in expected ways with bonus and limit valuations and with survey measures of addiction and desire to reduce screen time. This evidence consistently points toward perceived self-control problems, implying $\tilde{\gamma} > 0$ in our model.

When we add the interaction between Bonus and Limit group indicators to equation (4), the main effects are similar and the interaction terms are not statistically significant; see Appendix Figure A13.

5.5 Substitution

Figure 5 presents usage effects of the bonus (in period 3 only) and the limit (across periods 2–5) separately by app. Among the FITSBY apps, Facebook sees the largest reductions, followed by web browsers, YouTube, Instagram, Twitter, and Snapchat. The effects on other apps (the right-most coefficients) provide evidence on the extent to which participants substituted FITSBY time to alternative apps. The bonus has no statistically detectable effect on use of other apps in period 3, and the confidence intervals rule out any substantial substitution relative to the 56 minutes per day reduction in FITSBY use. The limit induces substitution of 12 minutes per day, so that roughly half of the FITSBY screen time that the limit eliminates moves to other apps where people had been less likely to set limits.

One important limitation is that we cannot directly monitor FITSBY use on devices other than the participant's smartphone. We screened out potential participants who reported using more than one smartphone regularly, but our remaining participants may still have used desktops, tablets, or other devices. To provide some evidence on this substitution, survey 4 asked participants to estimate their FITSBY use on other devices in period 3 compared to the three weeks before they joined the study. The results, shown in Appendix Figure A14, imply that the limit increased FITSBY use on other devices by a marginally significant 4.2 minutes per day. The bonus *reduced* the amount of time they spent on FITSBY on other devices by 8.1 minutes per day, suggesting that time on other devices was a mild complement in this case.

The differences in substitution induced by the bonus versus limit are notable. In a simple model where other apps and devices are either complements or substitutes for smartphone FITSBY use, the substitution effects described above might have the same sign for both the bonus and limit and might be in proportion to their direct effects on smartphone FITSBY use. In contrast, a much smaller share of the effect on FITSBY use is substituted to other smartphone apps for the bonus compared to the limit, and the self-reported effects on FITSBY use on other devices have opposite signs for the bonus versus the limit. This is an interesting result to understand in future work.

5.6 Predicted versus Actual Use

Figure 6 presents predicted and actual FITSBY use in the Control condition, where participants had neither the bonus nor the limit functionality. As specified in our pre-analysis plan, we winsorize predicted use at no more than 60 minutes per day more or less than actual use in the corresponding period. Within each period, the left-most spike is actual average use. The spikes to the right are average predictions. The point estimates show that people consistently underestimate their use in all future periods, even though actual use is fairly stable throughout the experiment and the surveys had reminded them of their past use before eliciting predictions. This is consistent with naivete, implying $\tilde{\gamma} < \gamma$ in our model.

Figure 7 presents predicted versus actual habit formation. Within each period, the left-most point is the treatment effect of the bonus on actual use, reproduced from Figure 4. Recall that before the multiple price list for the Screen Time Bonus on survey 2, we asked people to report the percent by which they thought the bonus would reduce their FITSBY use. Their estimates (translated into minutes using their status quo predictions) are almost exactly correct on average: 52 minutes per day. Then on survey 3, we asked people to predict their use in future periods. Figure 7 also presents treatment effects of the bonus on predicted use, estimated from equation (4). The figure shows that people correctly predict that the bonus will reduce their consumption in period 3 and that this reduction will persist even after the incentive is no longer in effect. If anything, comparing the time path of actual versus predicted effects suggests that people overestimate the extent of habit formation. Overall, these results suggest that people are well aware of habit formation.

Appendix D.1 presents additional results that validate that the usage predictions are meaningful. Predicted use lines up well with actual use, and the higher ($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.

5.7 Bonus and Limit Valuations

On the survey 3 multiple price list, the average Limit group participant was willing to give up a $4.20 fixed payment for three weeks of access to the limit functionality. About 58 percent of participants were willing to give up at least some money for the limits, and 20 percent were willing to give up more than $10; see Appendix Figure A17. This willingness to pay for a commitment device is consistent with perceived self-control problems ($\tilde{\gamma} > 0$) and unmet market demand for digital self-control tools.

On the survey 2 multiple price list, people who perceive self-control problems should prefer the Screen Time Bonus over higher fixed payments, as the incentive helps bring future use in line with current preferences. We show in Appendix E.5 that participants' average valuation of the bonus is consistent with perceived self-control problems ($\tilde{\gamma} > 0$).

Appendix D.1 presents additional results that validate that the MPL responses are meaningful. First, participants' valuations of the bonus are correlated with the amount of money they could expect to earn. Second, the bonus and limit valuations are correlated with each other and with *limit tightness*, *ideal use*

change, *addiction scale*, *SMS addiction scale*, and other variables in expected ways. 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 higher average valuation.

5.8 Effects on Survey Outcomes

Figure 8 presents the effects of the bonus and limit treatments on the survey outcomes described in Section 3.7. The outcome variables are signed so more positive effects always correspond to less addiction and/or higher subjective well-being. Following our pre-analysis plan, when estimating effects on survey outcomes, we constrain the limit effect to be the same for surveys 3 and 4 (because we correctly anticipated similar "first stage" effects on FITSBY use in both periods 2 and 3) and we report the bonus effect only for survey 4 (because we correctly anticipated negligible "first stage" effects on FITSBY use in period 2).[16]

Figure 8 shows that both interventions significantly reduced self-reported measures of addiction. Appendix Table A6 presents coefficient estimates and p-values. The bonus effect is larger than the limit effect for five of the six variables, consistent with the bonus's larger effects on FITSBY use. The bonus decreased *ideal use change* by 0.41 standard deviations (about 9 percentage points), while the limit decreased it by 0.23 standard deviations (about 5 percentage points). Both interventions reduced *addiction scale* and *SMS addiction scale* by 0.08 to 0.16 standard deviations, or about 0.21–0.44 points on the 16-point *addiction scale*. Both interventions statistically significantly reduced the chance that people reported using their smartphones to relax to go to sleep, losing sleep from use, using longer than intended, using to distract from anxiety, having difficulty putting down their phone, using mindlessly, and other specific measures from the addiction scales; see Appendix Figures A23 and A24. The limit treatment statistically significantly increased the extent to which people thought their smartphone use made their life better, while the bonus did not.

The bonus and limit treatments increased subjective well-being (SWB) by 0.09 standard deviations ($p \approx 0.026$) and 0.04 standard deviations ($p \approx 0.18$) respectively. The sharpened False Discovery Rate-adjusted p-values (see Benjamini and Hochberg 1995) are 0.09 and 0.24, respectively. These SWB effects appear to be driven particularly by improved concentration and reduced distraction; see Appendix Figure A25. The effects of the bonus and limit on happiness, life satisfaction, depression, and anxiety are individually and collectively insignificant, while the effects of the bonus (but not the limit) on concentration, distraction, and sleep quality are collectively significant. Both interventions improved *survey index*, the inverse covariance-weighted average of the five survey outcome variables, by about 0.2 standard deviations.

One point of comparison for the SWB effects is Allcott, Braghieri, Eichmeyer, and Gentzkow (2020). They find that deactivating subjects' Facebook accounts for a four week period increased an index of SWB

[16] Appendix Figure A22 presents the treatment effects on survey outcomes separately for surveys 3 and 4. The limit effects on surveys 3 and 4 are statistically indistinguishable. Although the bonus did not substantially affect consumption in period 2, the Bonus group reported more ideal use reduction and more addiction on survey 3. One potential explanation is that the Bonus group hoped to reduce FITSBY use in anticipation of the period 3 incentive, and these survey responses reflect their failure to do so.

by a statistically significant 0.09 standard deviations. Although the two interventions had similar effects on time use—deactivation in Allcott, Braghieri, Eichmeyer, and Gentzkow (2020) reduced Facebook use by 60 minutes per day for 27 days, while our Screen Time Bonus reduced FITSBY use by 56 minutes per day for 20 days—they differed on a number of dimensions including the apps affected and the time period in which the study took place.

Appendix Figure A26 presents effects on *survey index* in subgroups with above- and below-median values of our six pre-specified moderators. There is little heterogeneity with respect to the first four moderators, other than that the limit seems to have larger effects on women. However, the effects of both interventions are 2–3 times larger for people with above-median baseline values of *restriction index*, which measures interest in restricting smartphone time use, and *addiction index.* This implies that the interventions are well-targeted: they have larger effects for people who report wanting and needing them the most. Consistent with this, point estimates suggest that the bonus and limit both have larger effects on FITSBY use for people with higher *restriction index* and *addiction index*, although the differences are not as significant; see Appendix Figure A27. This targeting result need not have been the case: for example, it could have been that more addicted people were less likely to feel that the limit functionality worked well for them.

6 Estimating the Model

6.1 Setup

We now turn to our model to simulate the effect of temptation on steady-state FITSBY use. In the model from Section 2, temptation and habit formation interact, because the current consumption increase caused by temptation also increases future consumption. The long-run effect of temptation could therefore be different than any effects identified during the experiment. In this section, we estimate the model's structural parameters. In the next section, we simulate steady-state FITSBY use with counterfactual self-control and habit formation parameters.

We estimate the model using indirect inference: we derive equations that characterize how a consumer from our model would behave in our experiment, and we solve for the structural parameters consistent with the data. In our baseline estimates, we assume that all parameters other than ξ are homogeneous across consumers, although we relax this assumption in an extension that allows heterogeneity in temptation γ and perceived temptation $\tilde{\gamma}$.

In describing the estimation strategy, we focus on a "restricted model" where we set the anticipatory bonus effect τ_2^B to zero. This implies full projection bias ($\alpha = 1$), and thus that consumption decisions maximize current-period flow utility with no dynamic considerations. This substantially simplifies the exposition and, as we will show, has little impact on the results. Appendix E presents our "unrestricted model," in which we use the empirical τ_2^B and allow partial projection bias.

In the restricted model, consumers maximize current-period flow utility from equation (3), giving equi-

librium consumption

$$x_t^*(s_t, \gamma, \boldsymbol{p}_t) = \frac{\zeta s_t + \xi_t - p_t + \gamma}{-\eta}. \tag{6}$$

We define $\lambda \coloneqq \frac{\partial x_t^*}{\partial s_t}$ as the effect of habit stock on consumption; $\lambda = -\zeta/\eta$ in the restricted model. In a steady state with constant s, ξ, and p, we must have $s_{ss} = \rho(s_{ss} + x_{ss})$, and thus $s_{ss} = \frac{\rho}{1-\rho} x_{ss}$.

We model the Screen Time Bonus as a price $p^B = \$2.50$ per hour in period 3 plus a fixed payment.[17] We model the limit functionality as an intervention that elimates share ω of perceived and actual temptation. We conservatively assume $\omega = 1$ in our primary estimates, and we consider alternative assumptions below. We assume that when predicting period t consumption on the survey at the beginning of period t, consumers use perceived temptation $\tilde{\gamma}$ but are aware of projection bias, so the prediction is denoted $x_t^*(s_t, \tilde{\gamma}, \boldsymbol{p}_t)$.

Figure 9 illustrates temptation, naivete, and our identification strategies. The three demand curves are desired demand $x_t^*(s_t, 0, \boldsymbol{p}_t)$ according to preferences before period t, predicted demand $x_t^*(s_t, \tilde{\gamma}, \boldsymbol{p}_t)$ as of survey t, and actual demand $x_t^*(s_t, \gamma, \boldsymbol{p}_t)$. The actual equilibrium at $p = 0$ is point L, and the predicted equilibrium is at point C, so the distance CL is Control group misprediction $m^C \coloneqq x_t^*(s_t, \gamma, \boldsymbol{p}_t) - x_t^*(s_t, \tilde{\gamma}, \boldsymbol{p}_t)$. The bonus moves the equilibrium to point J in period 3, so the contemporaneous bonus effect τ_3^B is the distance JK. The limit treatment moves the equilibrium to point G, so the limit treatment effect τ^L is the distance GL.

6.2 Estimating Equations

Unlike many applications of indirect inference, we derive equations that allow us to directly solve for the model parameters, so we do not need to use an optimization routine to search for parameters that fit the data. We estimate the parameters in stages, as parameters estimated in the first few equations are used as inputs to subsequent equations. We estimate confidence intervals by bootstrapping. Appendix G presents formal derivations and additional details.

Habit Formation

We first estimate ρ from the decay of the bonus treatment effects. Taking the expectations over ξ in the Bonus and Bonus Control groups, we can write the average treatment effect of the bonus on period 4 consumption as the result of the decayed period 3 effect. Similarly, the average treatment effect in period 5 results from the cumulative decayed effects from periods 3 and 4:

[17] Modeling the bonus as a linear price simplifies the model substantially, although it is an approximation: 13 percent of the Bonus group hit the \$150 payment limit because they reduced period 3 FITSBY use by more than three hours per day relative to their Bonus Benchmark, and 3.5 percent used more than their Bonus Benchmark. These two subgroups in practice faced zero subsidy for marginal screen time reductions, although they may not have known that.

$$\tau_4^B = \lambda \rho \tau_3^B \tag{7}$$

$$\tau_5^B = \lambda \left(\rho \tau_4^B + \rho^2 \tau_3^B \right). \tag{8}$$

Dividing those two equations gives

$$\rho = \frac{\tau_5^B}{\tau_4^B} - \frac{\tau_4^B}{\tau_3^B}. \tag{9}$$

This equation shows that if the bonus effect is more persistent between periods 4 and 5, we infer that habit stock is more persistent (a larger ρ).

In the unrestricted model in Appendix E, we also estimate λ, because it is useful in estimating the other parameters. To provide a comparison, we also estimate λ in the restricted model by rearranging equation (7) and inserting the ρ from equation (9): $\lambda = \frac{\tau_4^B}{\rho \tau_3^B}$.

Price Response and Habit Stock Effect on Marginal Utility

After estimating ρ, we estimate η and ζ from the magnitude and decay of the bonus treatment effects. For each of periods 3 and 4, we take the expectations over ξ of equilibrium consumption in the Bonus and Bonus Control groups, difference the two, and rearrange, giving

$$\eta = \frac{p^B}{\tau_3^B} \tag{10}$$

$$\zeta = \frac{-\eta \tau_4^B}{\rho \tau_3^B}. \tag{11}$$

Figure 9 illustrates the first equation: the inverse demand slope η is just the ratio of p^B (the vertical distance KL) to τ_3^B (the horizontal distance JK). The second equation shows that if the bonus effect is more persistent between periods 3 and 4, we infer that habit stock has a larger effect on marginal utility (a higher ζ).

Naivete about Temptation

Next, we estimate naivete about temptation $\gamma - \tilde{\gamma}$ using misprediction in the Control group. To solve for $\gamma - \tilde{\gamma}$, we take the expectations over ξ of actual consumption and consumption as predicted at the beginning of the period, difference the two, and rearrange, giving

$$\gamma - \tilde{\gamma} = -\eta m^C. \tag{12}$$

Figure 9 illustrates: naivete $\gamma - \tilde{\gamma}$ is the vertical distance *HC* between actual and predicted marginal utility, and this can be inferred by multiplying Control group average misprediction m^C (the horizontal distance *CL* between actual and predicted demand) by the inverse demand slope η.

Temptation

To estimate temptation γ, we take the expectations over ξ of equilibrium consumption in the Limit and Limit Control groups, difference the two, and rearrange, giving

$$\gamma = \eta \tau_2^L. \tag{13}$$

Figure 9 illustrates: temptation γ is the vertical distance *LM* between desired and actual demand, and this can be inferred by multiplying the effect of removing temptation (τ_2^L, the horizontal distance *GL* between long-run and present demand) by the inverse demand slope η. We then substitute the estimated γ into equation (12) to infer $\tilde{\gamma}$.

Intercept

Finally, we back out the distribution of ξ that fits the distribution of baseline consumption. We assume that participant i's baseline consumption x_{i1} was in a steady state. Substituting $s_{ss} = \frac{\rho}{1-\rho} x_{ss}$ into equilibrium consumption from equation (6) and rearranging gives

$$\xi_i = p - \gamma + x_{i1} \left(-\eta - \zeta \frac{\rho}{1-\rho} \right). \tag{14}$$

This equation shows that we infer larger ξ_i for people with higher baseline consumption x_{i1}.

In the unrestricted model in Appendix E, equilibrium consumption also depends on ϕ, the direct effect of habit stock on utility. Our data do not allow us to separately identify ϕ from ξ, so we estimate an "intercept" $\kappa_i := (1-\alpha)\delta\rho(\phi - \xi_i) + \xi_i$ that includes both of these structural parameters. In the restricted model with $\alpha = 1$, this simplifies to $\kappa_i = \xi_i$.

6.3 Empirical Moments

Table 3 presents the moments used to estimate the restricted model. The bonus and limit effects τ_t^B and τ_2^L are as displayed in Figure 4. Control group misprediction m^C is the average across periods 2–4 of the difference between actual period t FITSBY use and the prediction for period t elicited on survey t, as displayed in Figure 6. The unrestricted model and our robustness checks also use the anticipatory bonus effect τ_2^B and additional parameters presented in Appendix Table A7. In light of the discussion in Section 5.3, we omit the first half of period 2 when we estimate τ_2^B.[18]

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

6.4 Parameter Estimates

Table 4 presents our point estimates and bootstrapped 95 percent confidence intervals. Column 1 presents the restricted model described above (fixing $\tau_2^B = 0$ and $\alpha = 1$), while column 2 presents the unrestricted model described in Appendix E. Since the estimated τ_2^B is close to zero and $\hat{\alpha}$ is close to one, the estimates in the two columns are very similar.

In column 1, we estimate $\hat{\lambda} \approx 1.15$ and $\hat{\rho} \approx 0.299$. In our model, this implies that an exogenous consumption increase of 1 minute per day over a three week period will cause consumption to increase by $\hat{\lambda}\hat{\rho} \approx 0.34$ minutes per day in the next three-week period, and $\hat{\lambda}\hat{\rho}^2 \approx 0.10$ minutes per day in the period after that.

Consistent with the small and statistically insignificant anticipatory bonus effect τ_2^B in the second half of period 2, we estimate $\hat{\alpha} \approx 0.897$ in the unrestricted model in column 2, which is marginally significantly different from one. The point estimate suggests that participants were attentive to only $(1-\hat{\alpha}) \times 100\% \approx$ 10.3 percent of habit formation. Inserting the estimates of λ, ρ, η, and ζ into equation (24) in Appendix E, we calculate that τ_2^B would have needed to be -16.1 minutes per day (compared to the actual point estimate of -1.96 minutes per day in the second half of period 2) to estimate zero projection bias ($\alpha = 0$). In other words, the anticipatory bonus effect is only 12 percent of what our model would predict with fully forward-looking ("rational") habit formation. This is striking when combined with the evidence from Figure 7 that participants correctly predicted habit formation. It is consistent with a model in which people are intellectually aware of habit formation but consume as if they are inattentive to it.

Since the restricted model estimating equations are so simple, one can easily calculate the point estimates in column 1 with the moments from Table 3. For example, the Control group underestimated FITSBY use by an average of 6.13 minutes per day on surveys 2–4. Inserting that into equation (12) gives a naivete of $\widehat{\gamma - \tilde{\gamma}} = -\hat{\eta} \cdot m^C \approx -(-2.68) \cdot (6.13/60) \approx 0.274$ \$/hour in column 1.

The limit changed period 2 FITSBY use by -24.3 minutes per day. Inserting that into equation (13) gives temptation $\hat{\gamma} = \hat{\eta}\,\tau_2^L \approx (-2.68) \cdot (\text{-}24.3/60) \approx 1.09$ \$/hour in column 1. This estimate implies that a tax on FITSBY use of \$1.09 per hour would reduce consumption to the level our participants would choose for themselves in advance. Dividing estimated naivete $\widehat{\gamma - \tilde{\gamma}}$ by this $\hat{\gamma}$ suggests that our participants underestimate temptation by $0.274/1.09 \times 100\% \approx 25$ percent.

Appendix E.5 presents alternative estimates of temptation γ in the restricted and unrestricted models. First, we infer perceived temptation using participants' valuations of the limit functionality and the Screen Time Bonus, following Acland and Levy (2012), Augenblick and Rabin (2019), Chaloupka, Levy, and White (2019), Allcott, Kim, Taubinsky, and Zinman (2021), and Carrera et al. (2021). Second, we generalize the model to include multiple temptation goods, using the self-reports of substitution to FITSBY use on other devices discussed in Section 5.5. Third, we assume that the limit treatment eliminates share $\omega \in [0,1]$ of temptation, relaxing the assumption of $\omega = 1$ in our primary estimates; we estimate ω from differences in self-reported *ideal use change* between the Limit and Limit Control groups. Finally, we allow for individual-specific heterogeneity in γ, using the distribution of *limit tightness* set by Limit group participants. These

alternative approaches all imply temptation γ between about \$1 and \$3 per hour, and our primary estimate of \$1.09 per hour is relatively conservative.

7 Counterfactuals: Effects of Temptation on Time Use

7.1 Methodology

Using the parameter estimates from the previous section, we can predict the effects of changes in temptation and habit formation on steady-state FITSBY use. Equation (21) in Appendix E characterizes steady-state consumption in the unrestricted model. Using that equation, we can predict participant i's steady-state FITSBY use at $p = 0$ as a function of any values of habit formation, temptation, and steady-state misprediction parameters $\{\zeta, \gamma, \tilde{\gamma}, m_{ss}\}$:

$$\hat{x}_{i,ss}(\zeta, \gamma, \tilde{\gamma}, m_{ss}) = \frac{\hat{\kappa}_i + (1-\hat{\alpha})\delta\hat{\rho}\left[(\zeta - \hat{\eta})\, m_{ss} - \left(1 + \hat{\tilde{\lambda}}\right)\tilde{\gamma}\right] + \gamma}{-\hat{\eta} - (1-\hat{\alpha})\delta\hat{\rho}(\zeta - \hat{\eta}) - \zeta\frac{\hat{\rho} - (1-\hat{\alpha})\delta\hat{\rho}^2}{1-\hat{\rho}}}. \tag{15}$$

The sample average prediction is denoted $\bar{\hat{x}}_{ss}(\zeta, \gamma, \tilde{\gamma}, m_{ss})$. As discussed in Appendix E.3, we assume that the predicted $\tilde{\lambda}$ equals the estimated $\hat{\lambda}$, that steady-state misprediction m_{ss} equals observed Control group misprediction m^C, and that the discount factor is $\delta = 0.997$ per three-week period, consistent with a five percent annual discount rate.

Since we can't identify ϕ (the direct effect of habit stock on utility), we must hold constant each participant's intercept $\kappa_i := (1-\alpha)\delta\rho(\phi - \xi_i) + \xi_i$ across counterfactuals in the restricted model. Since this intercept contains ρ and α, we can't predict consumption with counterfactual values of ρ or α.

In the restricted model with $\alpha = 1$, equation (15) simplifies to

$$\hat{x}_{i,ss}(\rho, \gamma) = \frac{\hat{\xi}_i + \gamma}{-\hat{\eta} - \hat{\zeta}\frac{\rho}{1-\rho}}, \tag{16}$$

which could also be derived from substituting $s_{ss} = \frac{\rho}{1-\rho}x_{ss}$ into equation (6). Steady-state misprediction m_{ss} and perceived temptation $\tilde{\gamma}$ do not affect steady-state consumption in the restricted model because consumers simply maximize current-period flow utility.

7.2 Counterfactual Results

Figure 10 presents point estimates and bootstrapped 95 percent confidence intervals for predicted average FITSBY use at counterfactual parameter values. For each counterfactual, we present predictions from the restricted model ($\alpha = 1$) and unrestricted model ($\alpha = \hat{\alpha}$). We label the restricted model predictions as our primary results, because they are simpler and more conservative.

The first "counterfactual" is the baseline at our point estimates: $\hat{x}_{ss}\left(\hat{\zeta}, \hat{\gamma}, \hat{\tilde{\gamma}}, \hat{m}^C\right)$. This mechanically

matches baseline average FITSBY use of 153 minutes per day. The second counterfactual removes naivete: $\bar{\hat{x}}_{ss}\left(\hat{\zeta},\hat{\gamma},\hat{\gamma},0\right)$.[19] As described above, naivete has no effect when $\alpha = 1$. Because naivete is so small and projection bias is so strong, the point estimate with $\alpha = \hat{\alpha}$ is very close to the baseline.

The third counterfactual removes temptation: $\bar{\hat{x}}_{ss}\left(\hat{\zeta},0,0,0\right)$. Relative to baseline, removing temptation reduces predicted FITSBY use by 48 minutes per day (31 percent) with $\alpha = 1$. Thus, our primary estimate is that smartphone FITSBY use would be 31 percent lower without self-control problems.

The fourth and fifth counterfactuals remove habit formation, first with temptation and then without: $\bar{\hat{x}}_{ss}\left(0,\hat{\gamma},\hat{\hat{\gamma}},\hat{m}^C\right)$ and then $\bar{\hat{x}}_{ss}\left(0,0,0,0\right)$. We emphasize that habit formation on its own is not a departure from rationality (Becker and Murphy 1988), and it could capture forces such as learning and investment that increase consumer welfare. Relative to baseline, removing habit formation reduces predicted FITSBY use by 75 minutes per day with $\alpha = 1$. Without habit formation, the effect of removing temptation (going from the fourth to the fifth counterfactual) is just the limit treatment effect ($\tau_2^L \approx$ -24.3 minutes per day), which is about half of the effect of removing temptation with habit formation (47.5 minutes per day with $\alpha = 1$).[20] This quantifies how habit formation magnifies the effects of temptation, because current temptation increases current consumption and thus future demand.

We highlight one important tension in our results: Figure 4 shows that the limit effects decay slightly over periods 2–5, while our model predicts that the limit effects should grow over time as the Limit group's habit stock diminishes. One potential explanation is that habit formation works differently in response to prices versus self-control tools. Another potential explanation is that motivation to use the limit functionality decays enough that it outweighs the habit stock effect.

Appendix Table A15 presents 19 alternative estimates of the effects of temptation on steady-state FITSBY use across the restricted and unrestricted models. Consistent with the fact that our primary estimates of γ are smaller than most alternative estimates, our primary estimates of the steady-state temptation effects are also relatively conservative. Furthermore, weighting our sample on observables to look more like the U.S. adult population also increases the predicted effects of temptation on consumption. This means that while our sample may still be non-representative on unobservable characteristics, sample selection bias captured by observables causes us to *understate* the effects of temptation on FITSBY use.[21]

Since we don't identify ϕ (the direct effect of habit stock on utility), we can't do a full welfare analysis. The relatively elastic demand—from Section 5.3, 39 percent of consumption is worth less than \$2.50 per hour—suggests that participants do not have strong preferences over how to spend this marginal time, so the welfare losses from self-control problems might be limited. On the other hand, even small individual-level losses might be substantial when aggregated over many social media users. In a static model, the deadweight

[19] Since Figure 7 shows that participants predicted habit formation fairly accurately, we attribute all of steady-state misprediction m_{ss} to naivete about temptation.

[20] Without habit formation, the effect of removing temptation on $\bar{\hat{x}}_{ss}$ is $\frac{\hat{\gamma}}{-\hat{\eta}}$, which equals τ_2^L after substituting $\hat{\gamma} = \tau_2^L \hat{\eta}$.

[21] Appendix Tables A11–A13 present the demographics, moments, and parameter estimates in the weighted sample. Appendix Table A14 presents the numbers plotted in Figure 10. Appendix Figure A35 presents the distribution of modeled temptation effects across participants, using the Limit group's distribution of *limit tightness* to identify heterogeneity in temptation. The effect is less than 10 minutes per day for 26 percent of participants, and over 100 minutes per day for 13 percent.

loss from temptation would be the triangle *GLM* on Figure 9: $-\tau^L\gamma/2 \approx -(\text{-}24.3/60) \times 1.09/2 \approx \$0.22$ per day, or \$4.62 per three-week period. This is closely consistent with the average valuation of \$4.20 for three weeks of access to the limit functionality. Aggregated across 240 million American social media users (Pew Research Center 2021), this would be $\$4.62 \times (52/3) \times 0.24 \approx \$19.2$ billion per year in welfare losses from overuse of social media caused by self-control problems. For comparison, Facebook's total global profits in 2020 were \$29 billion (United States Securities and Exchange Commission 2020). However, we don't know how these effects would cumulate over time, as represented by ϕ: for example, after a longer period of reduced screen time, people might find more peace of mind or regret the loss of online interactions with friends and family.

8 Conclusion

While digital technologies provide important benefits, some argue that they can be addictive and harmful. We formalize this argument in an economic model and transparently estimate the parameters using data from a field experiment. The Screen Time Bonus intervention had persistent effects after the incentives ended, suggesting that smartphone social media use is habit forming. Participants predicted these persistent effects on surveys but did not reduce FITSBY use before the bonus was in effect, suggesting that they are aware of but inattentive to habit formation. Participants used the screen time limit functionality when we offered it in the experiment, and this functionality reduced FITSBY use by over 20 minutes per day, suggesting that social media use involves self-control problems. The Control group repeatedly underestimated future use, suggesting slight naivete. Many participants reported indicators of smartphone addiction on surveys, and both the bonus and limit interventions reduced this self-reported addiction. Looking at these facts through the lens of our economic model implies that self-control problems magnified by habit formation might be responsible for 31 percent of social media use. These results suggest that better aligning digital technologies with well-being should be an important goal of users, parents, technology workers, investors, and regulators.

Our results raise many additional questions; here are two. First, what are the underlying mechanisms and microfoundations that generate the persistent bonus treatment effects? We model this persistence simply through a capital stock of past consumption, but it could be driven by learning (followed by forgetting), network investments (e.g. connections with friends ebb and flow if maintained or neglected), or more nuanced habit formation mechanisms involving cues or automaticity (e.g. Laibson 2001; Bernheim and Rangel 2004; Steiny Wellsjo 2021). Second, if so many of our participants perceive self-control problems and use (and are willing to pay for) the Phone Dashboard time limit functionality, why isn't there higher demand for commercial digital self-control tools? Only 5 percent of our sample reported using any apps to limit their smartphone use at baseline. Potential explanations include that our experimental setting or selected set of participants overstates demand for commitment, that commercial self-control tools are too expensive or are ineffective because it's too easy to evade them or substitute across devices, that people aren't aware of existing tools, that the time misallocated due to temptation is not very valuable, or that the

commitment and flexibility features we built into Phone Dashboard were better suited to people's needs. We leave these questions for future work.

References

Acland, Dan and Vinci Chow. 2018. "Self-Control and Demand for Commitment in Online Game Playing: Evidence from a Field Experiment." *Journal of the Economic Science Association* 4 (1):46–62.

Acland, Dan and Matthew R. Levy. 2012. "Naivete, Projection Bias, and Habit Formation in Gym Attendance." Working Paper: GSPP13-002.

———. 2015. "Naiveté, Projection Bias, and Habit Formation in Gym Attendance." *Management Science* 61 (1):146–160.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110 (3):629–76.

Allcott, Hunt, Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman. 2021. "Are High-Interest Loans Predatory? Theory And Evidence From Payday Lending." Review of Economic Studies, forthcoming.

Allcott, Hunt and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* 104 (10):3003–37.

Alter, Adam. 2018. *Irresistible: the Rise of Addictive Technology and the Business of Keeping Us Hooked*. Penguin Press.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484):1481–1495.

Andreassen, Cecilie Schou, Torbjørn Torsheim, Geir Scott Brunborg, and Ståle Pallesen. 2012. "Development of a Facebook Addiction Scale." *Psychological Reports* 110 (2):501–517.

Andreoni, James and Charles Sprenger. 2012a. "Estimating Time Preferences from Convex Budgets." *American Economic Review* 102 (7):3333–3356.

———. 2012b. "Risk Preferences Are Not Time Preferences." *American Economic Review* 102 (7):3357–3376.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *The Quarterly Journal of Economics* 121 (2):673–697.

Augenblick, Ned. 2018. "Short-Term Discounting of Unpleasant Tasks." Working Paper.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working Over Time: Dynamic Inconsistency In Real Effort Tasks." *The Quarterly Journal of Economics* 130 (3):1067–1115.

Augenblick, Ned and Matthew Rabin. 2019. "An Experiment on Time Preference and Misprediction in Unpleasant Tasks." *The Review of Economic Studies* 86 (3):941–975.

Auld, M Christopher and Paul Grootendorst. 2004. "An Empirical Analysis of Milk Addiction." *Journal of Health Economics* 23 (6):1117–1133.

Bai, Liag, Benjamin Handel, Edward Miguel, and Gautam Rao. 2018. "Self-Control and Demand for Preventive Health: Evidence from Hypertension in India." NBER Working Paper No. 23727.

Banerjee, Abhijit and Sendhil Mullainathan. 2010. "The Shape of Temptation: Implications for the Economic Lives of the Poor." NBER Working Paper No. 15973.

Becker, Gary S, Michael Grossman, and Kevin M Murphy. 1994. "An Empirical Analysis of Cigarette Addiction." *The American Economic Review* 84 (3):396–418.

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

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1):289–300.

Bernedo, Maria, Paul J Ferraro, and Michael Price. 2014. "The Persistent Impacts of Norm-Based Messaging and Their Implications for Water Conservation." *Journal of Consumer Policy* 37 (3):437–452.

Bernheim, B. Douglas and Antonio Rangel. 2004. "Addiction and Cue-Triggered Decision Processes." *American Economic Review* 94 (5):1558.

Beshears, John, James J. Choi, Christopher Harris, David Laibson, Brigitte C Madrian, and Jung Sakong. 2015. "Self-Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits?" NBER Working Paper No. 21474.

Beshears, John and Katherine Milkman. 2017. "Creating Exercise Habits Using Incentives: The Tradeoff between Flexibility and Routinization." Working Paper, available at https://www.semanticscholar.org/paper/Creating-Exercise-Habits-Using-Incentives-

Bianchi, Adriana and James G Phillips. 2005. "Psychological Predictors of Problem Mobile Phone Use." *CyberPsychology & Behavior* 8 (1):39–51.

Brandon, Alec, Paul J Ferraro, John A List, Robert D Metcalfe, Michael K Price, and Florian Rundhammer. 2017. "Do The Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments." NBER Working Paper No. 23277.

Brown, Eileen. 2019. "Americans spend far more time on their smartphones than they think." Available at https://www.zdnet.com/article/americans-spend-far-more-time-on-their-smartphones-than-they-think/.

Bursztyn, Leonardo, Davide Cantoni, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2020. "Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements." NBER Conference Paper.

Busse, Meghan R, Devin G Pope, Jaren C Pope, and Jorge Silva-Risso. 2015. "The Psychological Effect of Weather on Car Purchases." *Quarterly Journal of Economics* 130 (1):371–414.

Carrera, Mariana, Heather Royer, Mark Stehr, and Justin Sydnor. 2018. "Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members." *Journal of Health Economics* 58:202–214.

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky. 2021. "Who Chooses Commitment? Evidence and Welfare Implications." Working Paper.

Carroll, Gabriel D, James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick. 2009. "Optimal Defaults and Active Decisions." *The Quarterly Journal of Economics* 124 (4):1639–1674.

Casaburi, Lorenzo and Rocco Macchiavello. 2019. "Demand and Supply of Infrequent Payments as a Commitment Device: Evidence from Kenya." *American Economic Review* 109 (2):523–555.

Chaloupka, Frank. 1991. "Rational Addictive Behavior and Cigarette Smoking." *Journal of Political Economy* 99 (4):722–742.

Chaloupka, Frank and Kenneth Warner. 1999. "The Economics of Smoking." NBER Working Paper No. 7047.

Chaloupka, Frank J, Matthew R Levy, and Justin S White. 2019. "Estimating Biases in Smoking Cessation: Evidence from a Field Experiment." NBER Working Paper No. 26522.

Charness, Gary and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica* 77 (3):909–931.

Collis, Avinash and Felix Eggers. 2019. "Effects of Restricting Social Media Usage." Available at SSRN: https://ssrn.com/abstract=3518744.

DellaVigna, Stefano and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3):694–719.

Deloitte. 2018. "2018 Global Mobile Consumer Survey: US Edition." Available at https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-tmt-global-mobile-consumer-survey-exec-summary-2018.pdf.

Do, Quy Toan and Hanan G Jacoby. 2020. "Sophisticated Policy with Naive Agents: Habit Formation and Piped Water in Vietnam." Available at SSRN: https://ssrn.com/abstract=3571024.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101 (6):2350–2390.

Ericson, Keith Marzilli and David Laibson. 2019. *Intertemporal Choice*, vol. 2, chap. 1. Elsevier, 1 ed.

Exley, Christine L. and Jeffrey K. Naecker. 2017. "Observability Increases the Demand for Commitment Devices." *Management Science* 63 (10):3147–3529.

Eyal, Nir. 2020. *Indistractable: How to Control Your Attention and Choose Your Life*. Bloomsbury Publishing PLC.

Fang, Hanming and Dan Silverman. 2004. "Time Inconsistency and Welfare Program Participation: Evidence from the NLSY." Cowles Foundation Discussion Paper No. 1465.

Ferraro, Paul J, Juan Jose Miranda, and Michael K Price. 2011. "The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment." *American Economic Review* 101 (3):318–22.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Economic Journal: Applied Economics* 8 (4):160–188.

Gerber, Alan S, Donald P Green, and Ron Shachar. 2003. "Voting May be Habit-Forming: Evidence from a Randomized Field Experiment." *American Journal of Political Science* 47 (3):540–550.

Gine, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics* 2:213–235.

Goda, Gopi Shah, Matthew R. Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff. 2015. "The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings." NBER Working Paper No. 21482.

Gosnell, Greer K, John A List, and Robert D Metcalfe. 2020. "The Impact of Management Practices on Employee Productivity: A Field Experiment With Airline Captains." *Journal of Political Economy* 128 (4):1195–1233.

Griffiths, Mark. 2005. "A "Components" Model of Addiction Within a Biopsychosocial Framework." *Journal of Substance Use* 10 (4):191–197.

Gruber, Jonathan and Botond Köszegi. 2001. "Is Addiction "Rational"? Theory and Evidence." *Quarterly Journal of Economics* 116 (4):1261–1303.

Gul, Faruk and Wolfgang Pesendorfer. 2007. "Welfare without Happiness." *American Economic Review* 97 (2):471–476.

Hawley, Josh. 2019. "S. 2314 (116th): SMART Act." Available at https://www.govtrack.us/congress/bills/116/s2314/text.

Hoong, Ruru. 2021. "Self Control and Smartphone Use: An Experimental Study of Soft Commitment Devices." Harvard Working Paper.

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

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2019. "Rational Habit Formation: Experimental Evidence from Handwashing in India." Available at SSRN: https://ssrn.com/abstract=3040729.

Irvine, Mark. 2018. "Facebook Ad Benchmarks for YOUR Industry." https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks.

John, Anett. 2019. "When Commitment Fails - Evidence from a Field Experiment." *Management Science* 66 (2):503–529.

John, Leslie K, George Loewenstein, Andrea B Troxel, Laurie Norton, Jennifer E Fassbender, and Kevin G Volpp. 2011. "Financial Incentives for Extended Weight Loss: a Randomized, Controlled Trial." *Journal of General Internal Medicine* 26 (6):621–626.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-Control at Work." *Journal of Political Economy* 123 (6):1227–1277.

Kemp, Simon. 2020. "Digital 2020 Reports." Available at wearesocial.com/digital-2020.

Kuchler, Theresa and Michaela Pagel. 2018. "Sticking to Your Plan: The Role of Present Bias for Credit Card Paydown." NBER Working Paper No. 24881.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2):443–478.

———. 2001. "A Cue-Theory of Consumption." *The Quarterly Journal of Economics* 116 (1):81–119.

———. 2018. "Private Paternalism, the Commitment Puzzle, and Model-Free Equilibrum." *AEA Papers and Proceedings* 108:1–21.

Laibson, David, Peter Maxted, Andrea Repetto, and Jeremy Tobacman. 2015. "Estimating Discount Functions with Consumption Choices over the Lifecycle." Working Paper.

Levitt, Steven D, John A List, and Sally Sadoff. 2016. "The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment." NBER Working Paper No. 22107.

Liu, Zhuang, Michael Sockin, and Wei Xiong. 2020. "Data Privacy and Temptation." NBER Working Paper No. 27653.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118 (4):1209–1248.

Madrian, Brigitte C and Dennis F Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *The Quarterly Journal of Economics* 116 (4):1149–1187.

Makarov, Uliana. 2011. "Networking or not working: A model of social procrastination from communication." *Journal of Economic Behavior & Organization* 2011 (80):574–585.

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

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie. 2019. "The Economic Effects of Facebook." *Experimental Economics* :1–28.

New York Post. 2017. "Americans Check Their Phones 80 Times a Day: Study." Available at https://nypost.com/2017/11/08/americans-check-their-phones-80-times-a-day-study/.

Newport, Cal. 2019. *Digital Minimalism: Choosing a Focused Life in a Noisy World*. Penguin Random House.

O'Donoghue, Ted and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89 (1):103–124. URL https://www.aeaweb.org/articles?id=10.1257/aer.89.1.103.

Paserman, M. Daniele. 2008. "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation." *The Economic Journal* 118:1418–1452.

Pew Research Center. 2021. "Social Media Fact Sheet." URL https://www.pewresearch.org/internet/fact-sheet/social-media/.

Read, Danieal and Barbara Van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* 76 (2):189–205.

Rees-Jones, Alex and Kyle T. Rozema. 2020. "Price Isn't Everything: Behavioral Response around Changes in Sin Taxes."

Royer, Heather, Mark Stehr, and Justin Sydnor. 2015. "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." *American Economic Journal: Applied Economics* 7 (3):51–84.

Sadoff, Sally, Anya Savikhin Samek, and Charles Sprenger. 2020. "Dynamic Inconsistency in Food Choice: Experimental Evidence from a Food Desert." *Review of Economic Studies* 87 (4):1954–1988.

Sagioglu, Christina and Tobias Greitemeyer. 2014. "Facebook's Emotional Consequences: Why Facebook Causes a Decrease in Mood and Why People Still Use It." *Computers in Human Behavior* 35:359–363.

Schilbach, Frank. 2019. "Alcohol and Self-Control: A Field Experiment in India." *American Economic Review* 109 (4):1290–1322.

Shapiro, Jesse M. 2005. "Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle." *Journal of Public Economics* 89:303–325.

Shui, Haiyan and Lawrence M Ausubel. 2005. "Time Inconsistency in the Credit Card Market." Working Paper.

Skiba, Paige Marta and Jeremy Tobacman. 2018. "Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default." Working Paper.

Steiny Wellsjo, Alexandra. 2021. "Simple Actions, Complex Habits: Lessons from Hospital Hand Hygiene." Access on https://drive.google.com/file/d/1wbn6IuU0tMQ2VN6YHSWSCXv4v9pucKyK/view.

Strack, Philipp and Dmitry Taubinsky. 2021. "Dynamic Preference "Reversals" and Time Inconsistency." NBER Working Paper No. 28961.

Toussaert, Severine. 2018. "Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment." *Econometrica* 86 (3):859–889.

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

United States Securities and Exchange Commission. 2020. "Official Facebook 2020 10-K report as filed with SEC." Available at https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/4dd7fa7f-1a51-4ed9-b9df-7f42cc3321eb.pdf.

Van Soest, Daan and Ben Vollaard. 2019. "Breaking Habits." Working Paper.

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

Vox. 2020. "Tech Companies Tried to Help us Spend Less Time on Our Phones. It Didn't Work." Available at https://www.vox.com/recode/2020/1/6/21048116/tech-companies-time-well-spent-mobile-phone-usage-data.

World Health Organization. 2018. "Gaming Disorder." Available at https://www.who.int/features/qa/gaming-disorder/en/.

Wurmser, Yoram. 2020. "US Mobile Time Spent 2020." Available at https://www.emarketer.com/content/us-mobile-time-spent-2020.

Zenith Media. 2019. "Consumers Will Spend 800 Hours Using Mobile Internet Devices This Year." Available at https://www.zenithmedia.com/consumers-will-spend-800-hours-using-mobile-internet-devices-this-year/.

Table 1: **Experiment Timeline and Sample Sizes**

Phase	Date	Sample size
Recruitment and intake	March 22 - April 8	3,271,165 shown ads 26,101 clicked on ads 18,589 passed screen 8,514 consented 5,320 finished intake survey
Survey 1 (baseline)	April 12	4,134 began Survey 1 4,038 finished Survey 1 2,126 were randomized
Survey 2	May 3	2,068 began Survey 2 2,053 informed of treatment, of which: 2,048 were not in MPL group 2,032 finished Survey 2
Survey 3	May 24	1,993 began Survey 3 1,981 finished Survey 3
Survey 4	June 14	1,954 began Survey 4 1,948 finished Survey 4
Completion	July 26	1,938 kept Phone Dashboard through July 26, of which: 1,933 were not in MPL group ("analysis sample")

Table 2: **Sample Demographics**

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

Notes: Column 1 presents average demographics for our analysis sample, and column 2 presents average demographics of American adults using data from the 2018 American Community Survey.

Table 3: **Empirical Moments for Restricted Model Estimation**

Parameter	Description	(1) Point estimate	(2) Confidence interval
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	Control group misprediction (minutes/day)	6.13	$[4.52, 7.72]$
$\bar{x}_1$	Average baseline use (minutes/day)	153	$[149, 157]$

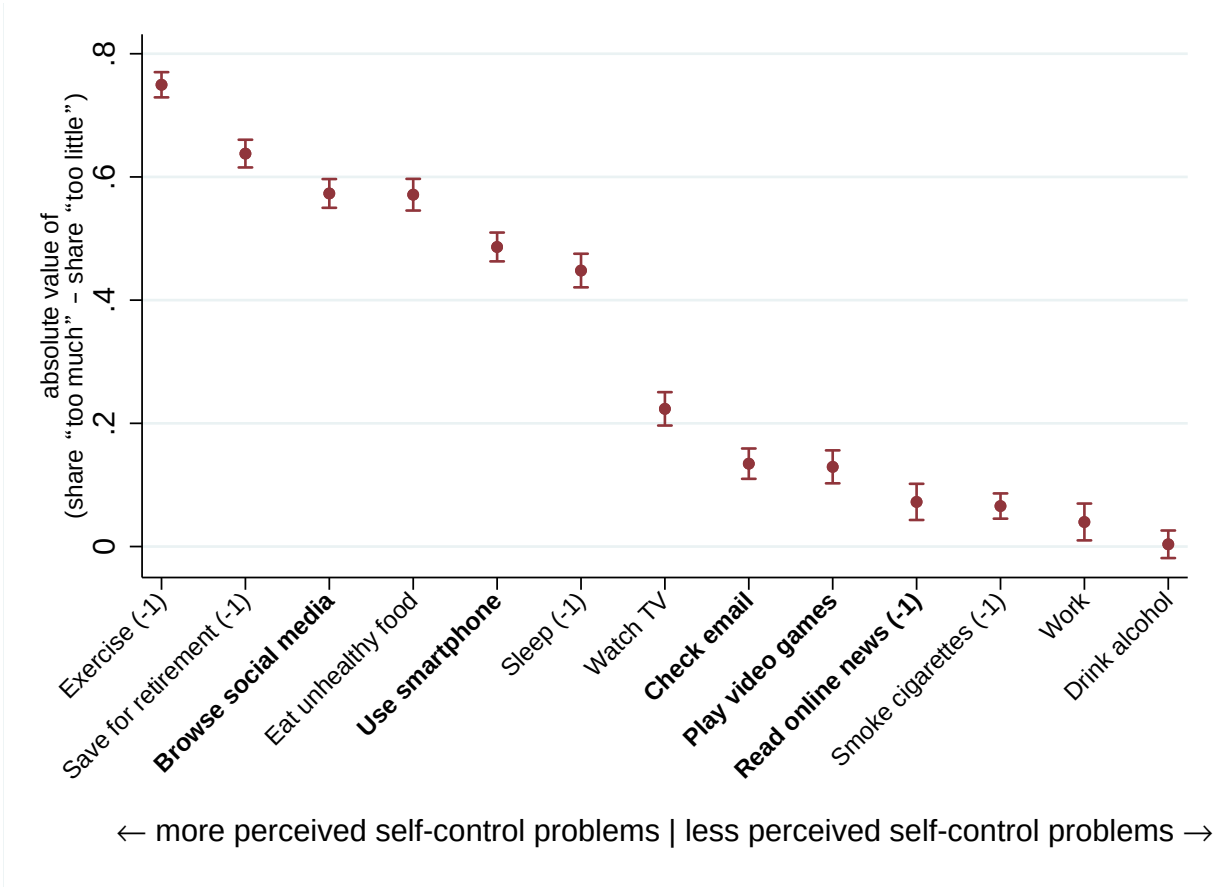
Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for the empirical moments used for our primary estimates of the restricted model.

Table 4: **Primary Parameter Estimates**

Parameter	Description (units)	(1) Restricted model ($\tau_2^B = 0$, $\alpha = 1$)	(2) Unrestricted model ($\alpha = \hat{\alpha}$)
λ	Habit stock effect on consumption (unitless)	1.15	1.12
		$[0.609, 3.31]$	$[0.572, 3.16]$
ρ	Habit formation (unitless)	0.299	0.302
		$[0.106, 0.493]$	$[0.106, 0.498]$
α	Projection bias (unitless)	1	0.897
			$[0.584, 1.00]$
η	Price coefficient (\$-day/hour^2)	-2.68	-2.75
		$[-2.98, -2.43]$	$[-3.04, -2.51]$
ζ	Habit stock effect on marginal utility (\$-day/hour^2)	3.08	3.01
		$[1.65, 8.97]$	$[1.55, 8.57]$
$\gamma - \tilde{\gamma}$	Naivete about temptation (\$/hour)	0.274	0.278
		$[0.201, 0.349]$	$[0.205, 0.354]$
γ	Temptation (\$/hour)	1.09	1.11
		$[0.884, 1.30]$	$[0.903, 1.33]$
$\bar{\kappa}$	Average intercept (\$/hour)	-2.41	-2.24
		$[-3.62, -1.10]$	$[-3.53, -0.803]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals from the estimation strategy described in Section 6.2 and Appendix E.3.

Figure 1: **Online and Offline Temptation**



Notes: This figure presents responses to the following question, which we asked participants in our experiment during the baseline survey, "For each of the activities below, please tell us whether you think you do it too little, too much, or the right amount." The bars are ordered from left to right in order of largest to smallest absolute value of (share "too little" – share "too much").

Figure 2: **Experimental Design**

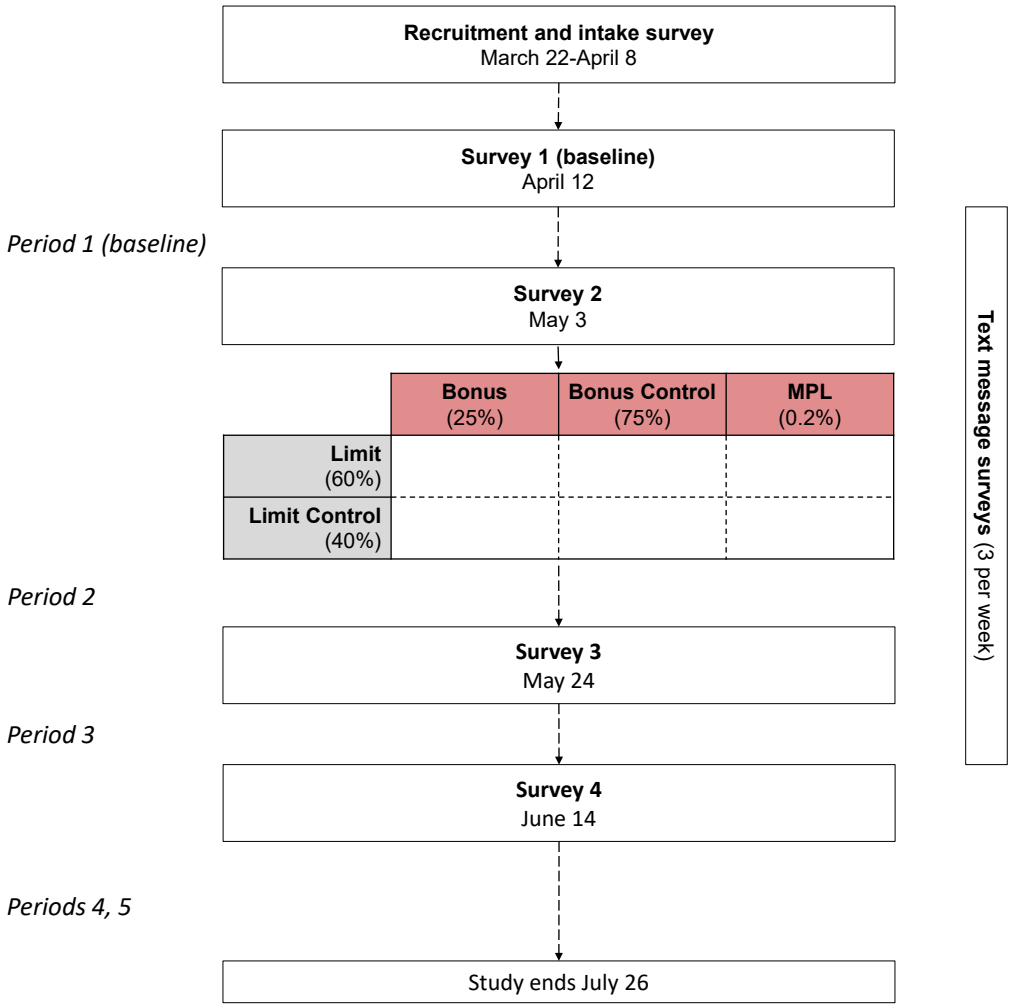
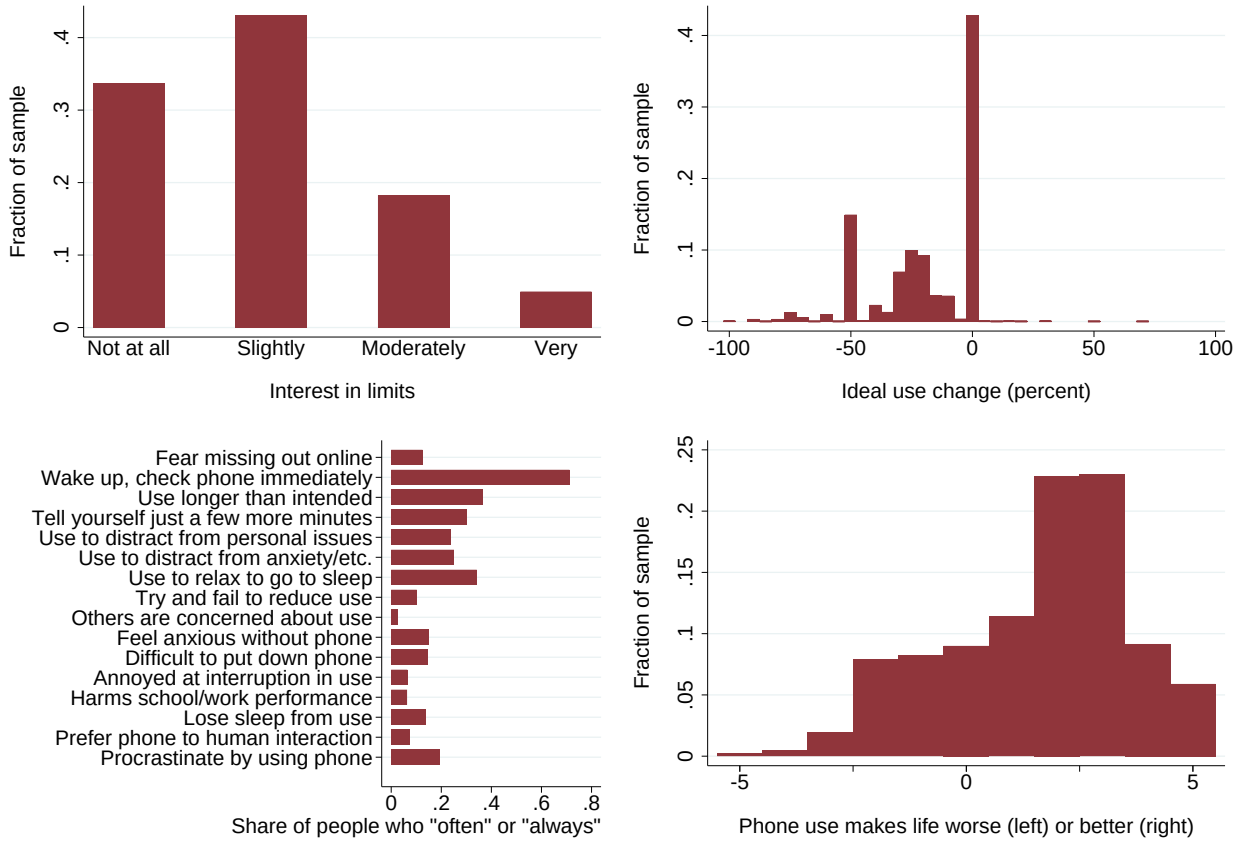
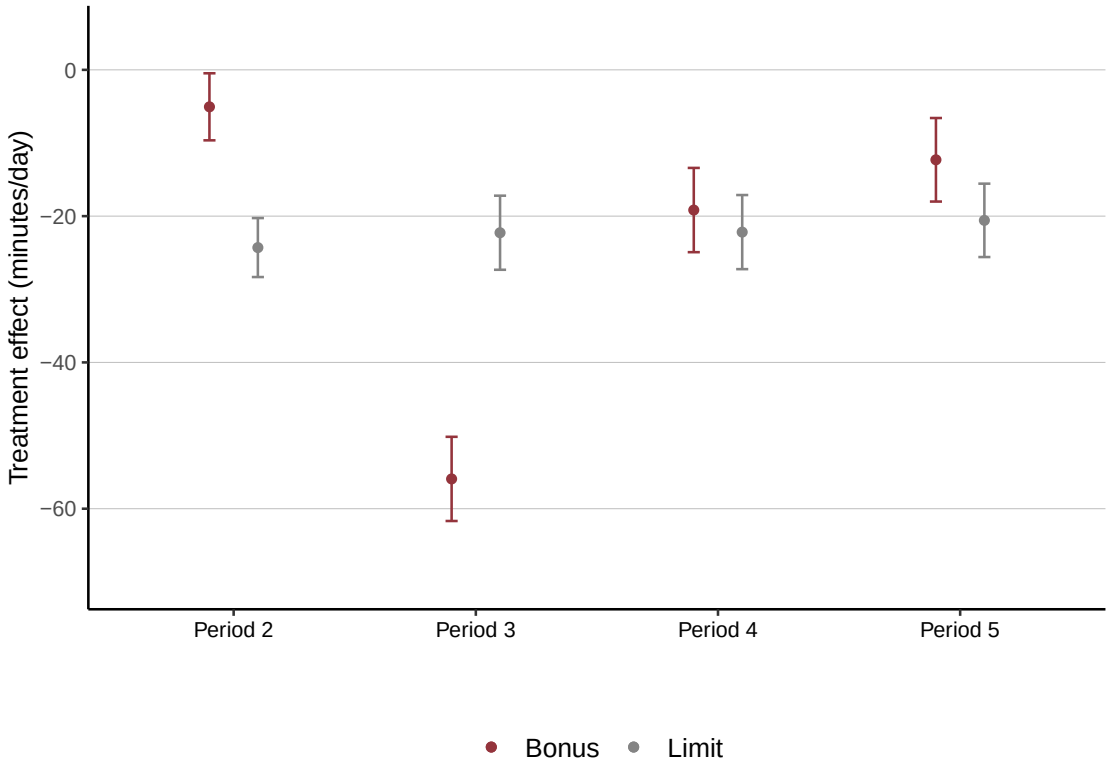


Figure 3: **Baseline Qualitative Evidence of Self-Control Problems**



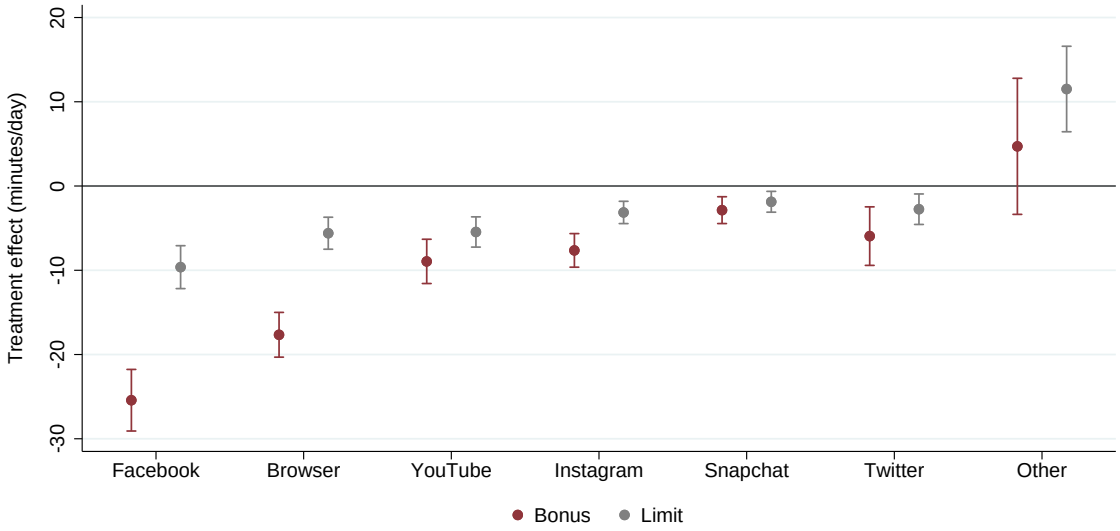
Notes: This figure presents the distributions of four measures of smartphone addiction from the baseline survey. *Interest in limits* is the answer to, "How interested are you to set limits on your phone use?" *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" The bottom left panel presents the share of participants who responded "often" or "always" to each of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *Phone use makes life worse or 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?"

Figure 4: **Treatment Effects on FITSBY Use**



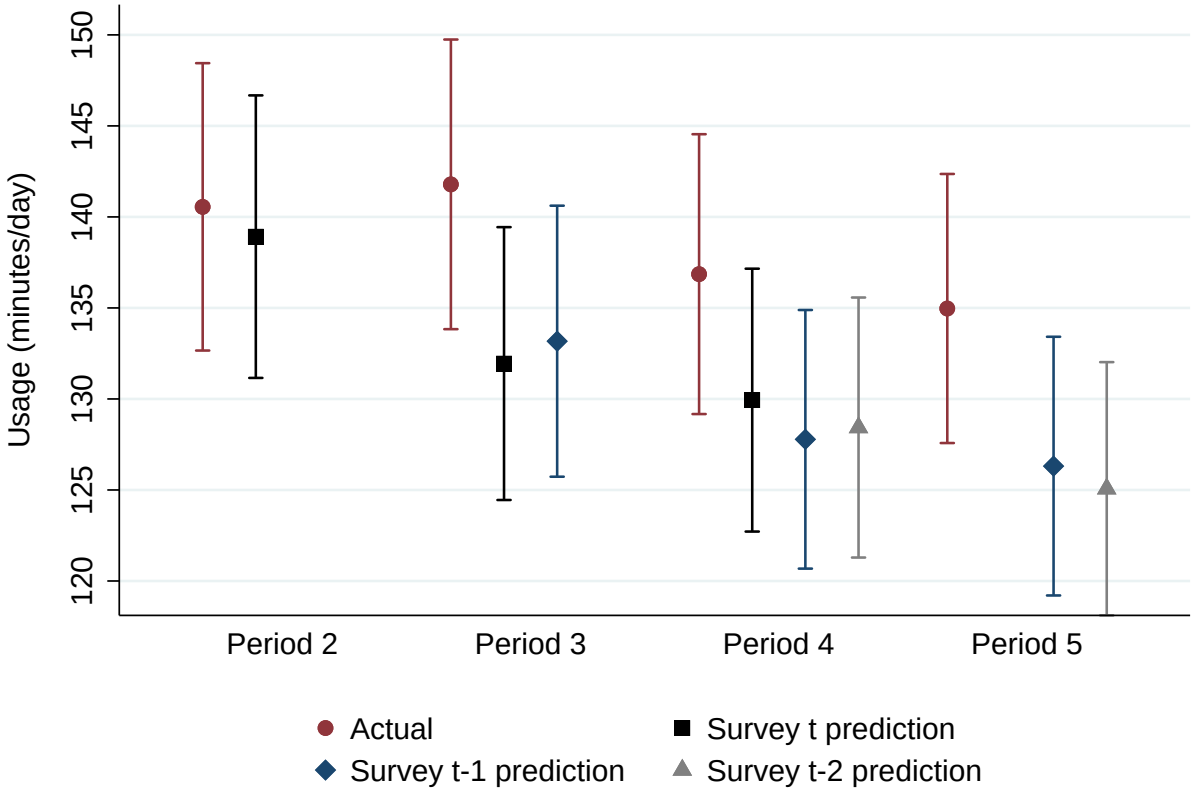
Notes: This figure presents effects of the bonus and limit treatments on FITSBY use using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure 5: **Effects on Smartphone Use by App**



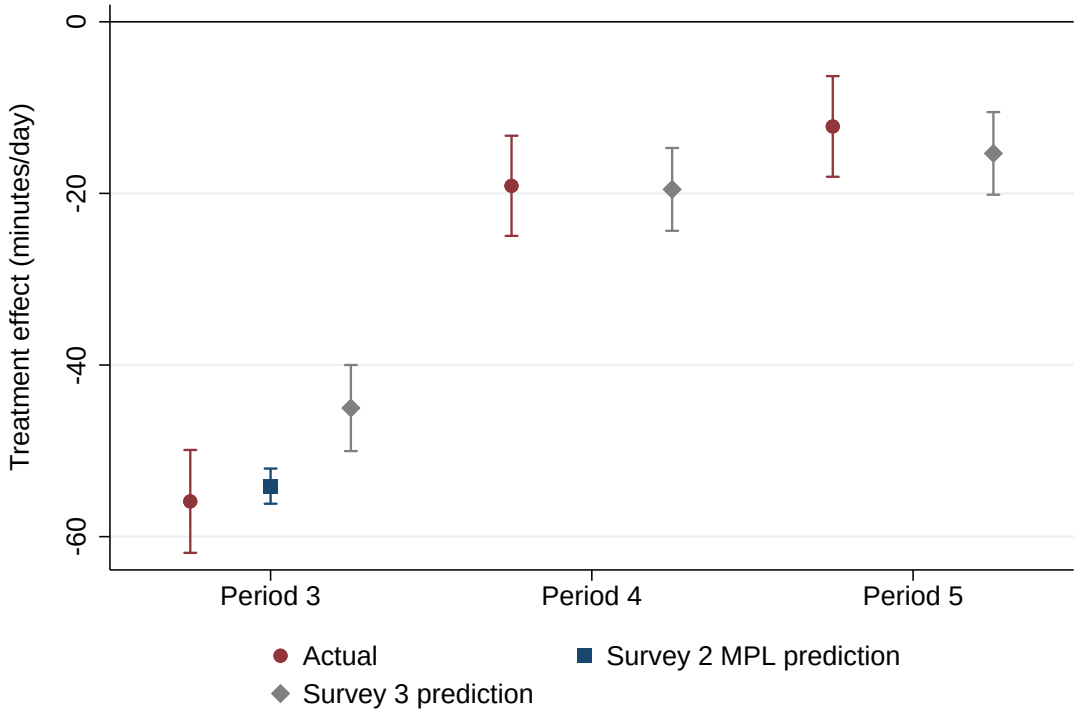
Notes: This figure presents effects of the bonus and limit treatments on smartphone use by app using equation (4). The bonus effects are measured in period 3, while the limit effects are measured in periods 2–5. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. FITSBY apps are in order of decreasing period 1 use.

Figure 6: **Predicted vs. Actual FITSBY Use in Control Conditions**



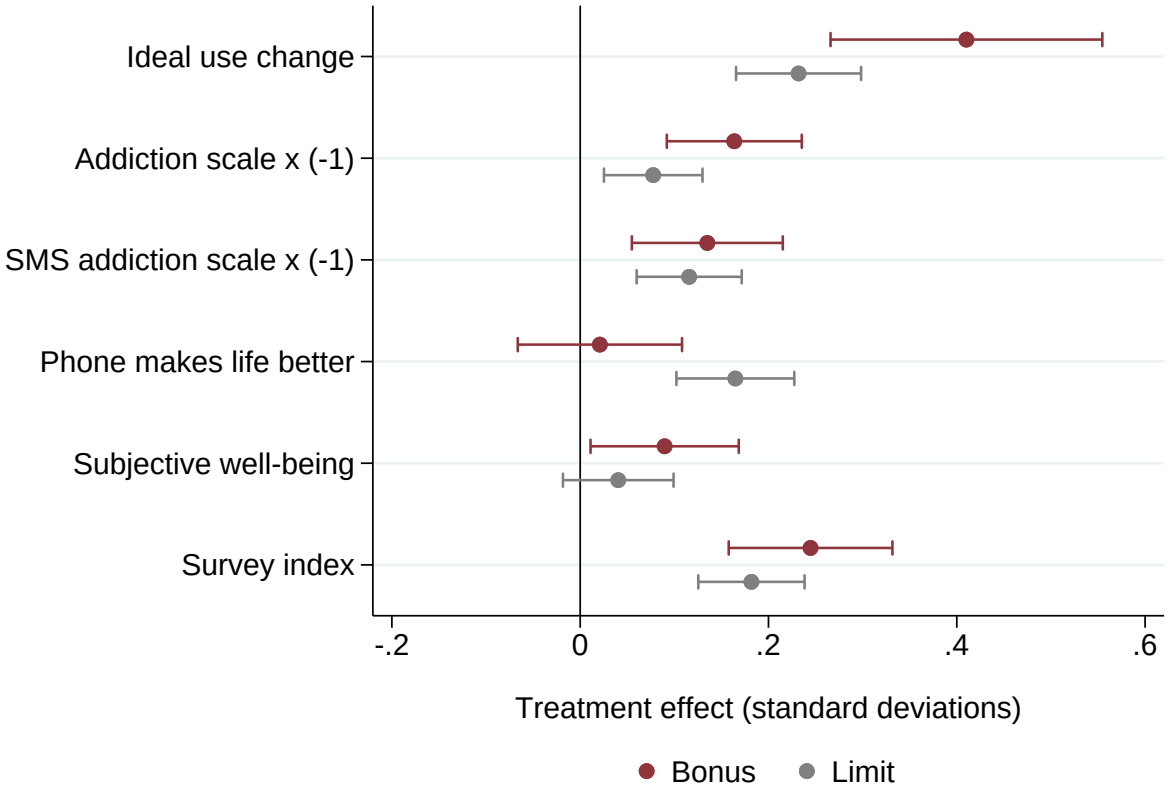
Notes: This figure presents average actual FITSBY use by period and average predicted FITSBY use for that period, for participants in the intersection of the Bonus Control and Limit Control groups. Period t is the three weeks immediately after survey t, so "survey t prediction" is the prediction for period t made just prior to period t. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure 7: **Predicted vs. Actual Habit Formation**



Notes: This figure presents the treatment effects of the bonus on FITSBY use and on predicted FITSBY use from survey 3 using equation (4), as well as the average predicted bonus treatment effect elicited on survey 2 before the bonus multiple price list. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure 8: **Effects of Limits and Bonus on Survey Outcome Variables**



Notes: This figure 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.

Figure 9: **Model Identification**

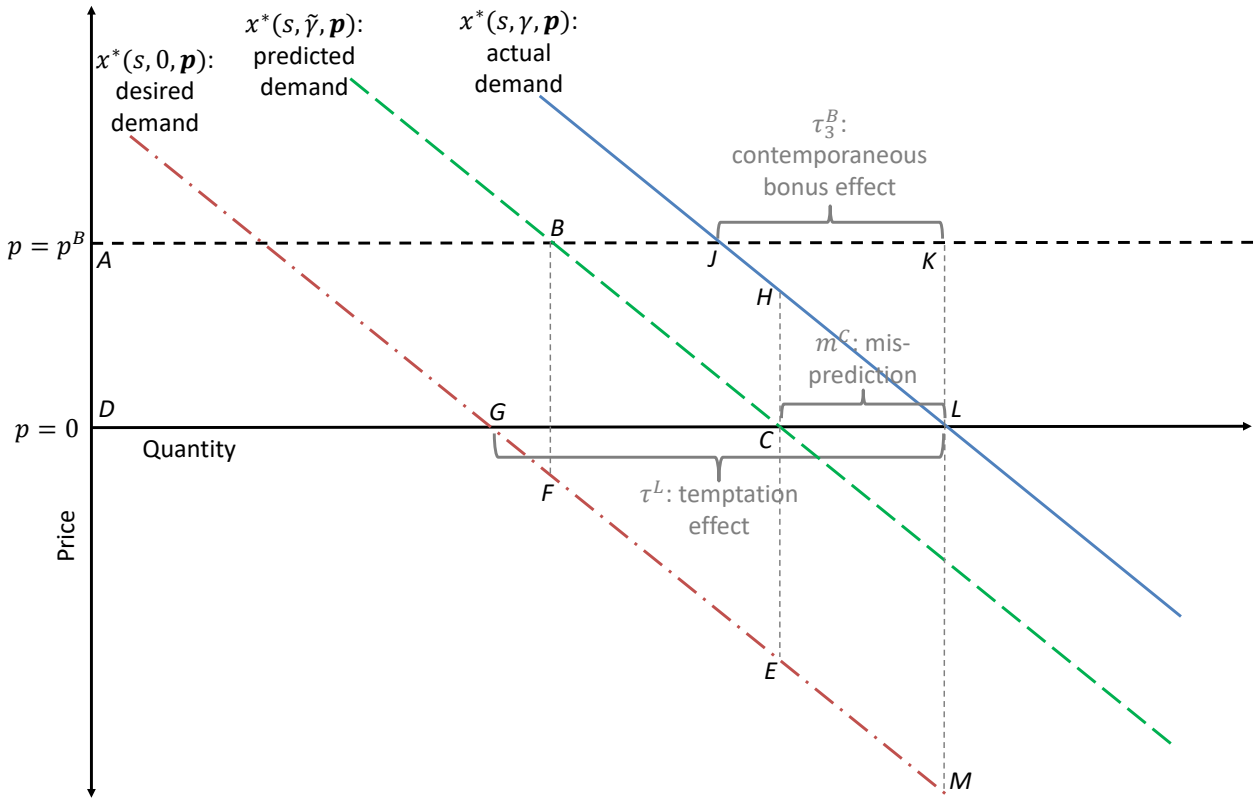
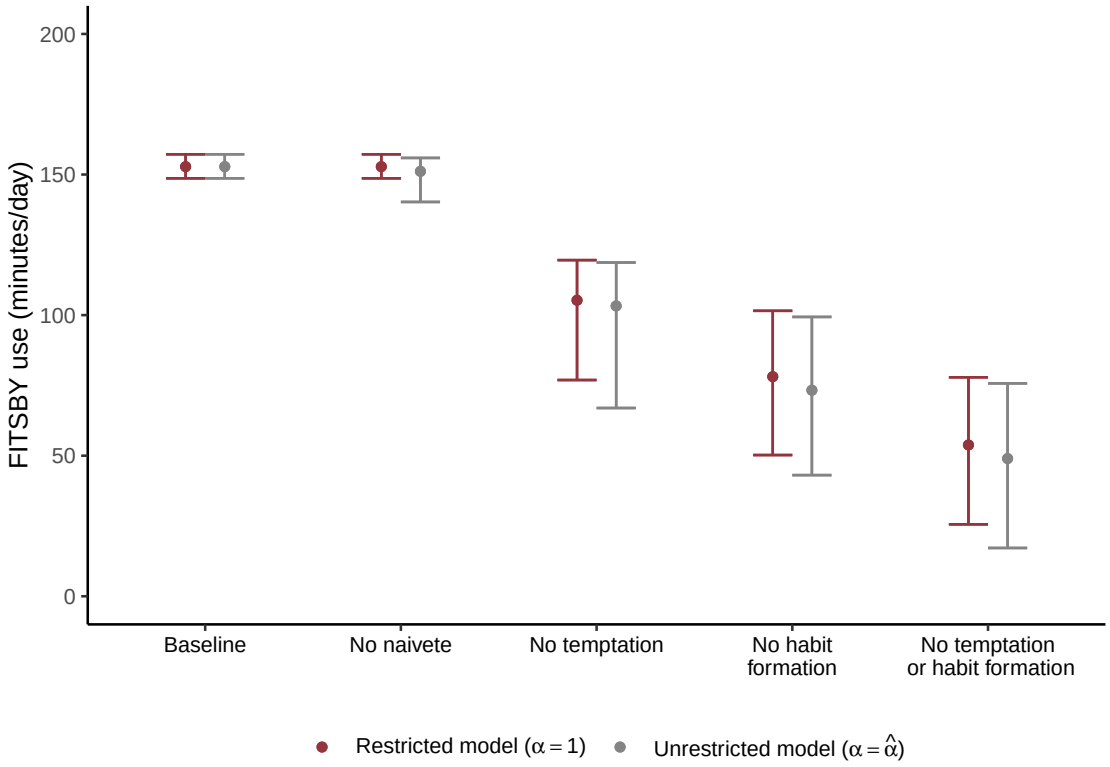


Figure 10: **Effects of Temptation and Habit Formation on FITSBY Use**



Notes: This figure presents point estimates and bootstrapped 95 percent confidence intervals for predicted steady-state FITSBY use with different parameter assumptions, using equation (15).

Online Appendix

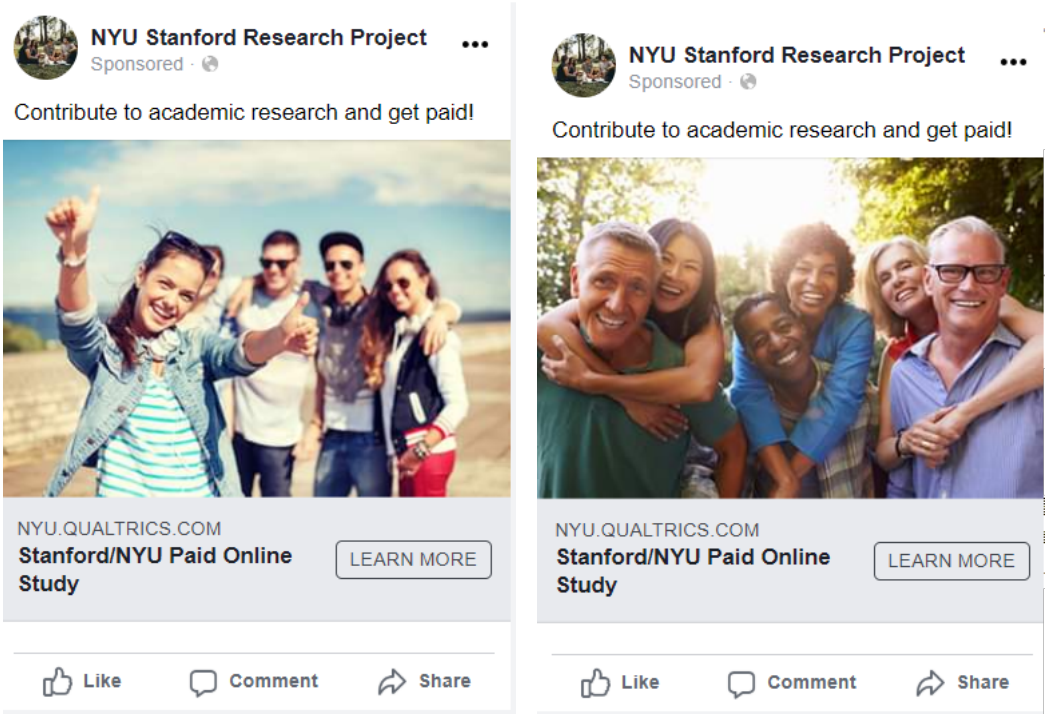
Digital Addiction

Hunt Allcott, Matthew Gentzkow, and Lena Song

Table of Contents

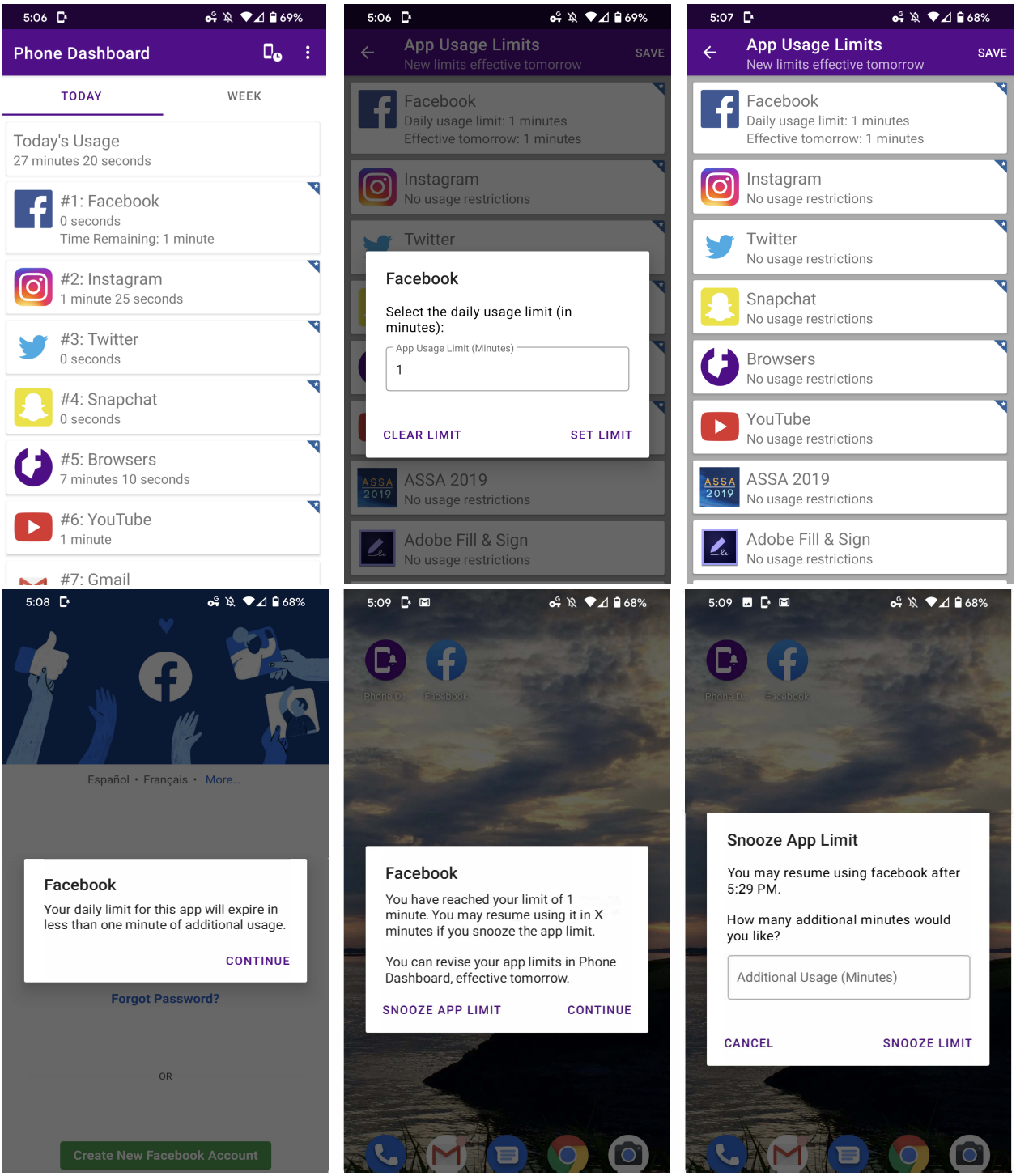
A Experimental Design Appendix

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**

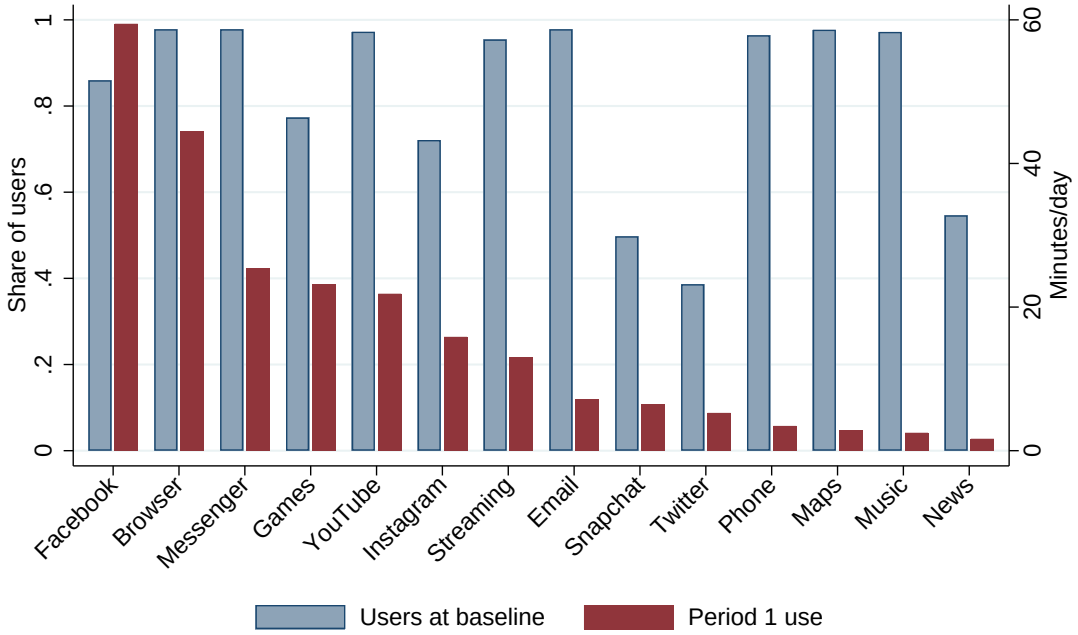
Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	t-test p-value (1)-(2)
Income ($000s)	40.15	41.76	0.35
	(36.22)	(37.84)	
College	0.67	0.67	0.72
	(0.47)	(0.47)	
Male	0.38	0.40	0.51
	(0.49)	(0.49)	
White	0.70	0.74	0.13
	(0.46)	(0.44)	
Age	33.61	33.79	0.76
	(12.33)	(12.35)	
Period 1 FITSBY use (minutes/day)	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

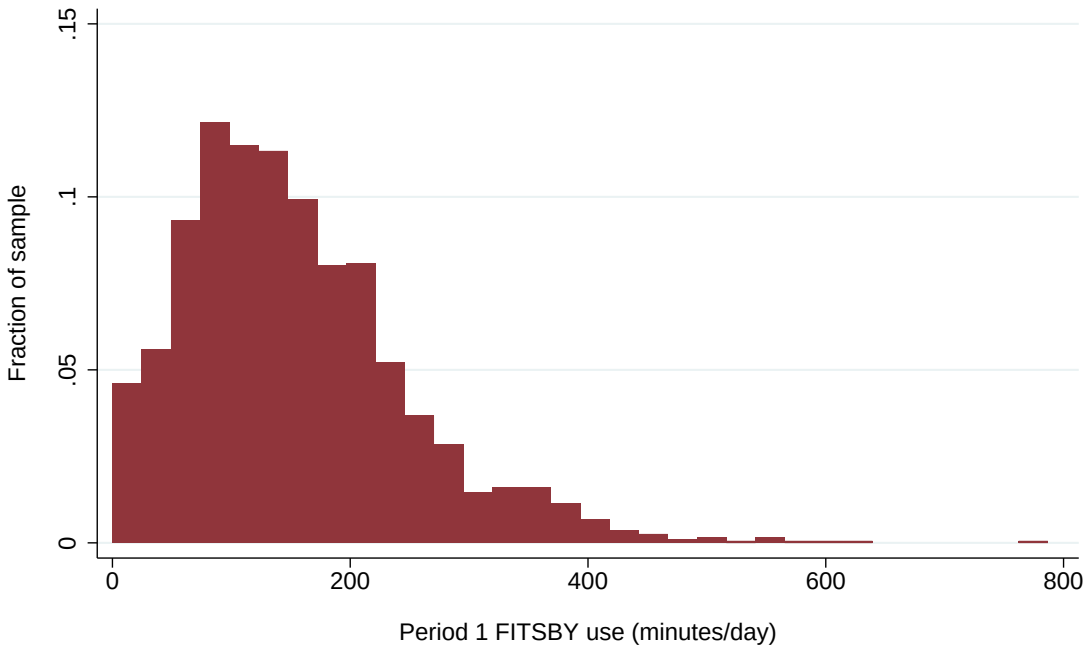
Notes: Panels (a) and (b) present tests of covariate balance for the Limit and Bonus treatment and control groups.

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.

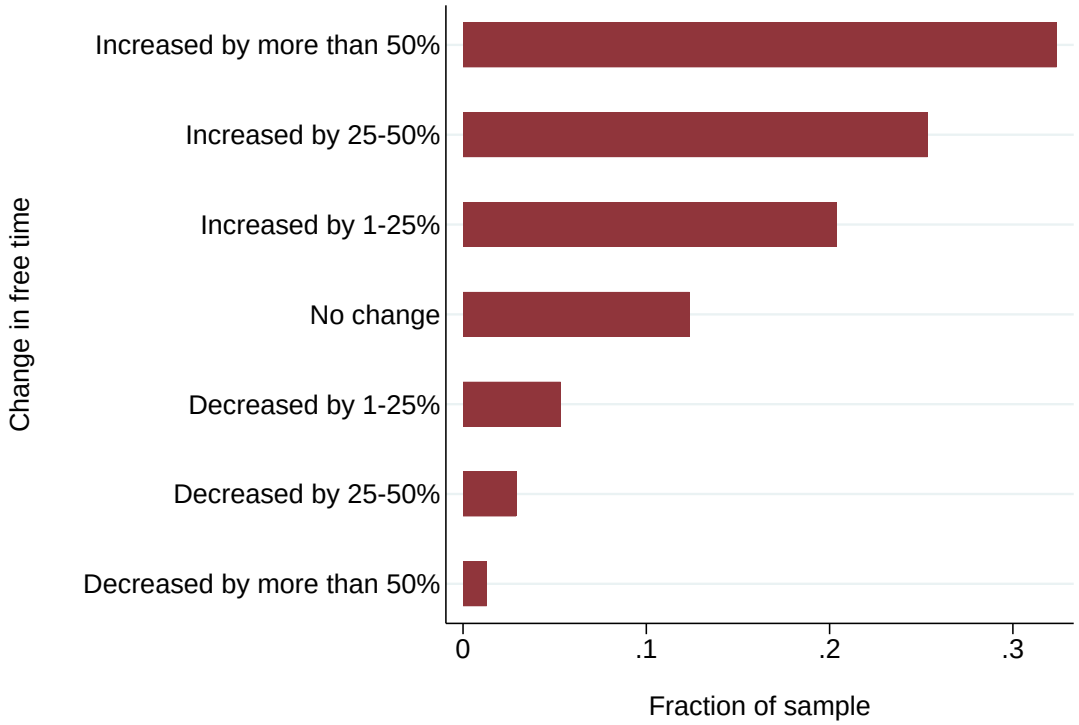
Table A3: **Descriptive Statistics for Survey Outcome Variables**

	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.

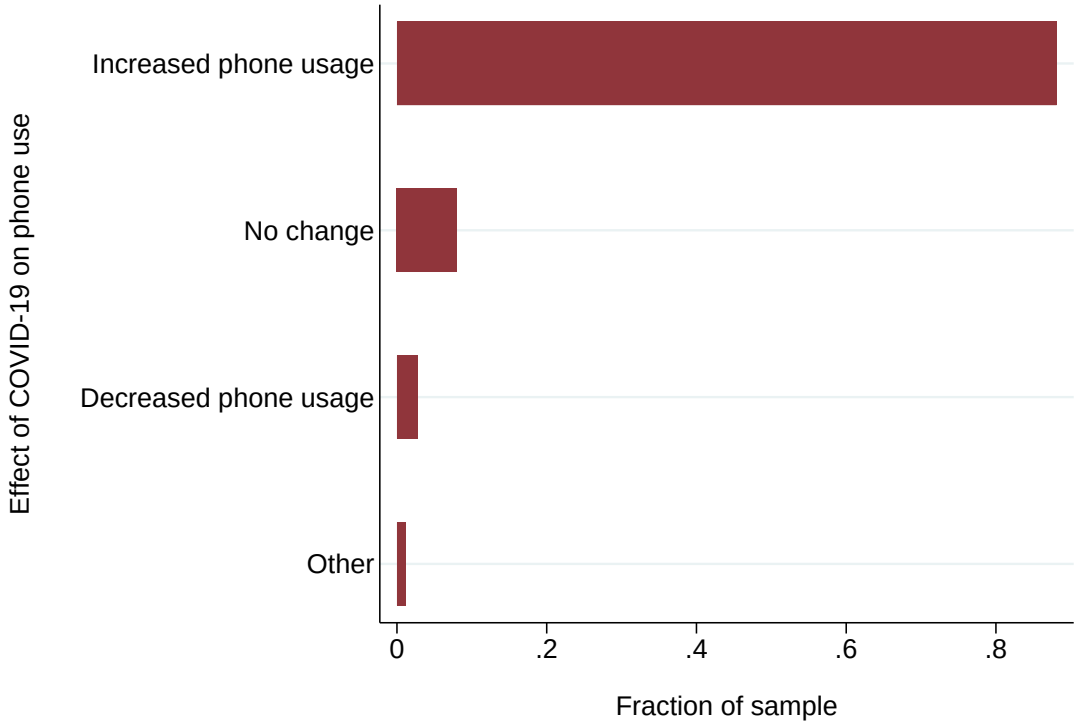
C Differences Between 2019 and the Study Period

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



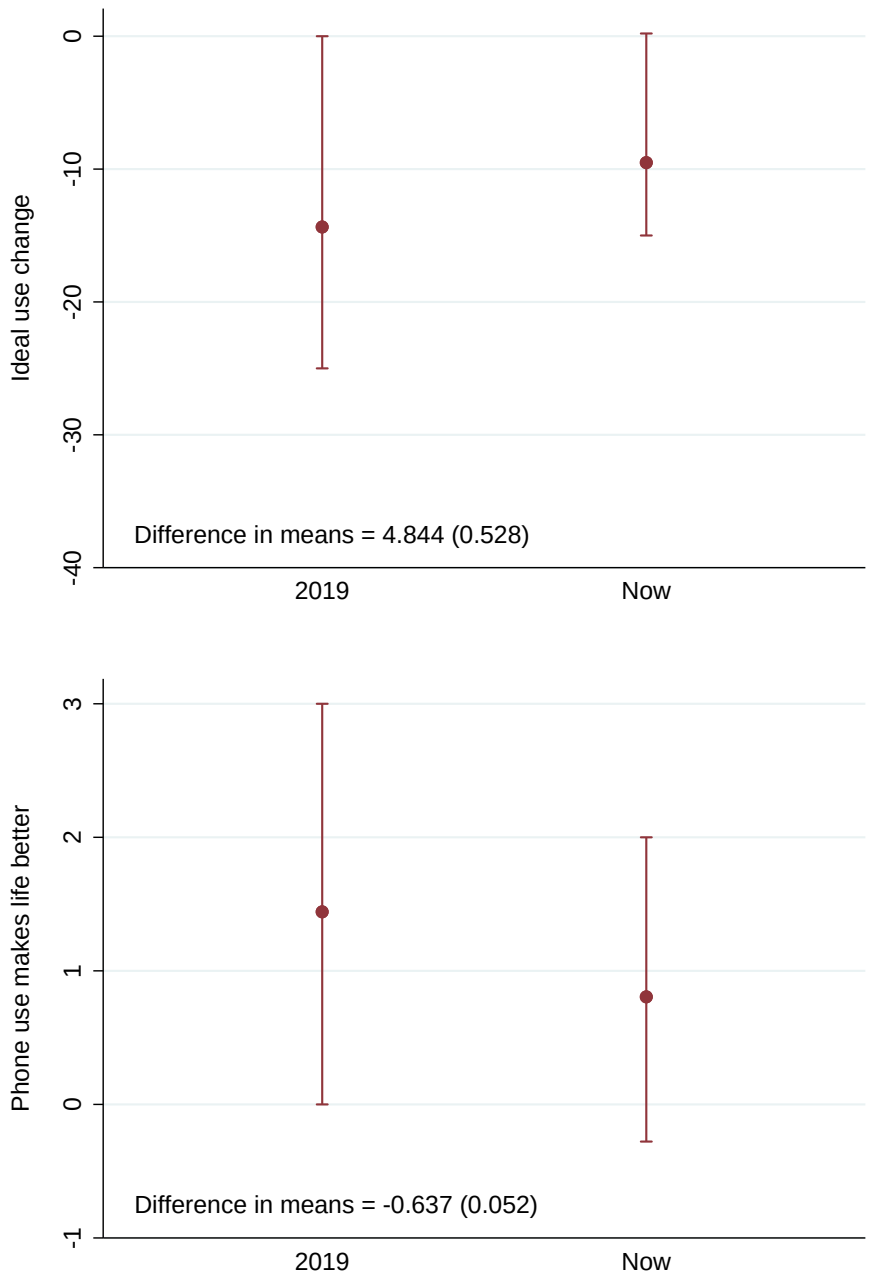
Notes: This figure presents the distribution of responses to the baseline survey question, "To what extent has the recent coronavirus outbreak changed how much free time you have?"

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.

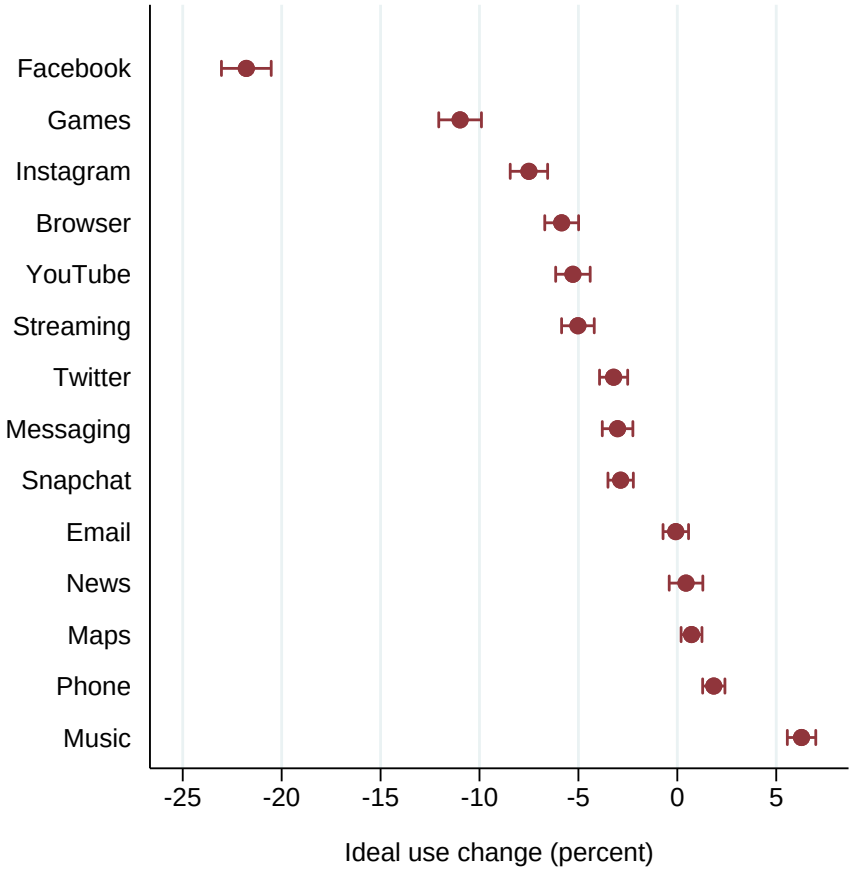
Figure A7: **Self-Control Problems in 2019 versus Now**



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]?"

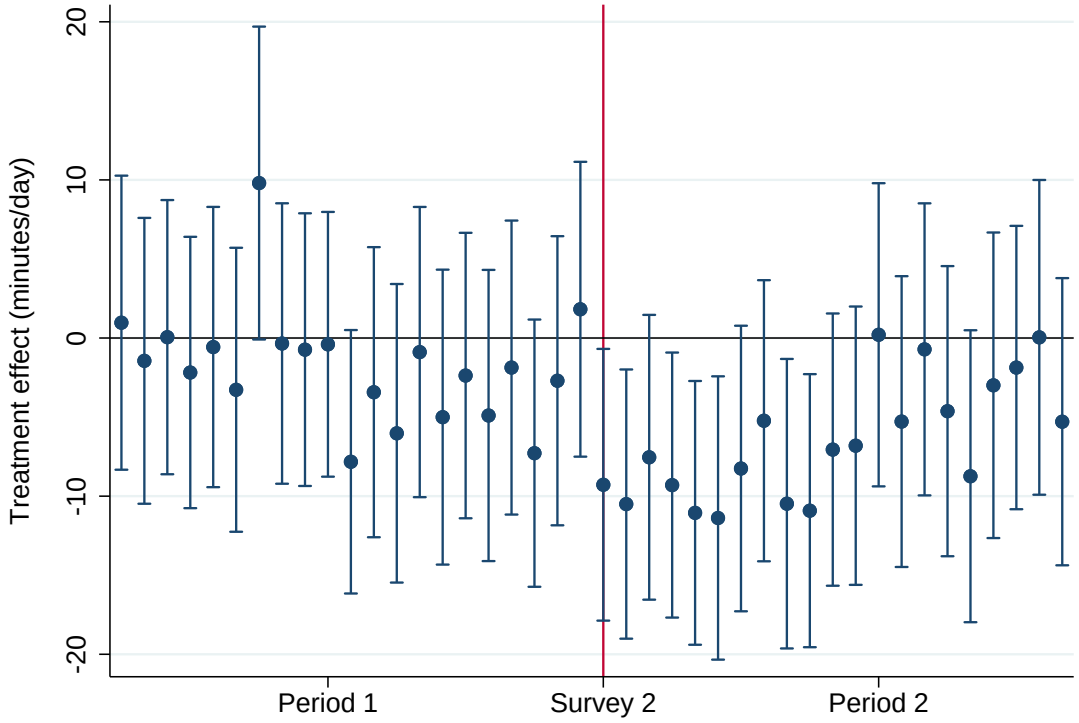
D Model-Free Results Appendix

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



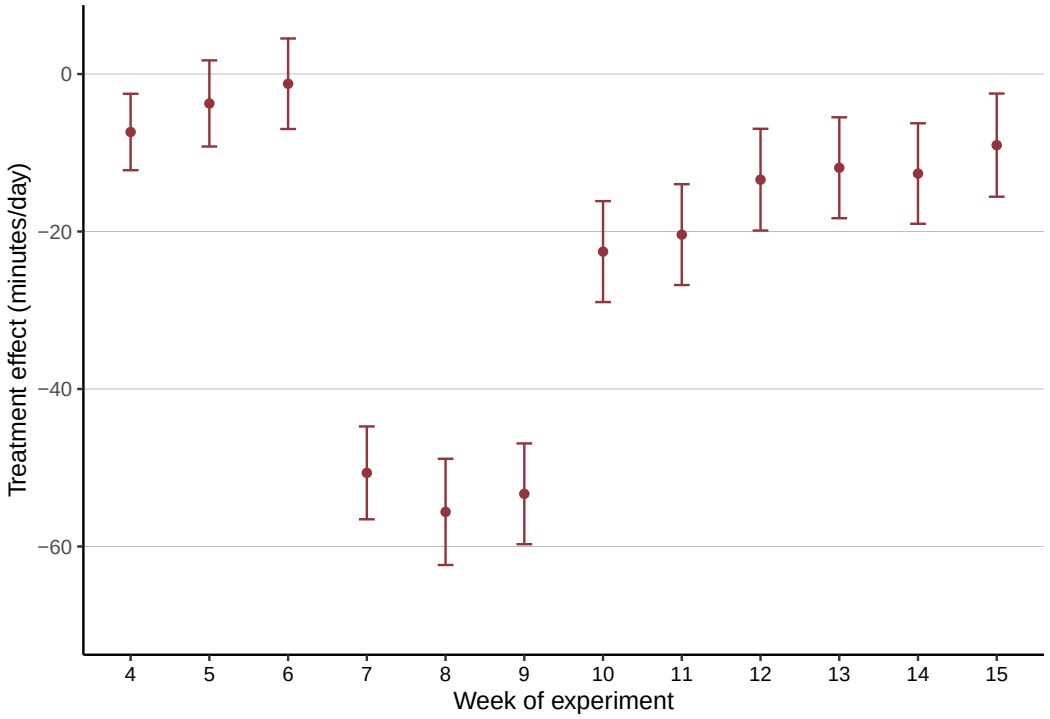
Notes: This figure presents mean *ideal use change* by app or app category at baseline. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" We code "I don't use this app at all" as 0, so these 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**



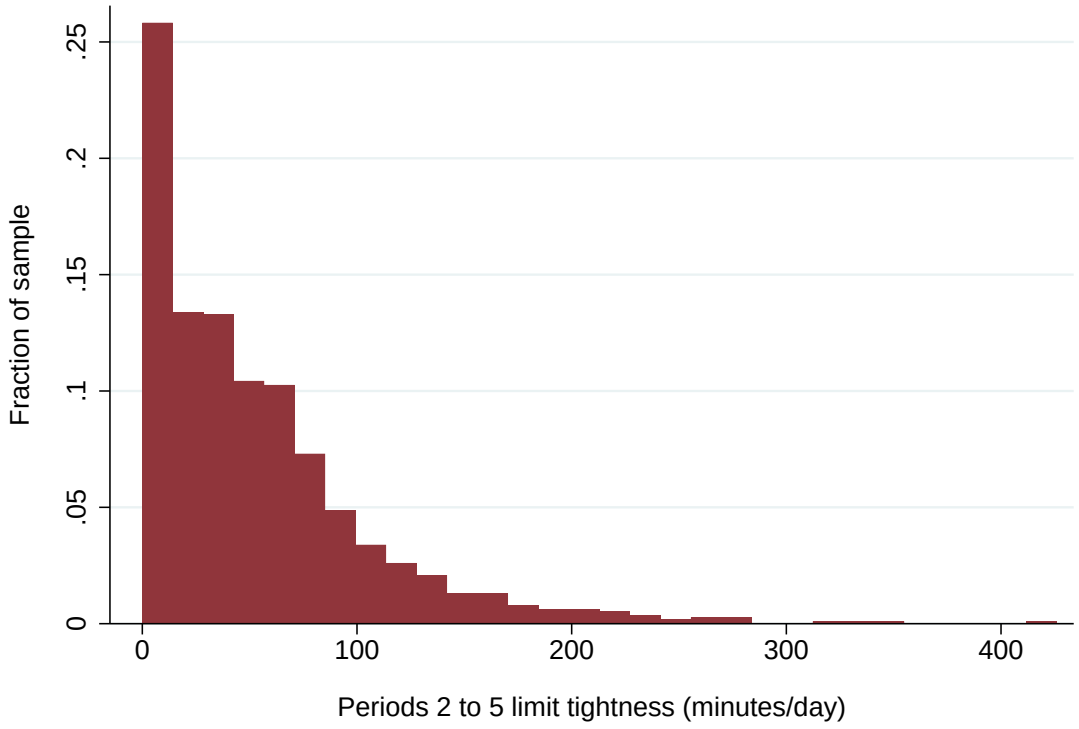
Notes: This figure presents differences in average FITSBY use between the Bonus and Bonus Control group for each day of periods 1 and 2. The vertical line indicates the day of survey 2, when the bonus was announced. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers,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

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



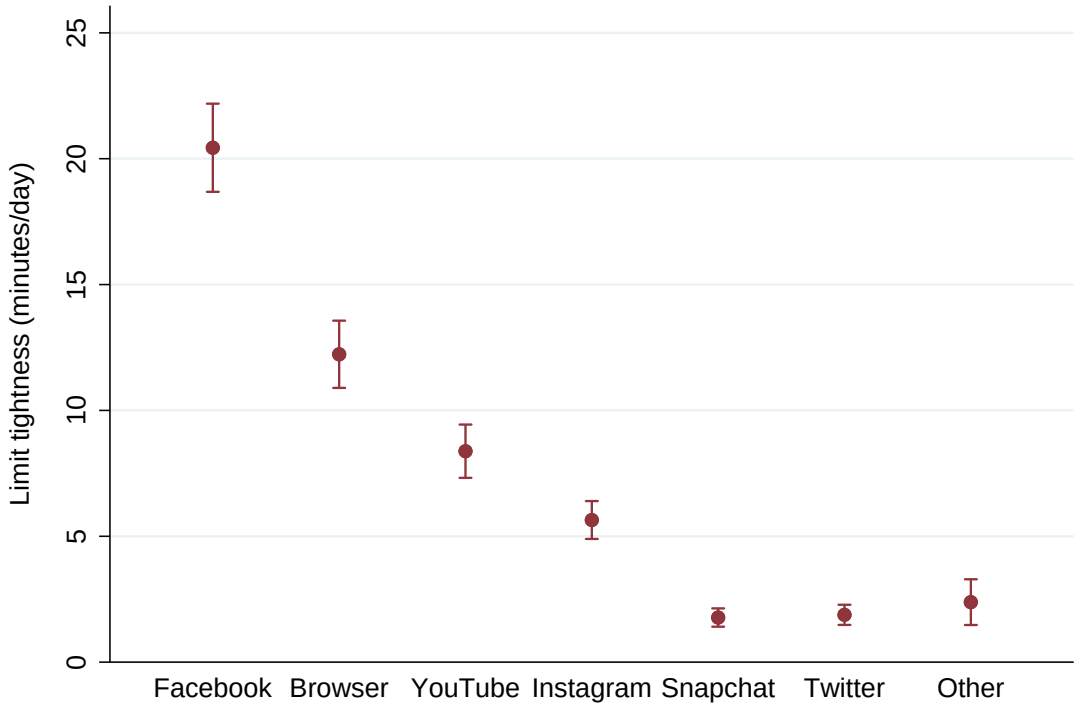
Notes: This figure presents effects of the bonus treatment on FITBSY use by week using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

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



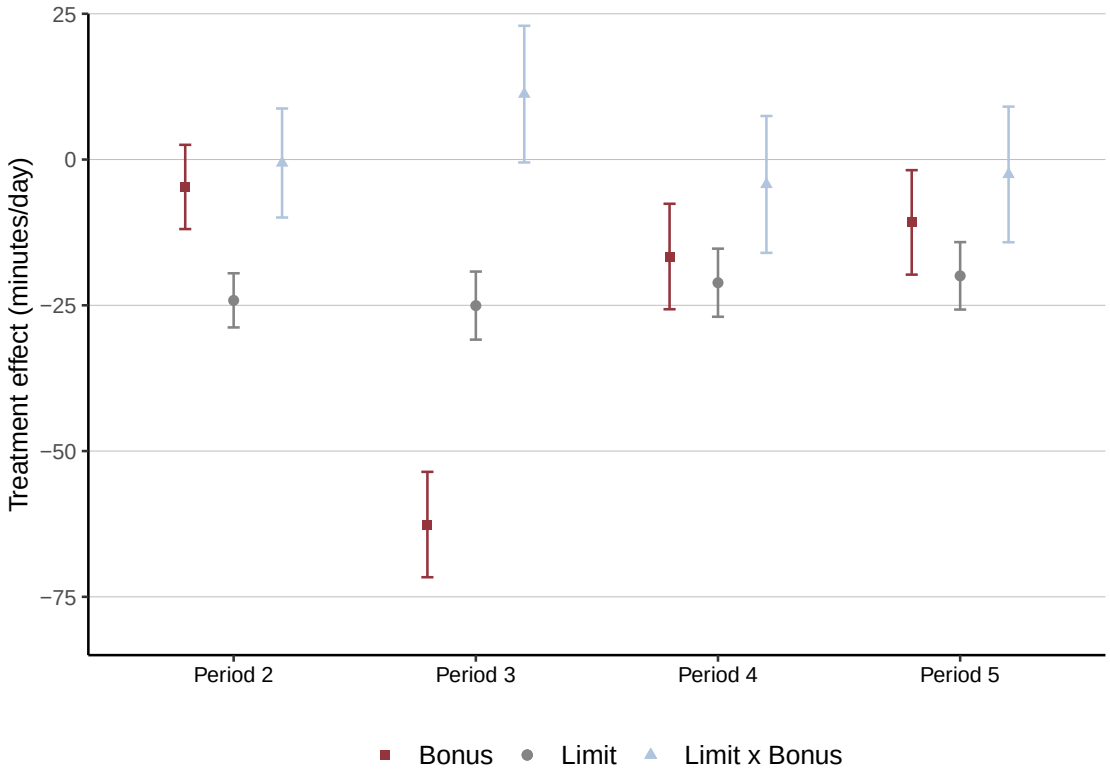
Notes: This figure presents mean *user-level limit tightness* over periods 2–5. *User-level limit tightness* is the amount by which a user's limits would have hypothetically reduced overall screen time if applied to their baseline use without snoozes; see equation (5).

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



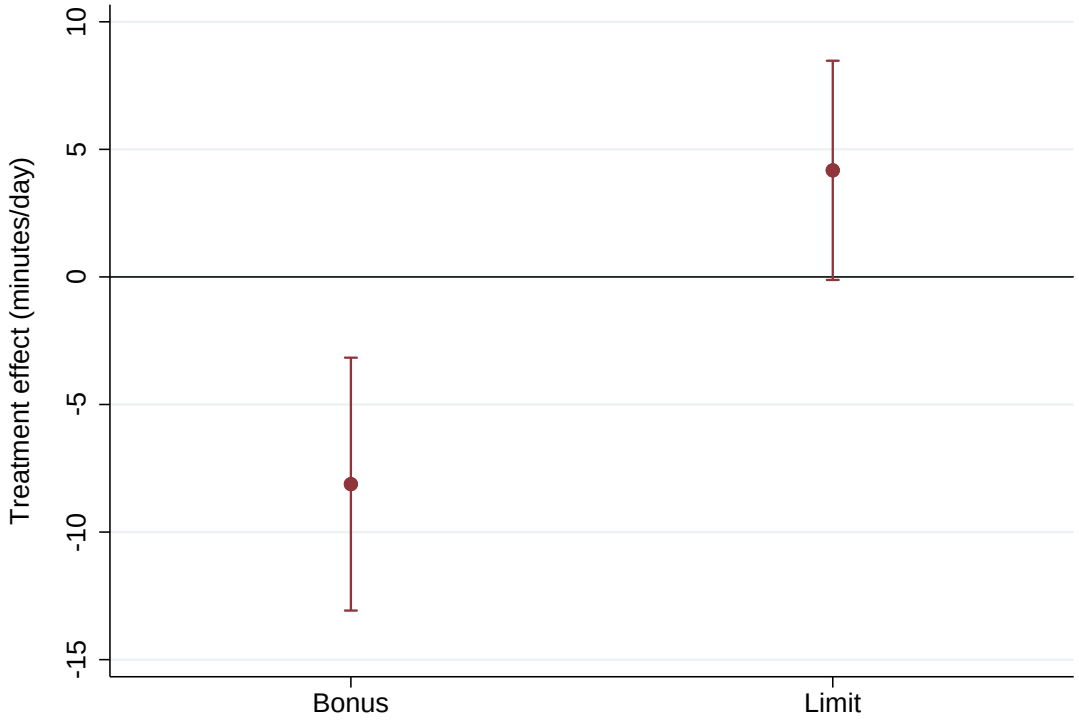
Notes: This figure presents average *limit tightness* by app over periods 2–5. *Limit tightness* is the amount by which a user's limits would have hypothetically reduced screen time if applied to their baseline use without snoozes; see equation (5). FITSBY apps are in order of decreasing period 1 use.

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**



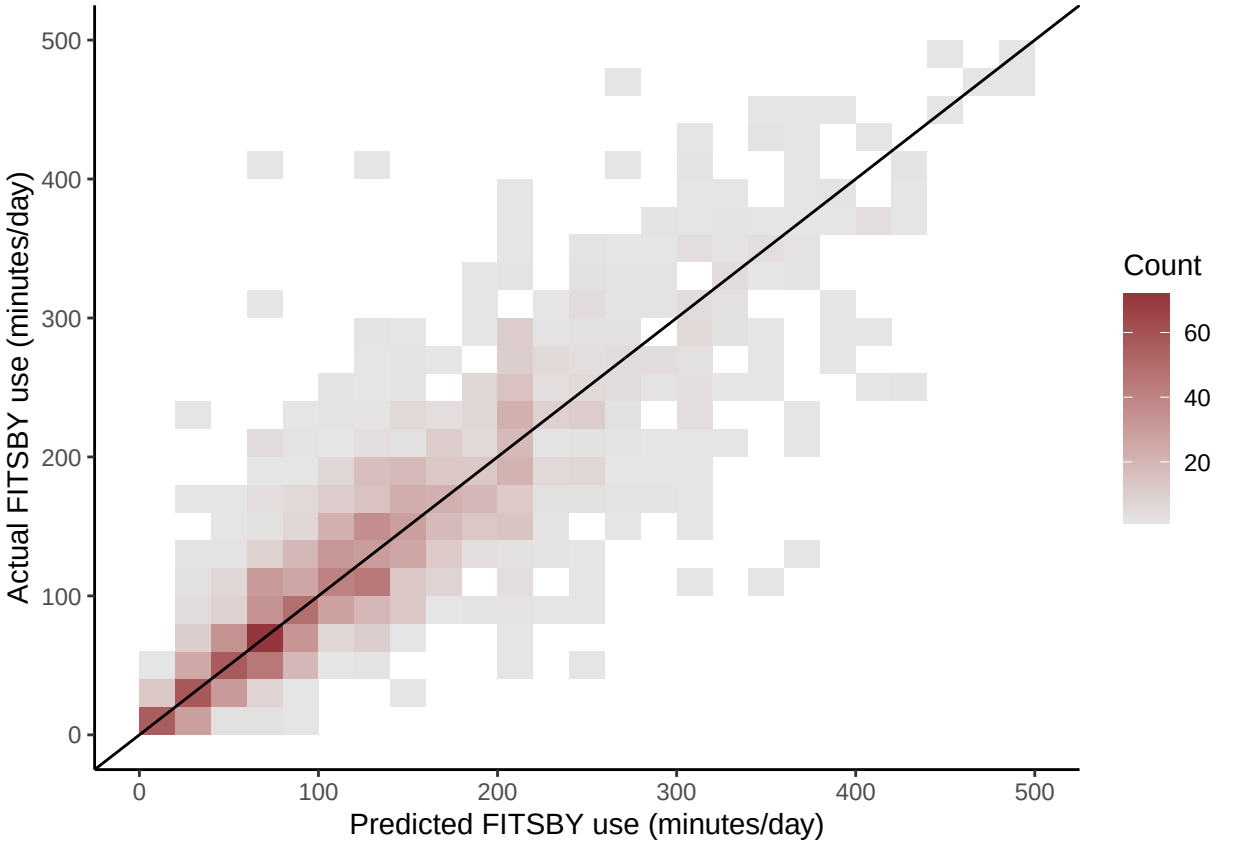
Notes: This figure presents the effects of bonus and limit treatments on self-reported change in FITSBY use on other devices relative to the three weeks before the study using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. Self-reported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

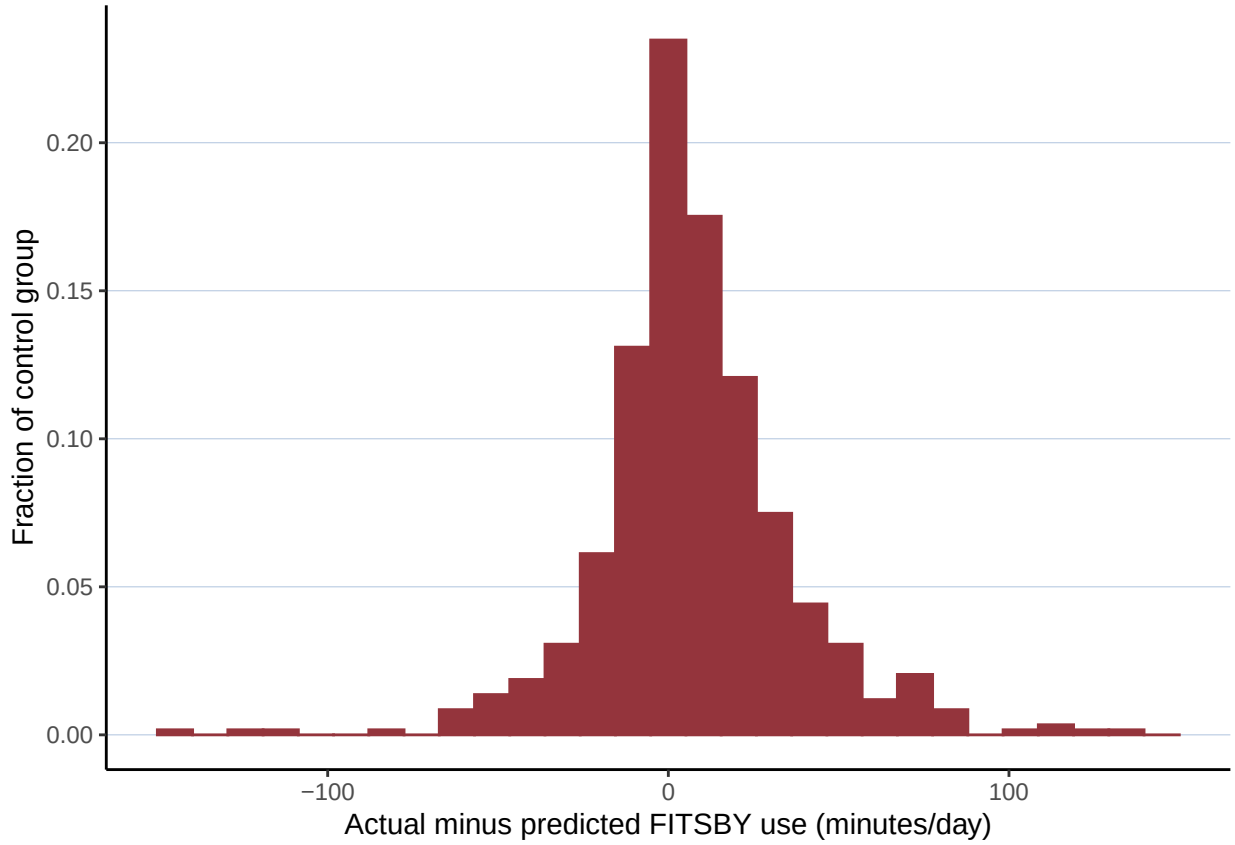
Multiple price lists are cognitively challenging, so we carry out several additional analyses to validate that these valuations are informative about people's preferences. First, participants' valuations of the bonus are correlated with the amount of money they could expect to earn; see Appendix Figure 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**



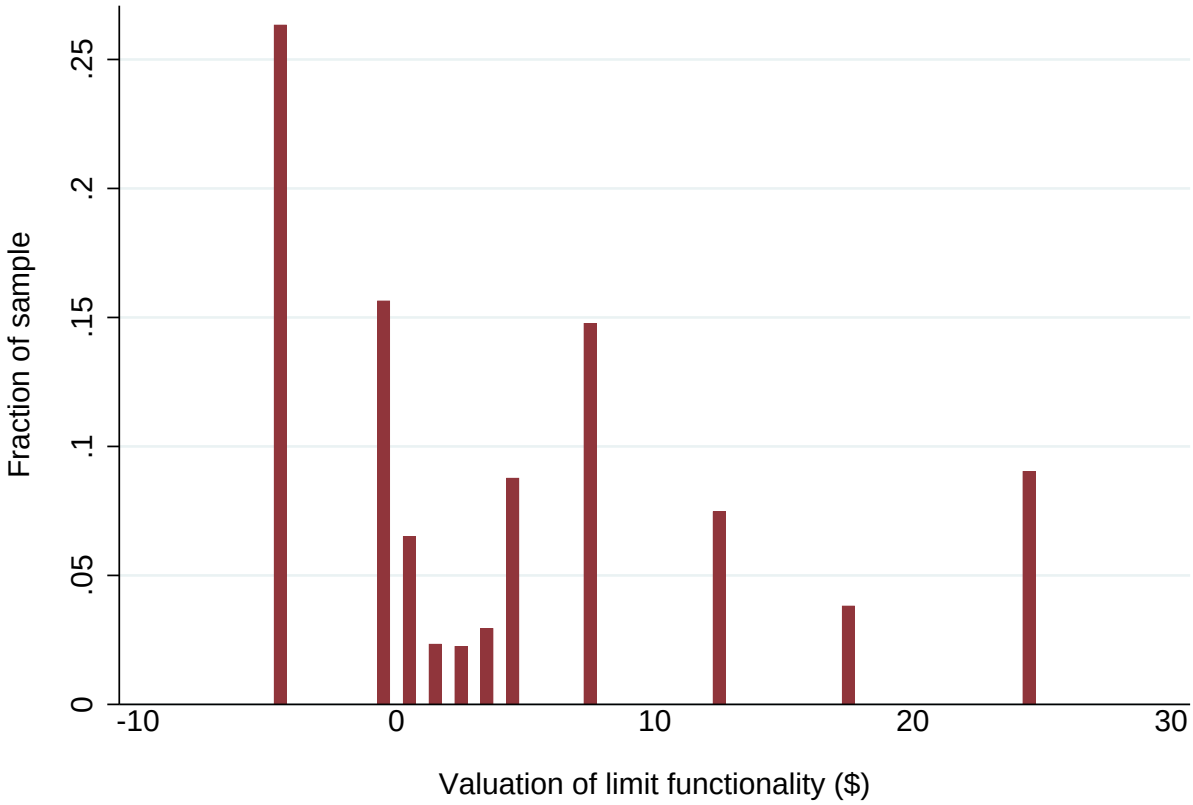
Notes: This figure presents the distribution of the difference between actual and predicted FITSBY use across periods 2–4 in the Control group, using predictions from the survey just before each period. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

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

	(1) Predicted use	(2) Actual use	(3) Predicted - actual use	(4) 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)

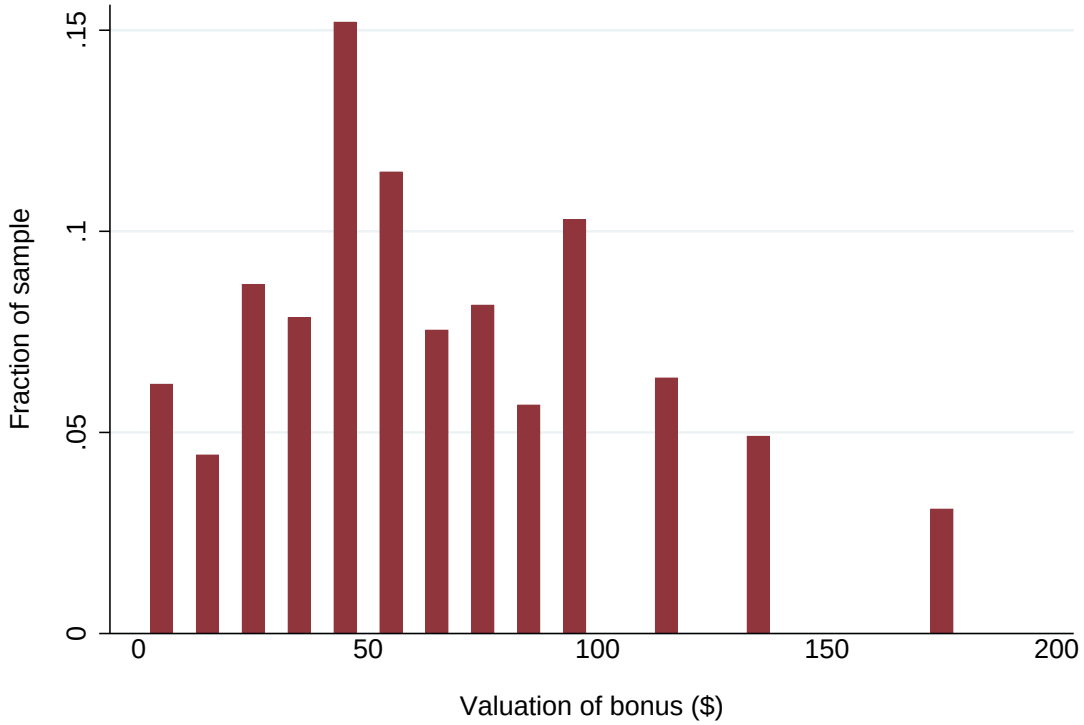
Notes: This table presents the effects of being offered the higher Prediction Reward ($5 instead of $1 for predicting within 15 minutes of actual screen time) on predicted and actual FITSBY use in minutes per day. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and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

Figure A17: **Valuation of Limit Functionality**



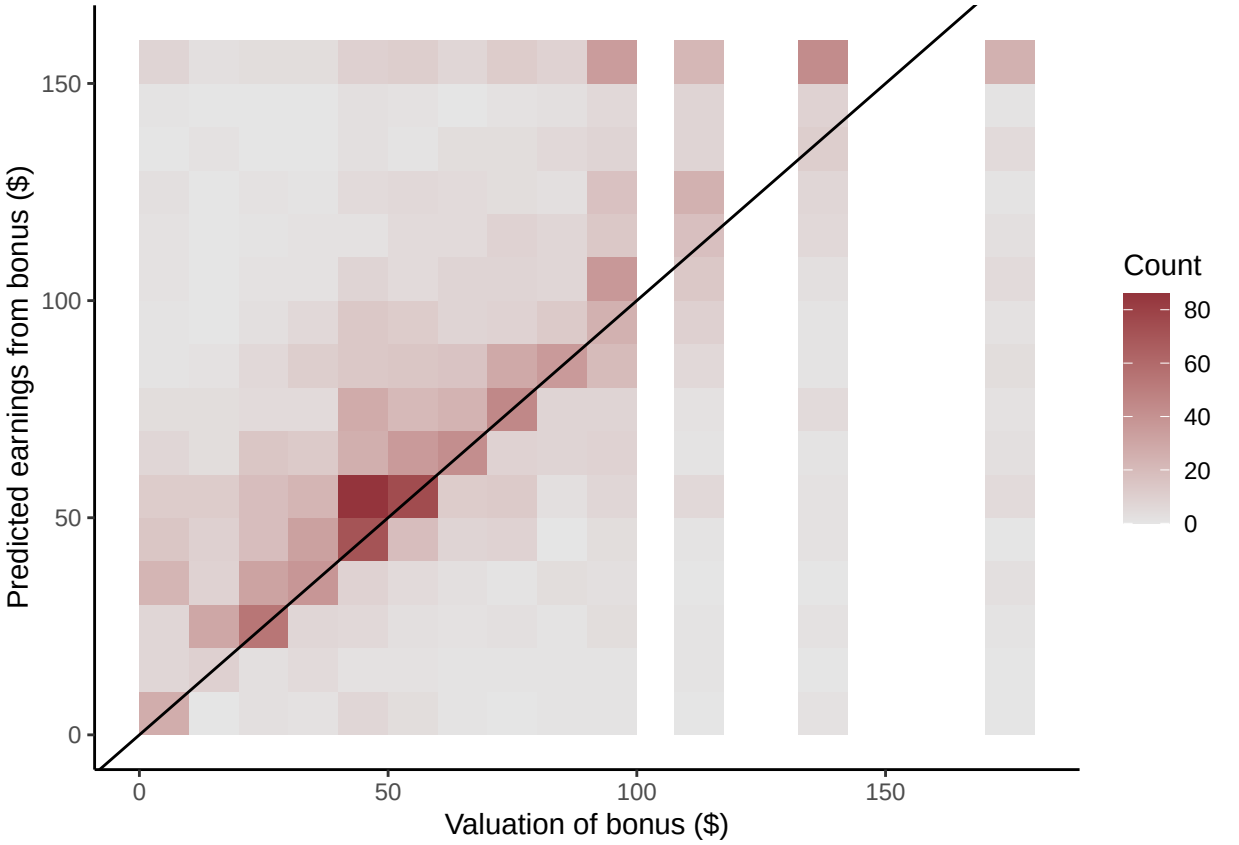
Notes: This figure presents the distribution of valuations of access to the limit functionality for the next three weeks, as elicited in a multiple price list on survey 3. Valuations above $20 are plotted at $25, and valuations below $-1 are plotted at $-5.

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**



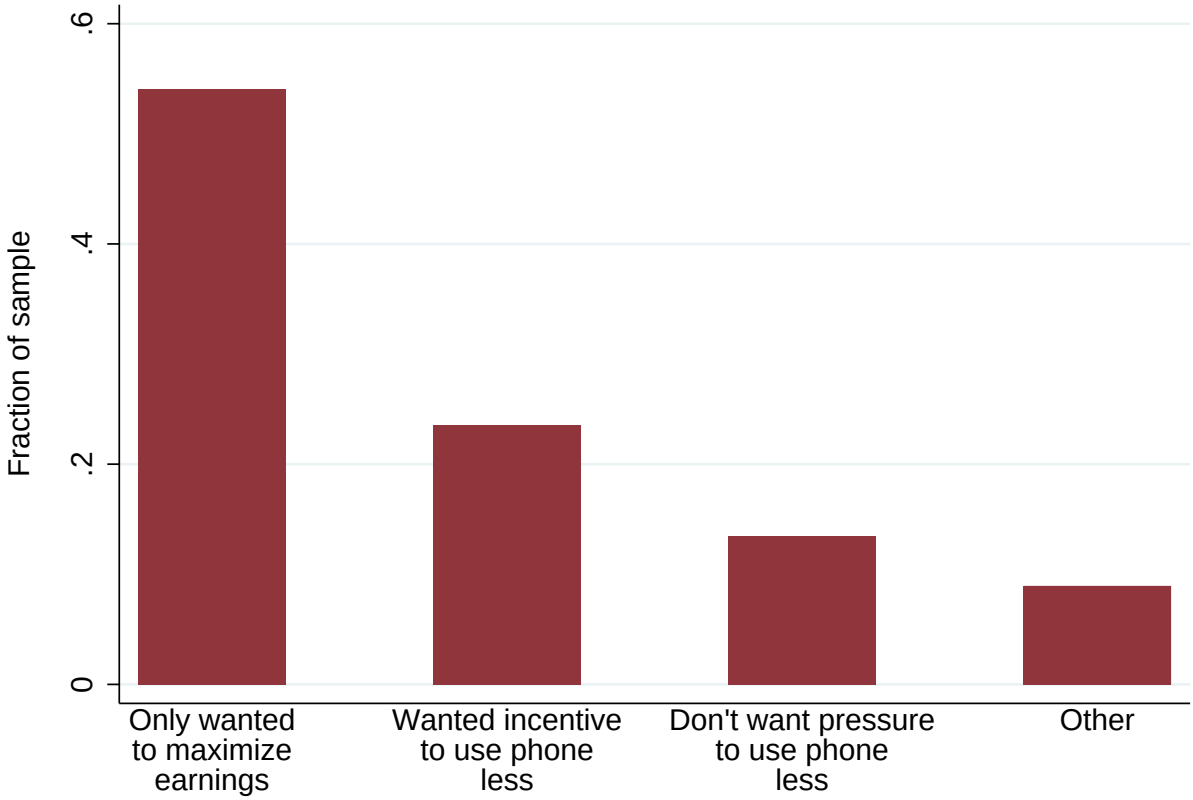
Notes: This figure presents the number of participants in each cell of predicted earnings from the Screen Time Bonus (given the participant's Bonus Benchmark and predicted FITSBY use) and valuation of the bonus, as elicited on survey 2.

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

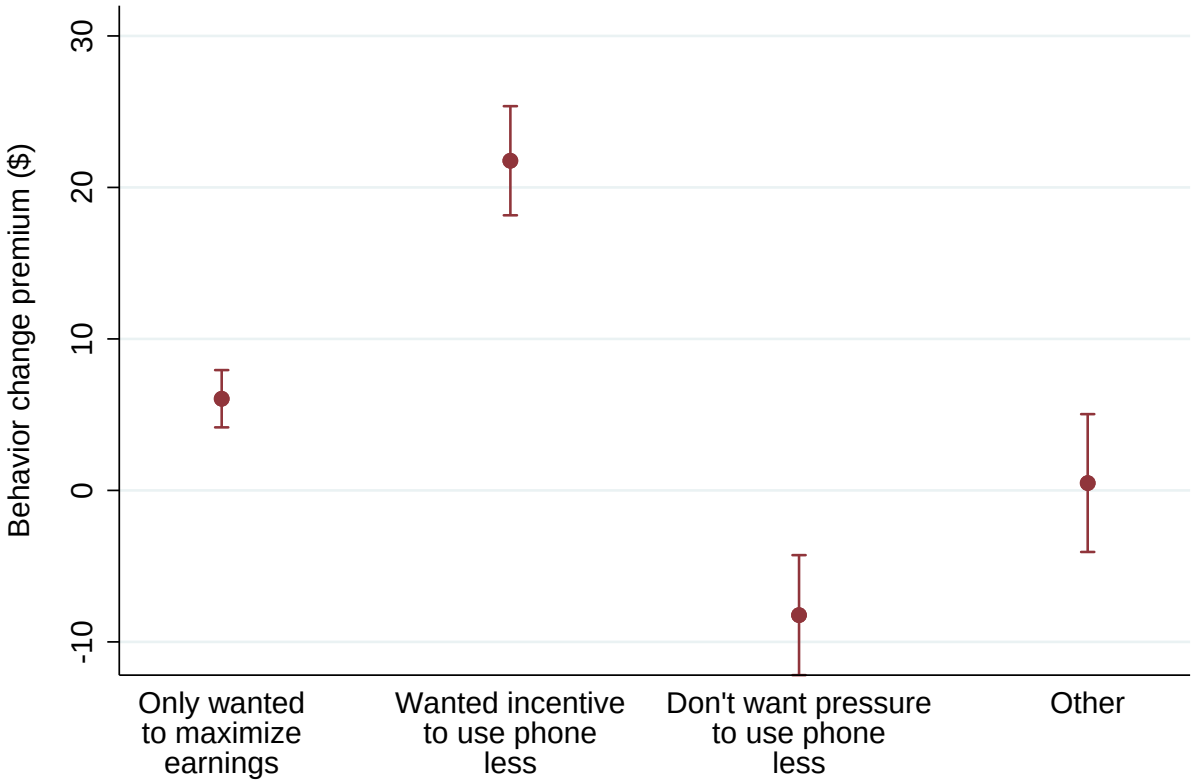
Note: The behavior change premium is the difference between the valuation of the Screen Time Bonus and the modeled valuation if the consumer believed herself to be time consistent. *Interest in limits*, *ideal use change*, *addiction scale*, *SMS addiction scale*, and *phone makes life better* are from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

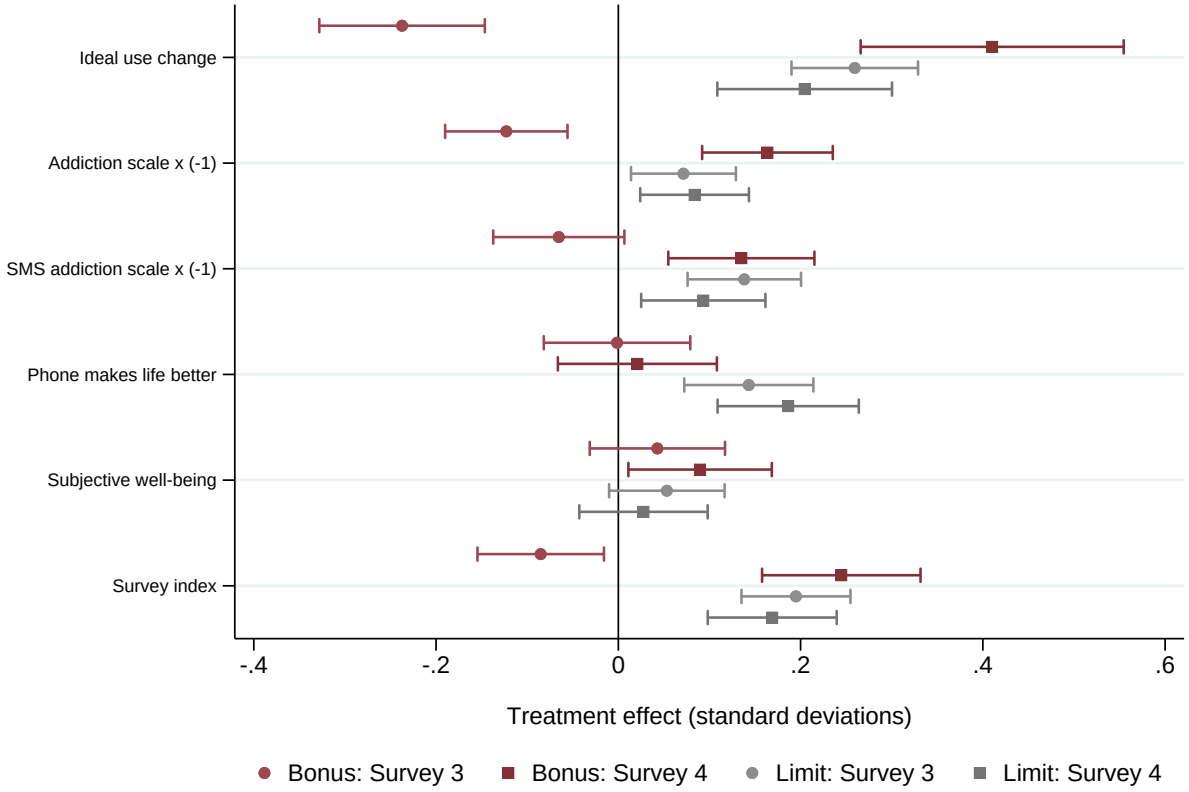
Figure A21: **Behavior Change Premium by Reported Reasoning**



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**

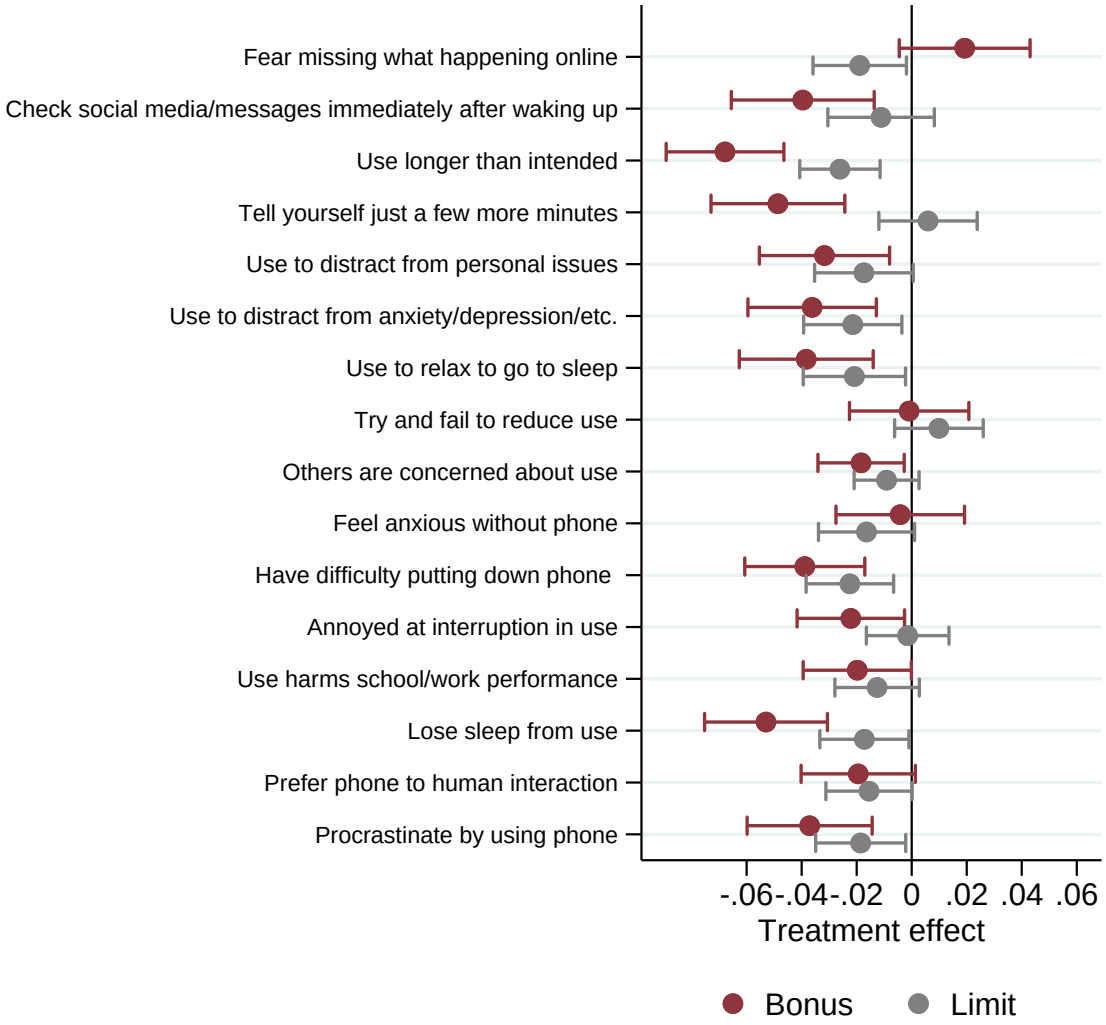
	(1)	(2)	(3)	(4)	(5)	(6)
	Treatment effect (original units)	Standard error (original units)	Treatment effect (SD units)	Standard error (SD units)	P-value	Sharpened FDR-adjusted q-value
Ideal use change	9.0	1.6	0.41	0.074	0.000	0.000
Addiction scale x (-1)	0.44	0.10	0.16	0.037	0.000	0.000
SMS addiction 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

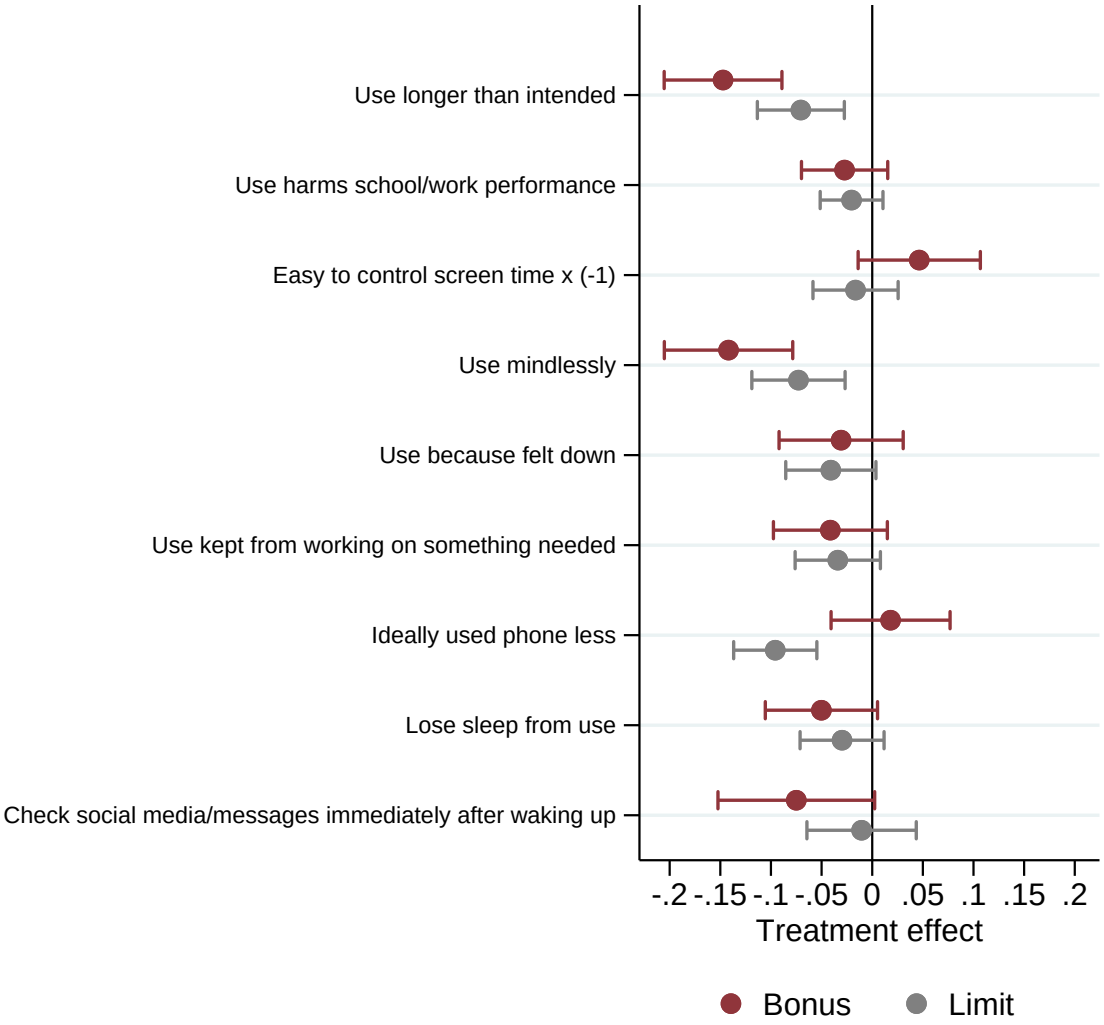
Notes: This table presents effects of the bonus and limit treatments on survey outcome variables using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 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**



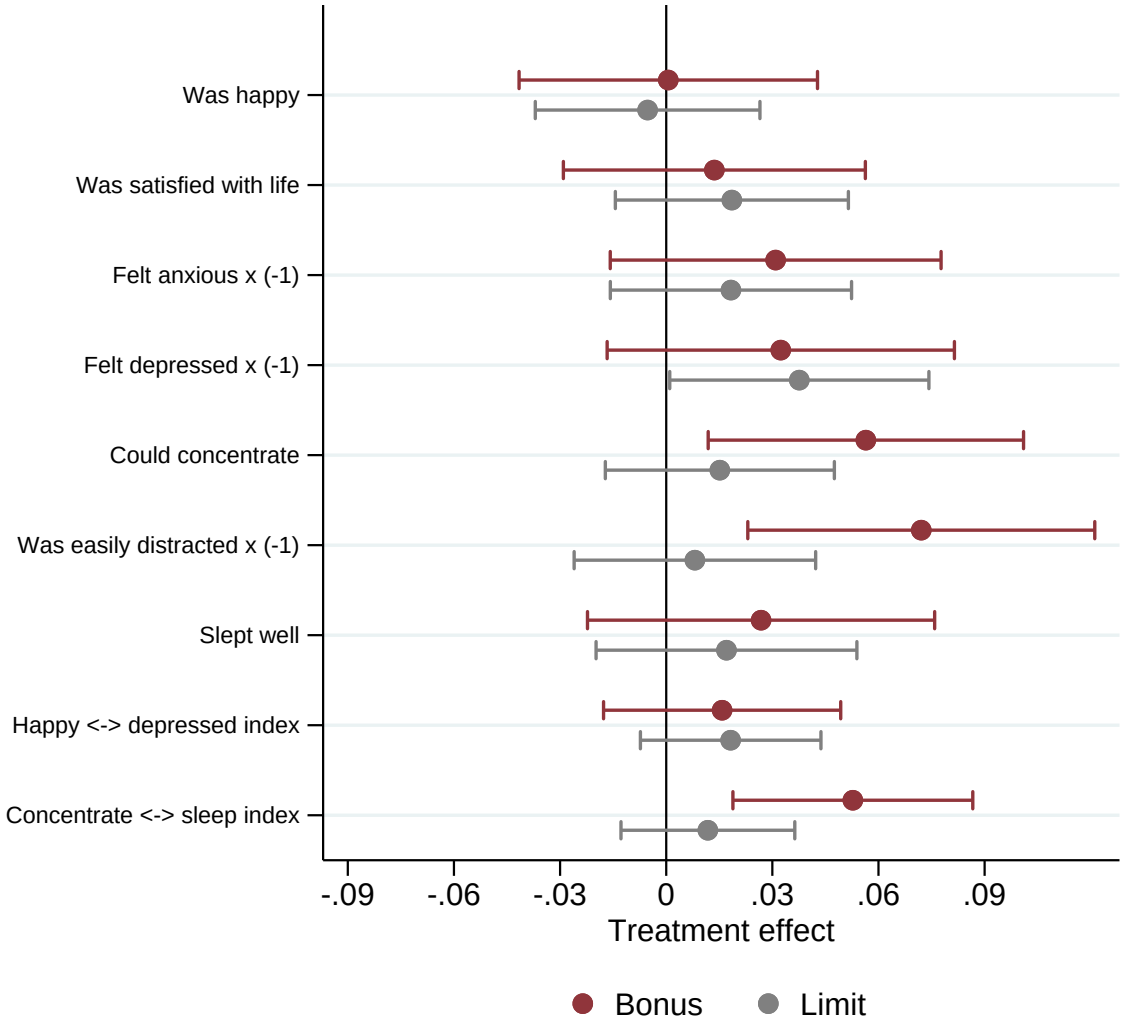
Notes: This figure presents the effects of the bonus and limit treatments on individual items in the *addiction scale* variable using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. The direction of the effects in this figure are opposite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

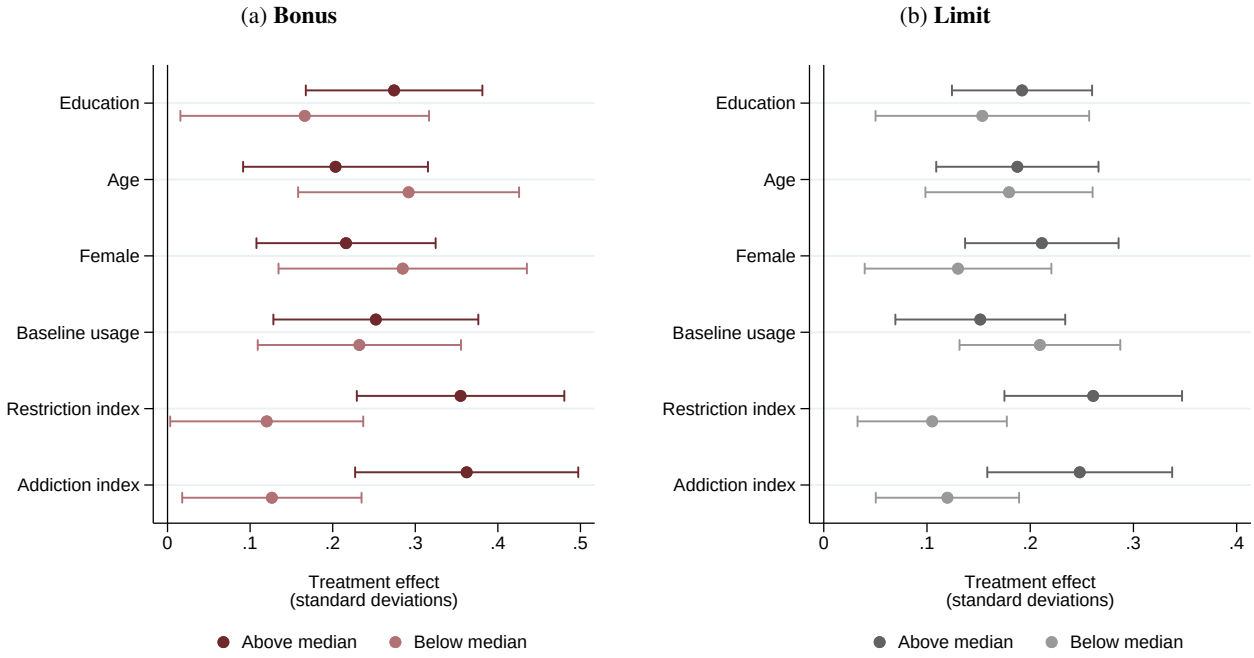
Figure A25: **Effects on Subjective Well-Being Responses**



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.

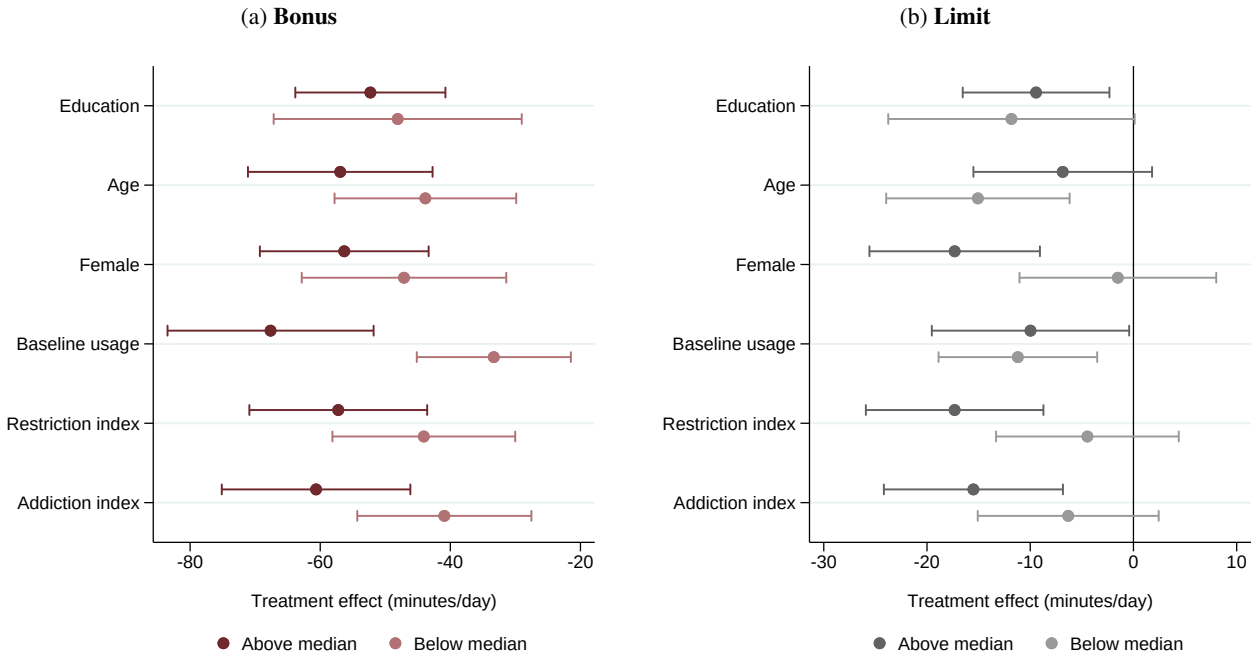
D.3 Heterogeneous Treatment Effects

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:

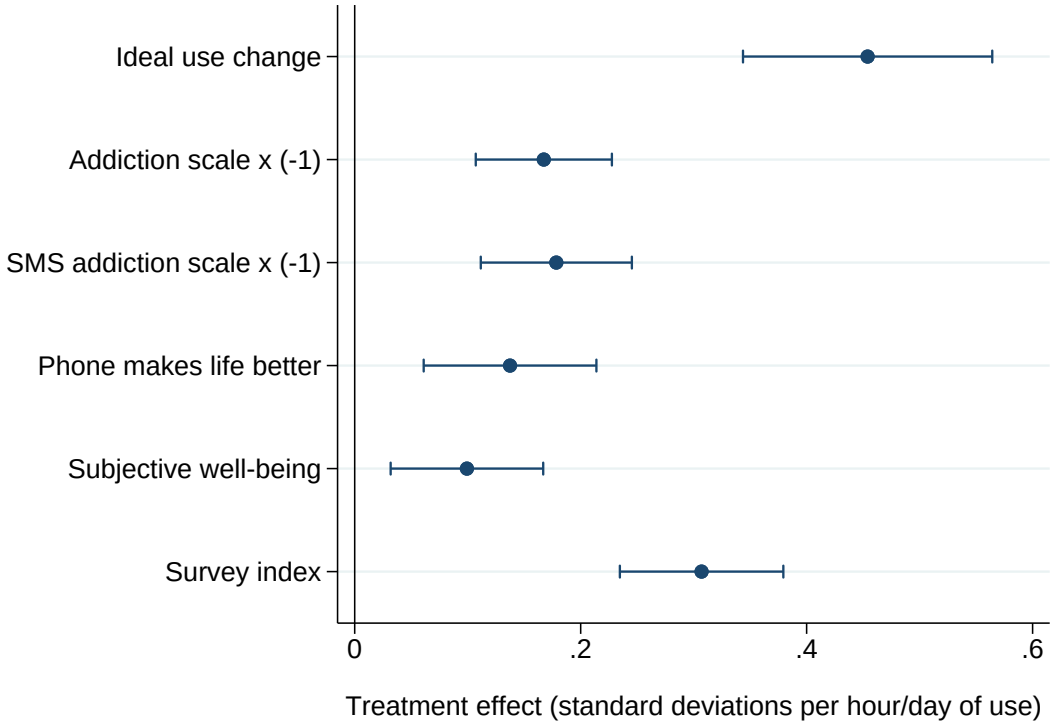
$$Y_{it} = \tau x_{i,t-1} + \beta_t \boldsymbol{X}_{i1} + \nu_{it} + \varepsilon_i, \tag{17}$$

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

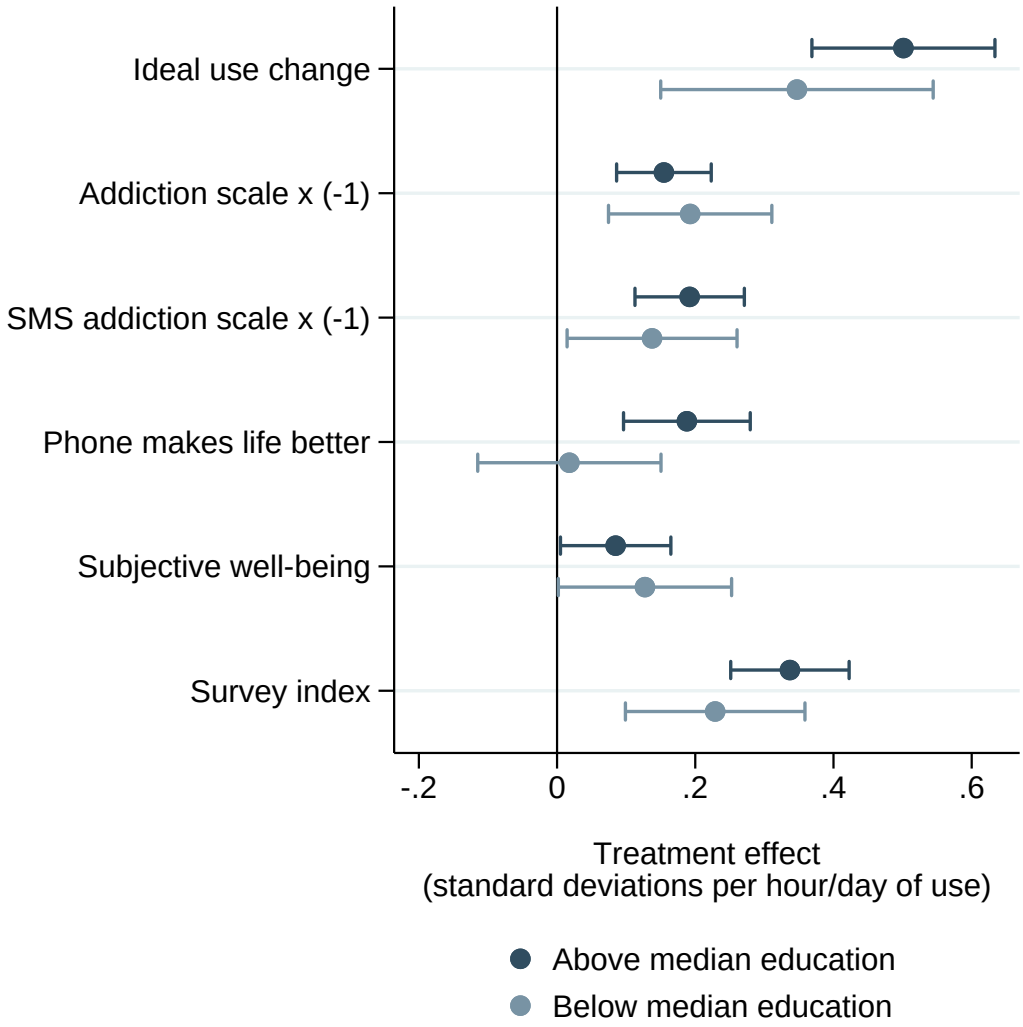
moderators. The results are qualitatively similar to Figures 8 and A26, except that the estimates are slightly more precise, as would be expected from combining effects of two interventions. Note that since the average effects of both interventions are about the same for people with low versus high baseline 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**



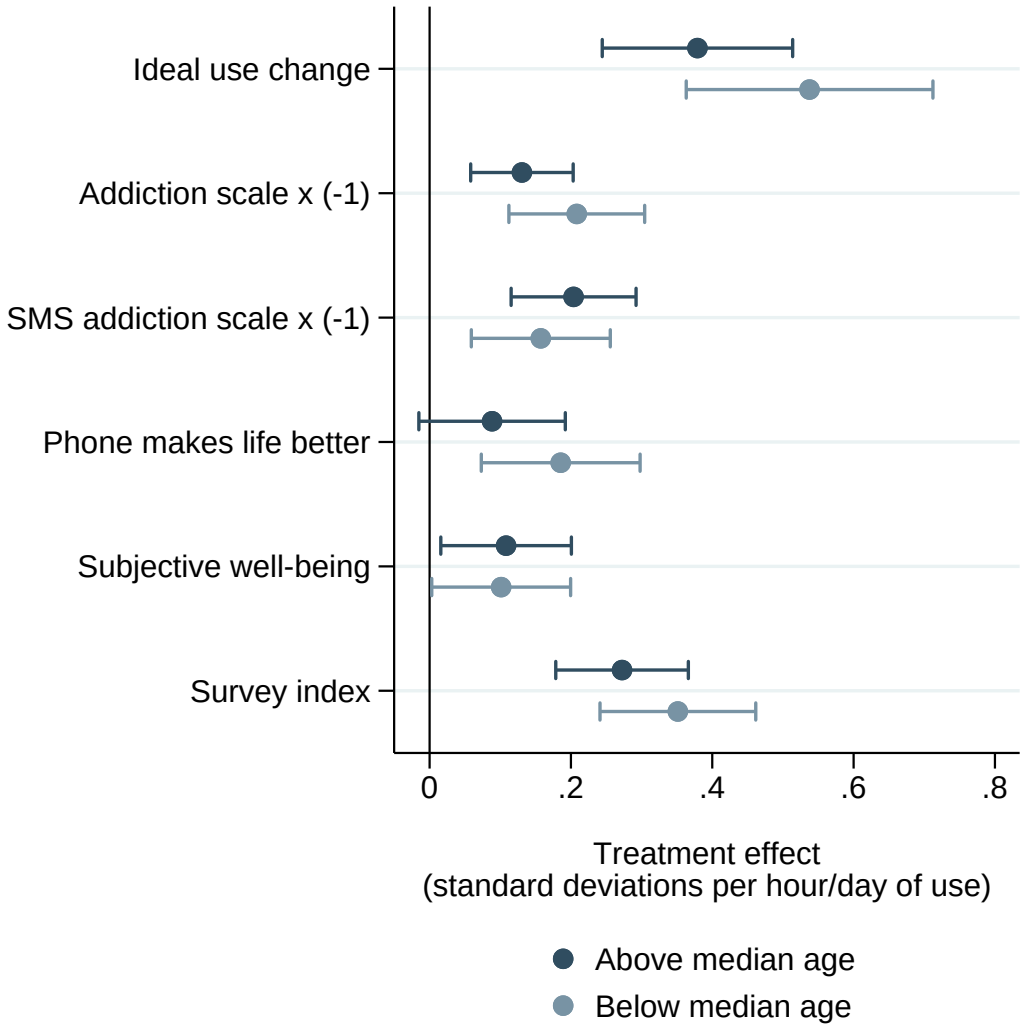
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17). We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, 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**



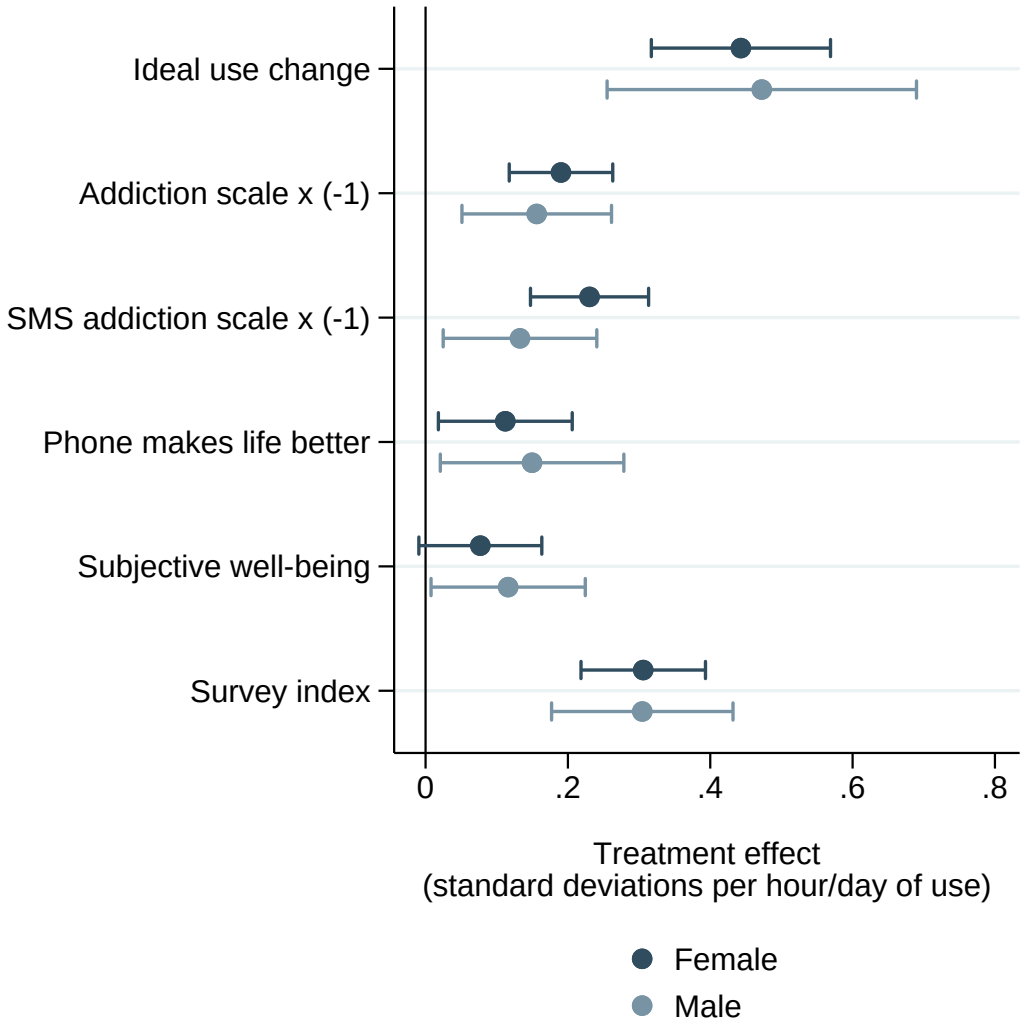
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median education. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, 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**



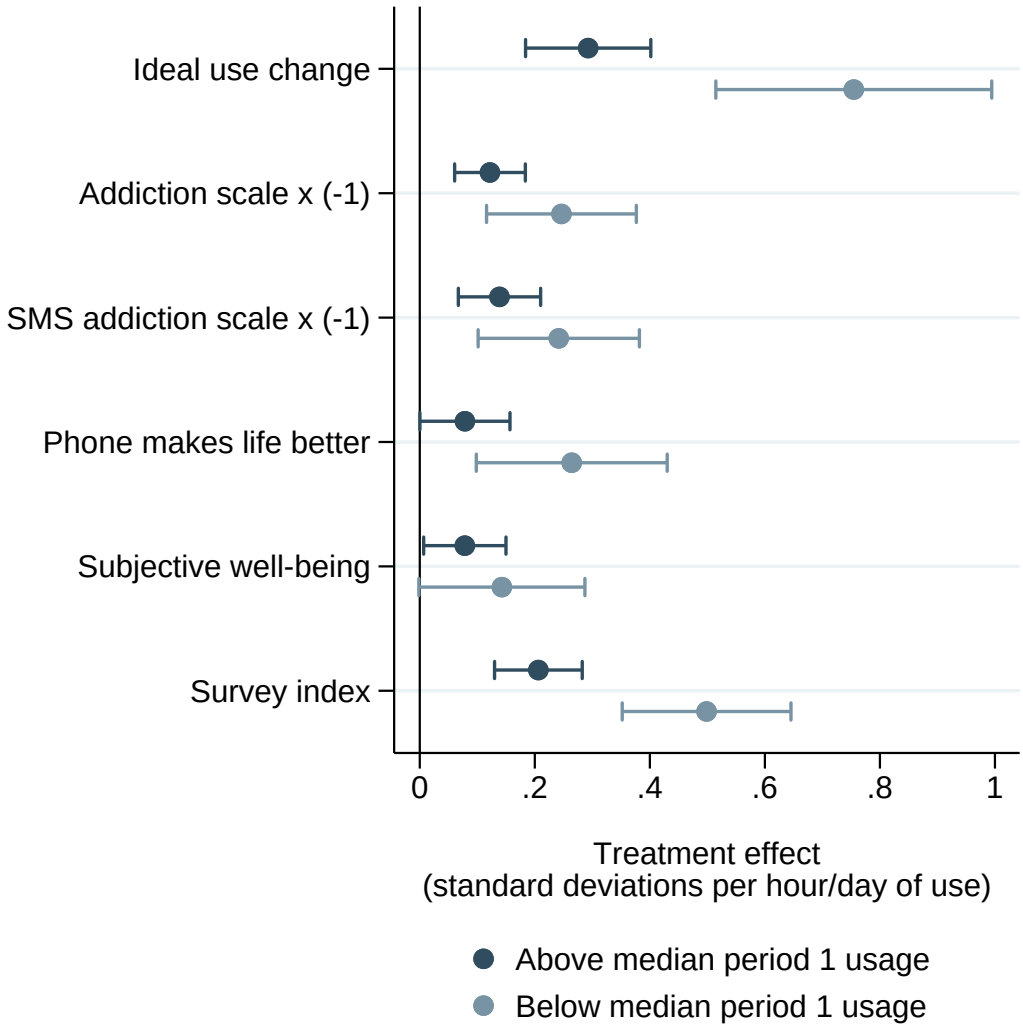
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median age. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, 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**



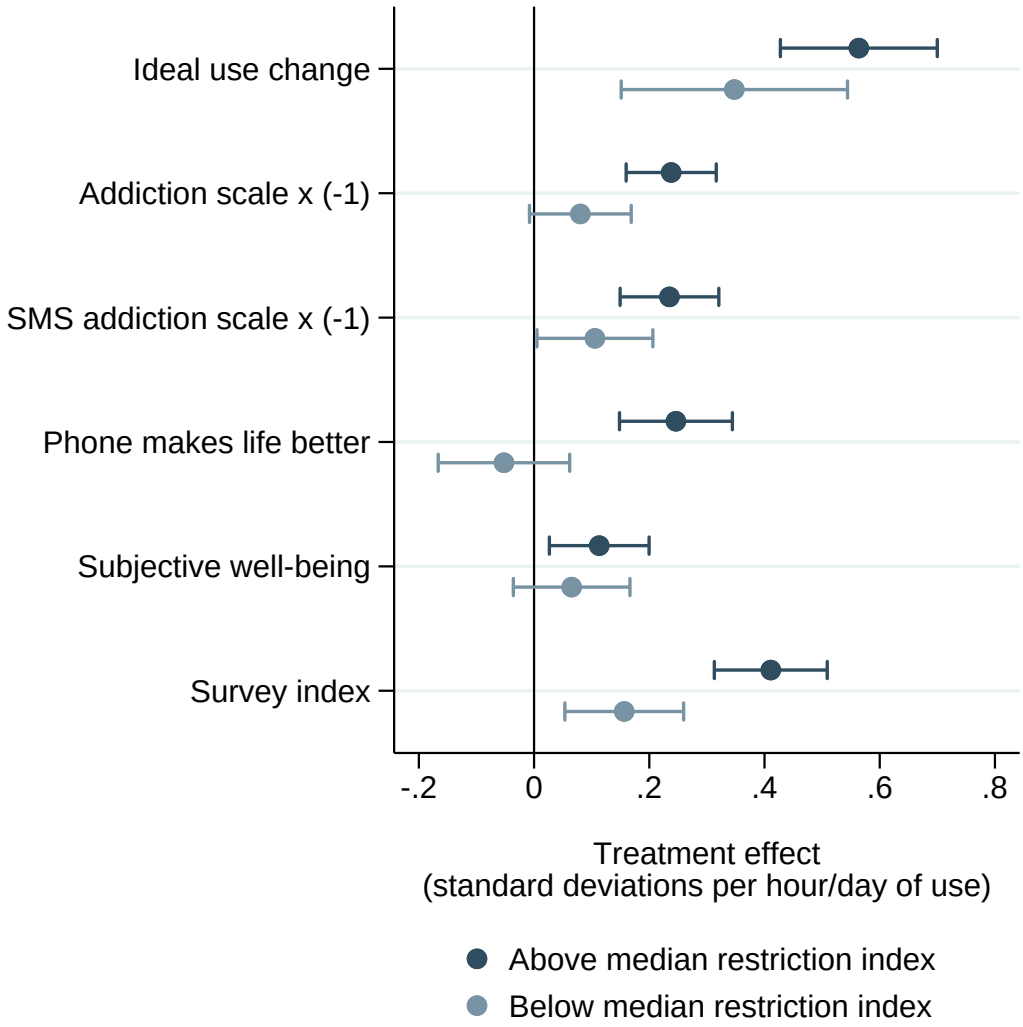
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for men versus women. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, 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**



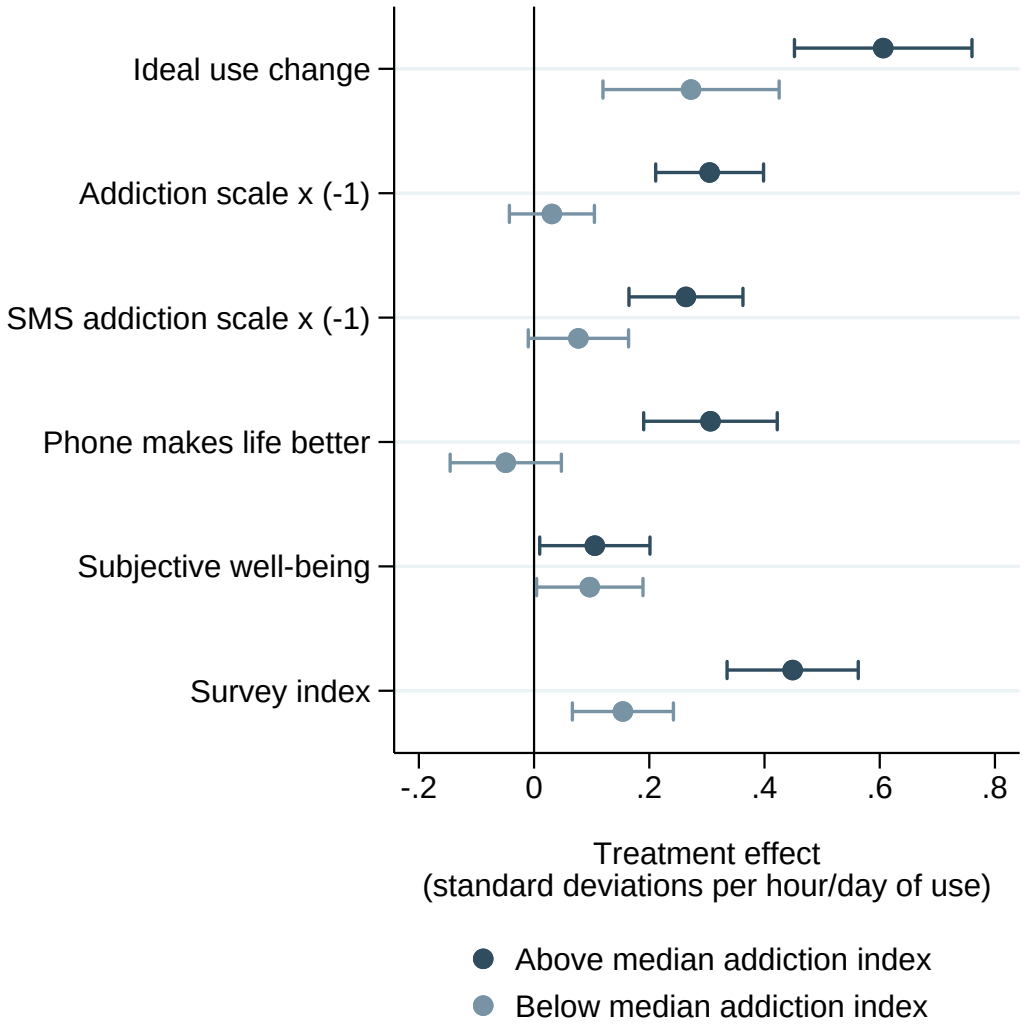
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median baseline FITSBY use. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen 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 $u_t \coloneqq u_t(x_t^*; s_t, p_t)$ as current utility, define $\tilde{x}_r \coloneqq \tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, \boldsymbol{p}_r)$ and $\tilde{u}_r \coloneqq u_r(\tilde{x}_r; \tilde{s}_r, p_r)$ as predicted consumption and utility for future periods $r > t$, and define $\tilde{\lambda}_r \coloneqq \frac{\partial \tilde{x}_r}{\partial \tilde{s}_r}$ 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 $t < T$,*

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

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 $\tilde{\gamma}\tilde{\lambda}_{t+1}$ 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 $\text{ceil}(\cdot)$ 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, $\bar{F}^B - p_3^B \bar{\tilde{x}}_3^{B+BC}$. 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 $\tilde{\gamma}\cdot\left(x_3^*(\tilde{\gamma}) - x_3^*(0)\right)\cdot\frac{1}{2}$. 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

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

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 $p_3^B \tilde{\bar{x}}_3 \approx \$2.50 \times \frac{1}{2}(2.5+1.6) \approx \$5.09$ 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 $\hat{\tilde{\gamma}} = \frac{\bar{v}^L}{-\tilde{\tau}_3^L/2} \approx \frac{0.210}{(-(-22.3)/60)/2} \approx 1.13$ \$/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 $\hat{\tau}^L_{it} = \hat{\tau}^L_t + \hat{\tau}^{HL} H_{i2}$ 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

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

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

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

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

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

$$\frac{\partial u_t}{\partial x_t} + \gamma = -(1-\alpha)\delta\frac{d\tilde{s}_{t+1}}{dx_t}\frac{dV_{t+1}(\tilde{s}_{t+1})}{d\tilde{s}_{t+1}} \tag{38}$$

$$= -(1-\alpha)\delta\frac{d\tilde{s}_{t+1}}{dx_{t+1}}\left[\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}} + \frac{\partial\tilde{u}_{t+1}}{\partial\tilde{s}_{t+1}}\right] - (1-\alpha)\delta^2\frac{d\tilde{s}_{t+2}}{dx_t}\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}} \tag{39}$$

$$= -(1-\alpha)\delta\rho\left[\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}} + \frac{\partial\tilde{u}_{t+1}}{\partial\tilde{s}_{t+1}}\right] - (1-\alpha)(\delta\rho)^2\left(1+\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\right)\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}}, \tag{40}$$

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

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

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:

$$0 = \frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}} + \tilde{\gamma} + \delta\frac{d\tilde{s}_{t+2}}{dx_{t+1}}\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}} \tag{42}$$

$$= \frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}} + \tilde{\gamma} + \delta\rho\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}} \tag{43}$$

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

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

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

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

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 $(-\infty, \bar{\bar{\tilde{\lambda}}})$,where $\bar{\bar{\tilde{\lambda}}} = \frac{-\eta+\delta\rho^2\zeta}{\delta\rho^2(\zeta-\eta)} > 0$. 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 $\tilde{\lambda} > \bar{\bar{\tilde{\lambda}}}$, $f(\tilde{\lambda})$ 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 $[0,\bar{\tilde{\lambda}})$ and $\lim_{\tilde{\lambda}\to\bar{\tilde{\lambda}}} f(\tilde{\lambda}) = \infty$ 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 $(-\infty, \bar{\tilde{\lambda}}]$, it cannot be true that an increasing, continuous, and convex function that diverges to infinity at $\bar{\tilde{\lambda}}$ only crosses the identity function once in $[0,\bar{\tilde{\lambda}})$. Hence, both solutions are in $[0,\bar{\tilde{\lambda}}]$.

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 $[\tilde{\lambda}_2, \bar{\tilde{\lambda}}]$ 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

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

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

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

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. □

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

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

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

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

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

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

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 $m_{ss} \coloneqq \tilde{x}_{t+1} - x_{ss}$.

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

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

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

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

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

$$x_{ss}\left[\eta + \frac{\rho\zeta}{1-\rho} - (1-\alpha)\delta\rho\left((\eta-\zeta) + \frac{\zeta\rho}{1-\rho}\right)\right] = p - \xi - \gamma + (1-\alpha)\delta\rho\left[(\eta-\zeta)m_{ss} + \xi - p + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \phi\right] \quad (106)$$

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

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

□

G Derivations of Estimating Equations in Appendix E.3

We define $y^g \coloneqq \mathbb{E}_{i\in g} y_i$ 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

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

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

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

Re-arranging gives equation (22).

Derivation of equation (23). Similarly, we have

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

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

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

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

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

Substituting this into equation (22) gives:

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

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

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

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

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

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:

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

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

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

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

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

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

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

This exactly parallels equation (30), except adjusting the denominator for substitution. The survey taker values the total temptation reduction $-\left(\tilde{\tau}_{x3}^B + \tilde{\tau}_{y3}^B\right)$ 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

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

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

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

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]

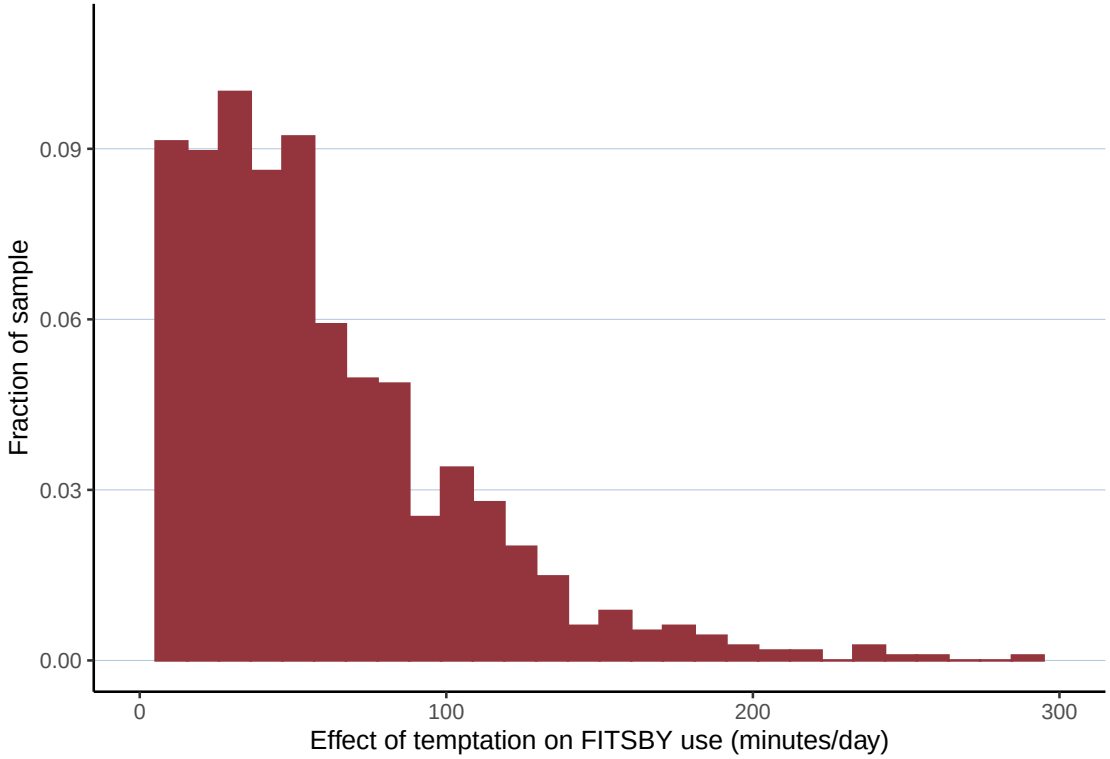
Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for predicted steady-state FITSBY use with different parameter assumptions, using equation (15). The numbers are as plotted in Figure 10.

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.