NBER WORKING PAPER SERIES

DIGITAL ADDICTION

Hunt Allcott
Matthew Gentzkow
Lena Song

Working Paper 28936
http://www.nber.org/papers/w28936

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2021

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, Chicago Harris, Di Tella University, London Business School, London School of Economics, the Marketplace Innovation Workshop, Microsoft Research, the New York City Media Seminar, the New York Fed, NYU, Paris School of Economics, Stanford Institute for Theoretical Economics, Trinity College Dublin, USC, and Wharton for helpful comments. We thank Michael Butler, Zong Huang, Zane Kashner, Uyseok Lee, 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. 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Digital Addiction
Hunt Allcott, Matthew Gentzkow, and Lena Song
NBER Working Paper No. 28936
June 2021
JEL No. D12,D61,D90,D91,I31,L86,O33

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.

Hunt Allcott
Microsoft Research
1 Memorial Drive
Cambridge, MA 02142
and NBER
allcott@post.harvard.edu

Matthew Gentzkow
Department of Economics
Stanford University
579 Jane Stanford Way
Stanford, CA 94305
and NBER
gentzkow@stanford.edu

Lena Song
New York University
lena.song@nyu.edu

A online appendix is available at http://www.nber.org/data-appendix/w28936

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 2020a). 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 matters 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[2] a good idea?

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 digital addiction 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) 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), Banerjee and Mullainathan (2010), 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. (2019) 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 also allows people to misperceive both habit formation and self-control. 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. 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, effective the next day. Unlike limits in existing tools such as the iPhone's Screen Time app, these limits forced participants to stop using the relevant app and in most cases could not be immediately overridden. 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 results provide clear evidence of digital addiction. The bonus treatment 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 next 3 weeks, after the incentive had ended, the bonus treatment group still used 19 minutes less per day. In the 3 weeks after that, they used 12 minutes less per day. The persistence in these later periods is consistent with habit formation.

Participants correctly predict habit formation: the effects of the bonus on predicted FITSBY use after the incentive ended 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 11 percent of what our model would predict for forward-looking habit formation. These results are consistent with a form of projection bias (Loewenstein, O'Donoghue, and Rabin 2003) in which consumers are aware of habit formation but choose as if they are inattentive to it.[3]

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 (17 percent) over 12 weeks. Although the experiment offered no incentive to set limits, 89 percent of participants set binding limits.

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

The average participant was willing to give up \$4.26 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.

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 more addiction measures in the baseline survey.

In the final section of the paper, we look at these results through the lens of our structural model. 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 inattention to habit formation, 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 47.5 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 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. Second, 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. 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, participants were aware that they were participating in an experiment, and we cannot rule out the possibility of experimenter demand effects.

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).[4] Ours is one of a small handful of papers that estimates the parameters of a present focus model with *partial* naivete using field (instead of laboratory) behavior.[5] The digital self-control problems we study are particularly interesting because this is one of the few domains where market forces have created commitment devices, such as smartphone app time use limits and email and web blockers (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.[6] A second set of papers test for forward-looking habit formation using belief elicitation or advance responses to future price changes.[7]

Finally, we extend two literatures that speak directly to digital addiction. The first is a set of experimental papers studying the effects of social media use on outcomes like subjective well-being and academic performance.[8] A second body of work studies the effects of digital self-control tools.[9] Hoong (2021) is particularly related, and is an important antecedent to our study. Ours is the first paper to formally model

[4]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), Augenblick (2018), Kuchler and Pagel (2018), Toussaert (2018), Augenblick and Rabin (2019), Schilbach (2019), John (2019), and Sadoff, Samek, and Sprenger (2020).

[5]To our knowledge, these are Allcott, Kim, Taubinsky, and Zinman (2020b), Bai et al. (2018), Carrera et al. (2019), Chaloupka, Levy, and White (2019), and Skiba and Tobacman (2018).

[6]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), Allcott, Braghieri, Eichmeyer, and Gentzkow (2020a), Bursztyn et al. (2020), Gosnell, List, and Metcalfe (2020), and Van Soest and Vollaard (2019).

[7]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). See Chaloupka and Warner (1999) and Auld and Grootendorst (2004) for discussions of empirical challenges in non-experimental tests.

[8]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 (2020a), and Collis and Eggers (2019).

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

digital addiction and estimate the effects on time use.

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

2 Model

In each period $t \leq T$, consumers choose consumption of a good x_t sold at price p_t which 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.

Consumers may misperceive habit formation, behaving as if the habit formation parameter will be $\tilde{\rho} \in [0,\rho]$ instead of ρ. We say that consumers are fully inattentive to habit formation if $\tilde{\rho} = 0$ and fully attentive if $\tilde{\rho} = \rho$. Inattention to habit formation captures why our experiment participants do not respond much to future price changes, and it is conceptually similar to the projection bias model of Loewenstein, O'Donoghue, and Rabin (2003).

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 simplifies the estimating equations and naturally matches our application to an addictive good instead of a consumption-savings problem.

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 O'Donoghue and Rabin (1999) and others, we solve for perception-perfect strategies, where consumers maximize current utility given perceptions 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 $\boldsymbol{p}_t = \{p_t, p_{t+1}, ..., p_T\}$. A strategy profile $(x_0^*, ..., x_T^*)$ is perception perfect if in each period t

$$x_t^*(s_t, \gamma, \boldsymbol{p}_t) = \arg\max_{x_t} u_t(x_t; s_t, p_t) + \gamma x_t + \sum_{\tau=t+1}^{T} \delta^{\tau-t} u_\tau(x_\tau^*(\tilde{s}_\tau, \tilde{\gamma}, \boldsymbol{p}_\tau); \tilde{s}_\tau, p_\tau), \tag{2}$$

where $\delta \leq 1$ is the discount factor. In period t, consumers know their current habit stock s_t and predict that future habit stock will evolve according to $\tilde{s}_{\tau+1} = \tilde{\rho}\left(\tilde{s}_\tau + x^*_\tau\left(\tilde{s}_\tau, \tilde{\gamma}, \boldsymbol{p}_\tau\right)\right)$. The "rational" habit formation model of Becker and Murphy (1988) is the special case where $\tilde{\rho} = \rho$ 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:

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

where $\eta < 0$ measures the demand slope, ζ allows habit stock to affect marginal utility, ϕ is the direct effect of habit stock on utility, 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).

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 $t = 1, 2, 3, \ldots$ 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 failed quality checks.[10] 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 now 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 was fielded, and after four days we 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.

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 this 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:

[10] Participants were dropped if they (i) reported on survey 1 that they already used another app to limit their phone use (five percent of the sample); (ii) did not to promise to "provide my best answers" on our surveys; (iii) reported having idiosyncratic bugs with Phone Dashboard; (iv) failed to answer more than two of our text message questions between survey 1 and survey 2; (v) had a device that was incompatible with Phone Dashboard; or (vi) were missing screen time data during the baseline period.

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 perceived habit formation (the parameter $\tilde{\rho}$), 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., not now, but 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 correctly describing 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 perceived habit formation $\tilde{\rho}$. 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 the limits as a commitment device that allowed participants to at least partially determine period t consumption before period t. Since participants were not required or incentivized to set limits, they would only have used the functionality if they perceived they had self-control problems.

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 are commitment devices that constrain 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 (2020b) and Carrera et al. (2019). 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 2020a). 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 present a version of our results in Section 7 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 coronavirus, and 88 percent of people reported that coronavirus 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. The decay of the effect in period 5 relative to period 4 is consistent with the exponential decay of the habit stock in the 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 evidence of 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, transitory imbalance, and salience effects following survey 2.[11]

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.[12] *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. 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 limits. Access to the limit functionality reduced use in periods 2–5 by an average of 22 minutes per day, or 17 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 effects suggest that the limit functionality helps address substantial self-control problems.

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 effects on use by app of the bonus (in period 3 only) and the limit (across periods 2–5). Among the FITSBY apps, Facebook sees the largest reductions, followed by web browsers, YouTube,

[11] Some evidence supports the possibility of transitory imbalance: while we control for baseline use when estimating equation (4), Appendix Figure A9 shows that consumption is slightly lower in the Bonus group compared to Bonus Control in the 11 days before survey 2. 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.

[12] 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 $H_{it} = \sum_a H_{iat}$.

Instagram, Twitter, and Snapchat. The effect on other apps (the right-most coefficient) 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 only include participants who report using a single phone, but they 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.

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.

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 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. On survey 3, we then asked people to predict their use in future periods. Figure 7 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. Thus, 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 $4.26 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 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 will show in Section 6 that participants' average valuation of the bonus is consistent with perceived self-control problems.

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).[13]

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

[13]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.

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 affected *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 (2020a). They find that deactivating subjects' Facebook accounts for a four week period increased an index of SWB 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 (2020a) 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.

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

In this section, we use data from the experiment to structurally estimate the model from Section 2.

6.1 Key Theoretical Results

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

Euler equation. The first-order conditions of equation (2) for periods t and $t+1$ can be re-arranged into an Euler equation characterizing the equilibrium relationship between consumption in periods t and $t+1$. To simplify notation, define $u_t := u_t(x_t^*; s_t, p_t)$ as current utility, define $\tilde{x}_\tau := x_\tau^*(\tilde{s}_\tau, \tilde{\gamma}, \boldsymbol{p}_\tau)$ and $\tilde{u}_\tau := u_\tau(\tilde{x}_\tau; \tilde{s}_\tau, p_\tau)$ as predicted consumption and utility for future periods $\tau > t$, and define $\lambda_\tau := \frac{\partial \tilde{x}_\tau}{\partial \tilde{s}_\tau}$ as the 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 $t < T$,*

$$\underbrace{\eta x_t^* + \zeta s_t + \xi_t - p_t}_{\partial u_t / \partial x_t} + \gamma = \delta \tilde{\rho} \left[\underbrace{\eta \tilde{x}_{t+1} + \zeta \tilde{s}_{t+1} + \xi_{t+1} - p_{t+1}}_{\partial \tilde{u}_{t+1} / \partial \tilde{x}_{t+1}} + \tilde{\gamma} + \tilde{\gamma} \lambda_{t+1} - \underbrace{(\zeta \tilde{x}_{t+1} + \phi)}_{\partial \tilde{u}_{t+1} / \partial \tilde{s}_{t+1}} \right] . \tag{6}$$

Proof. See Appendix E.1.

□

With full myopia ($\delta = 0$) or full inattention to habit formation ($\tilde{\rho} = 0$), consumers maximize current-period flow utility, setting the left-hand side of equation (6) to zero. In a "rational" habit formation model with $\tilde{\rho} = \rho$ and $\tilde{\gamma} = \gamma = 0$, the right-hand side adds two effects. First, there is an adjacent complementarity effect where people consume more in period t (driving down marginal utility $\partial u_t / \partial x_t$) if they expect to consume more in $t+1$ (i.e. if future marginal utility $\partial \tilde{u}_{t+1} / \partial x_{t+1}$ is lower). Second, there is a direct habit stock effect where people consume more in period t if the marginal utility from the resulting habit stock $\partial \tilde{u}_{t+1} / \partial s_{t+1}$ is higher.

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

Linear policy functions. With quadratic flow utility, equilibrium consumption is linear in habit stock with slope λ_t. Furthermore, if the consumer's objective function is concave, λ is 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,*

$$x_t^*(s_t, \gamma, \boldsymbol{p}_t) = \lambda_t s_t + \mu_t(\gamma) \tag{7}$$

where λ_t is a function of only $\{\eta, \zeta, \delta, \tilde{\rho}\}$ and μ_t is linear in p_t. Furthermore, if the objective function from equation (2) is concave, then $\lim_{T \to \infty} \lambda_t = \lambda$ 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 $-\eta > \delta \tilde{\rho} (\zeta - \eta) + \delta \tilde{\rho}^2 ((\zeta - \eta) \lambda - \zeta)$.

Proof. See Appendix E.2. That appendix also provides an explicit condition that guarantees concavity. $\square$

Importantly for our estimation, λ is the same for both predicted and actual equilibrium consumption. Consumers mispredict their future temptation γ, but this enters equation (7) only through μ.

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 E.3. $\square$

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*

$$x_{ss} = \frac{\alpha - p(1-\delta\tilde{\rho}) + \delta\tilde{\rho}\left[(\zeta-\eta)m_{ss} - (1+\lambda)\tilde{\gamma}\right] + \gamma}{-\eta - \delta\tilde{\rho}(\zeta-\eta) - \zeta\frac{\rho - \delta\tilde{\rho}^2}{1-\rho}}, \tag{8}$$

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

Proof. See Appendix E.4. $\square$

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 below in Table 4.

6.2 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.[14] Imagine a hypothetical "zero temptation" intervention that sets $\tilde{\gamma} = \gamma = 0$, generating treatment effects τ_t^0. We model the limit functionality as an intervention that reduces perceived and actual temptation by share $\omega \in [0,1]$, generating treatment effects $\tau_t^L = \omega\tau_t^0$.

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.

6.3 Identification and Estimation

Using the theoretical results from Section 6.1, we can now derive equations that characterize how a consumer from our model would behave in our experiment. When populated with data from our experiment, these equations allow us to transparently estimate the model parameters.

Figure 9 illustrates temptation, naivete, and our identification strategies. The three demand curves are desired demand according to preferences before period t, $x_t^*(s_t, 0, p_t)$, predicted demand $x_t^*(\tilde{s}_t, \tilde{\gamma}, p_t)$, and actual demand $x_t^*(s_t, \gamma, p_t)$. Demand is globally linear in price, per Proposition 2. The actual equilibrium at $p = 0$ is point L, and the predicted equilibrium is at point C, so misprediction m is the distance CL. The bonus moves the equilibrium to point J in period 3, so the contemporaneous bonus effect τ_3^B is the distance JK. A zero temptation intervention would move the equilibrium to point G, so the zero temptation treatment effect τ^0 is the distance GL.

For identification, we assume homogeneity: all consumers share the same value of $\{\delta, \rho, \tilde{\rho}, \eta, \zeta, \gamma, \tilde{\gamma}, \phi\}$, and thus the same λ, so consumption heterogeneity is explained by heterogeneity in ξ. We relax this in an extension that allows heterogeneity in γ and $\tilde{\gamma}$. We assume that the discount factor is $\delta = 0.997$ for each three-week period, consistent with a five percent annual discount rate. We estimate the remaining parameters in stages, as described below. Appendix F 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 (7), 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:

[14] 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.

$$\lambda = \frac{\tau_4^B}{\rho\tau_3^B + \rho^2\tau_2^B} \tag{9}$$

$$\rho = \frac{\tau_5^B}{\tau_4^B(1+\lambda)}. \tag{10}$$

The first 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 consumption (a larger λ). The second 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). 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 $\rho \geq 0$.

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

After estimating λ and ρ, we estimate $\tilde{\rho}$, 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 $\tilde{\rho}$, η, and ζ:

$$\tilde{\rho} = \frac{\eta\tau_2^B}{\delta\left[-p^B + (\eta-\zeta)\tilde{\tau}_3^B + \zeta\tilde{\rho}\tau_2^B\right]} \tag{11}$$

$$\eta = \frac{p^B - \zeta\rho\tau_2^B + \delta\tilde{\rho}^2\zeta(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)}{\tau_3^B - \delta\tilde{\rho}^2\lambda\left(\rho\tau_2^B + \tau_3^B\right)} \tag{12}$$

$$\zeta = \frac{-\eta\tau_4^B + \delta\tilde{\rho}^2\eta\lambda\left(\rho^2\tau_2^B + \rho\tau_3^B + \tau_4^B\right)}{\rho\tau_3^B + \rho^2\tau_2^B - \delta\tilde{\rho}^2(1-\lambda)\left(\rho^2\tau_2^B + \rho\tau_3^B + \tau_4^B\right)}. \tag{13}$$

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 (larger $\tilde{\rho}$).

For better intuition, consider the special case with no anticipatory demand response, i.e. $\tau_2^B = 0$. The first equation simplifies to $\tilde{\rho} = 0$: we infer that consumers are fully inattentive to habit formation. The second equation then simplifies to $\eta = \frac{p^B}{\tau_3^B}$: we infer η from the static inverse demand slope. Figure 9 illustrates: with $\tilde{\rho} = 0$, the inverse demand slope is just the ratio of p^B (the vertical distance *KL*) to τ_3^B (the horizontal distance *JK*). The third equation simplifies to $\zeta = \frac{-\eta\tau_4^B}{\rho\tau_3^B}$: if the bonus effect is more persistent between periods 3 and 4, we infer that habit stock has a larger effect on marginal utility (higher ζ).[15]

[15] λ (the effect of habit stock on consumption) and ζ (the effect of habit stock on current marginal utility) are closely related, and they are both primarily identified from the persistence of the bonus effect between periods 3 and 4.

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

$$\gamma-\tilde{\gamma}=m^{C}\cdot\left[-\eta+\delta\tilde{\rho}^{2}\left((\eta-\zeta)\lambda+\zeta\right)\right]. \tag{14}$$

With full inattention to habit formation ($\tilde{\rho}=0$), this equation simplifies to $\gamma-\tilde{\gamma}=m^{C}\cdot[-\eta]$: we infer naivete if people underestimate future consumption. Figure 9 illustrates: with $\tilde{\rho}=0$, naivete $\gamma-\tilde{\gamma}$ is the vertical distance *HC* between actual and predicted marginal utility, and this can be inferred by multiplying misprediction m (the horizontal distance *CL* between actual and predicted demand) by the inverse demand slope.

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 (14) 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

$$\gamma=\eta\tau_{2}^{L}/\omega-\delta\tilde{\rho}\left[(\eta-\zeta)\tilde{\tau}_{3}^{L}/\omega+\zeta\tilde{\rho}\tau_{2}^{L}/\omega-\tilde{\gamma}-\tilde{\gamma}\lambda\right]. \tag{15}$$

With full inattention to habit formation ($\tilde{\rho}=0$), this equation simplifies to $\gamma=\eta\tau_{2}^{L}/\omega$: we infer more temptation if the limit has a larger effect. Figure 9 illustrates: with $\tilde{\rho}=0$, 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}/ω, the horizontal distance *GL* between long-run and present demand) by the inverse demand slope.

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 (2020b).

Let $V_{t}(\tilde{s}_{t},\cdot)=\sum_{\tau=t}^{T}\delta^{\tau-t-1}u_{\tau}\left(x_{\tau}^{*}(\tilde{s}_{\tau},\tilde{\gamma},\boldsymbol{p}_{\tau});\tilde{s}_{\tau},p_{\tau}\right)$ be the period t continuation value function conditional on 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

$$\Delta V_3\left(p^B\right) := V_3\left(\tilde{s}_3, p_3 = p_3^B\right) - V_3\left(\tilde{s}_3, p_3 = 0\right) = -p_3^B \cdot \frac{1}{2}\left(\tilde{x}_3(p_3^B) + \tilde{x}_3(0)\right) - \tilde{\gamma} \cdot \left(\tilde{x}_3(p_3^B) - \tilde{x}_3(0)\right), \quad (16)$$

where $\tilde{x}_3(p_3) = x_3^*(\tilde{s}_3, \tilde{\gamma}, \boldsymbol{p}_3)$ 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:

$$\tilde{\gamma} = \frac{\bar{v}^B - \bar{F}^B + p_3^B \bar{\tilde{x}}_3^{B+BC}}{-\tilde{\tau}_3^B}. \quad (17)$$

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 (17)) 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

$$v^L = V_3\left(s_3, \tilde{\gamma}_3 = (1-\omega)\tilde{\gamma}\right) - V_3\left(s_3, \tilde{\gamma}_3 = \tilde{\gamma}\right) = \tilde{\gamma} \cdot \left(\tilde{x}_3(\tilde{\gamma}) - \tilde{x}_3((1-\omega)\tilde{\gamma})\right) \cdot \frac{2-\omega}{2}, \quad (18)$$

where $\tilde{x}_3(\tilde{\gamma}_3)$ is now shorthand for predicted period 3 consumption as a function of 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 ($\tilde{x}_3(0)$, point *G*) instead of the predicted amount ($\tilde{x}_3(\tilde{\gamma})$, point *C*). The height of this triangle is $\tilde{\gamma}$ and the width is $\tilde{x}_3(\tilde{\gamma}) - \tilde{x}_3(0)$, and thus the area is $\tilde{\gamma} \cdot (\tilde{x}_3(\tilde{\gamma}) - \tilde{x}_3(0)) \cdot \frac{1}{2}$. With $\omega < 1$, the valuation v^L equals the deadweight loss reduction trapezoid starting to the right of point *G* and bounded by segment *CE*.

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

$$\tilde{\gamma} = \frac{\bar{v}^L}{-\tilde{\tau}_3^L(2-\omega)/2}. \tag{19}$$

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

Intercept

Finally, we back out a heterogeneous intercept α_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 α_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 (8). Rearranging that equation gives

$$\begin{aligned}\alpha_i := \delta\tilde{\rho}\phi + (1-\delta\tilde{\rho})\,\xi_i = (1-\delta\tilde{\rho})\,p - \delta\tilde{\rho}\,[(\zeta-\eta)\,m_{ss} - (1+\lambda)\,\tilde{\gamma}] \\ -\gamma + x_{i1}\left[-\eta - \delta\tilde{\rho}(\zeta-\eta) - \zeta\frac{\rho-\delta\tilde{\rho}^2}{1-\rho}\right].\end{aligned} \tag{20}$$

With full inattention to habit formation ($\tilde{\rho}=0$), this equation simplifies to $\alpha_i = \xi_i = p - \gamma + x_{i1}\left(-\eta - \zeta\frac{\rho}{1-\rho}\right)$: we infer larger intercepts for people with higher baseline consumption x_{i1}, after adjusting for price p and temptation γ.

6.4 Empirical Moments and Estimation Details

Table 3 presents the moments and fixed parameter values that are inputs to our estimation. The bonus effects τ_t^B are as displayed in Figure 4, except that 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.[16] The period 2 limit effect τ_2^L is also from Figure 4.

Misprediction m^C is the average across surveys 2–4 of the difference between actual and predicted FITSBY use for participants in the Control group, as displayed in Figure 6. To be consistent with the model, we use only the predictions for period t from survey t, just before period t begins. 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 $\tilde{\tau}_3^L$.[17]

[16] Appendix Table A7 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.

[17] 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 (15), $\tilde{\tau}_3^L$ makes little difference because it is multiplied by $\tilde{\rho}$, which is small. However, in the limit valuation strategy in equation (19), $\tilde{\gamma}$ is inversely proportional to $\tilde{\tau}_3^L$, so a much smaller $\tilde{\tau}_3^L$ would make the estimated $\tilde{\gamma}$ much larger.

For our primary estimates, we assume that the limit fully eliminates temptation ($\omega = 1$). The intercepts α_i match the distribution of period 1 FITSBY use x_{i1} displayed in Appendix Figure A4.

We estimate two specifications: an "unrestricted" model and a "restricted" model where we fix $\tau_2^B = 0$ and thus $\tilde{\rho} = 0$. The restricted model allows very transparent identification, delivers very similar parameter estimates, and has more precise counterfactual predictions. We calculate standard errors by bootstrapping.

In each bootstrap draw, we winsorize all treatment effect moments at $\{\tau_t^B, \tilde{\tau}_t^B, \tau_t^L, \tilde{\tau}_t^L\} \leq 0$. This winsorizes 15 percent of τ_2^B draws but no draws of the other moments. We immediately winsorize parameter estimates (before estimating the remaining parameters in each bootstrap draw) according to the model's theoretical limits: $\{\lambda, \rho\} \geq 0$, $0 \leq \tilde{\rho} \leq \rho$, and $\eta \leq 0$. This affects 15 percent of $\tilde{\rho}$ draws in the unrestricted model but leaves the other parameters essentially unaffected. Finally, we drop any bootstrap draws in which the denominator of the steady-state equation is not positive. This drops 0.73 percent of draws in the unrestricted model, which are generally draws with large $\tilde{\rho}$.

7 Estimation Results and Counterfactual Simulations

7.1 Parameter Estimates

Primary Estimates

Table 4 presents our point estimates and bootstrapped 95 percent confidence intervals. We list the parameter groups in the order that they are estimated, as described in Section 6.3. Column 1 presents the restricted model (fixing $\tau_2^B = 0$ and $\tilde{\rho} = 0$), while column 2 presents the unrestricted model. Since our estimated $\hat{\tilde{\rho}}$ is close to zero, 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 $\hat{\tilde{\rho}} = 0.0492$ in column 2, which is not significantly different from zero. The point estimates suggest that participants were attentive to only $\tilde{\rho}/\rho \approx 0.0492/0.299 \times 100\% \approx 16$ percent of habit formation. Inserting the estimates of $\hat{\lambda}$, $\hat{\rho}$, $\hat{\eta}$, and $\hat{\zeta}$ into equation (11), we calculate that τ_2^B would have needed to be -17.6 minutes per day (compared to the actual point estimate of -1.96 minutes per day) to increase $\hat{\tilde{\rho}}$ to $\hat{\tilde{\rho}} = \hat{\rho} \approx 0.299$. In other words, the anticipatory bonus effect is only 11 percent of what our model would predict with forward-looking ("rational") habit formation. This is striking when combined with the evidence from Figure 7 that participants correctly predicted habit formation. This is consistent with a model where people are intellectually aware of habit formation but consume as if they are inattentive to it.

In the restricted model with $\tilde{\rho} = 0$, the estimating equations become so simple that one can easily calculate the point estimates in Table 4 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 (14) with $\tau_2^B = \tilde{\rho} = 0$ 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.

We think of the limit effect strategy as our primary specification for estimating γ, because it uses observed consumption instead of the survey multiple price lists. The limit changed period 2 FITSBY use by -24.3 minutes per day. Inserting that into equation (15) with $\tilde{\rho} = 0$ and $\omega = 1$ 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.

Alternative Temptation Estimates

Table 5 presents alternative estimates of temptation γ in the restricted and unrestricted models. After repeating the limit effect estimate, the table reports the bonus valuation estimate. Before the bonus MPL on survey 2, the average participant predicted that they would use FITSBY 2.5 and 1.6 hours per day without and with the bonus, respectively. Thus, the average survey taker would have predicted that the price increase would cause a consumer surplus loss from their period 3 self's perspective of $p_3^B \bar{\tilde{x}}_3 \approx \$2.50 \times \frac{1}{2}(2.5 + 1.6) \approx \$5.09$ per day of period 3. This is the trapezoid *ABCD* on Figure 9. The average bonus fixed payment was $\bar{F}^B \approx \$7.03$ per day. Thus, if the average participant perceived herself to be time consistent, she would have been indifferent between the bonus and a certain payment of $\$7.03 - \$5.09 \approx \$1.94$ per day.

In reality, the average participant was indifferent between the bonus and a certain payment of \$64, or $\bar{v}^B \approx \$64/20 \approx \$3.20$ per day over the 20-day period. This excess valuation implies a behavior change premium of $\$3.20 - \$1.94 \approx \$1.26$ per day. This is the parallelogram *BCEF* on Figure 9: the additional temptation reduction benefit that the period 2 survey taker perceives from the reduced FITSBY use caused by the bonus. Rearranging this logic into equation (17) 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 \$4.26, or $\bar{v}^L \approx \$4.26/20 \approx \$0.213$ 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 (19) with $\omega = 1$ gives perceived temptation $\hat{\tilde{\gamma}} = \frac{\bar{v}^L}{-\tilde{\tau}_3^L/2} \approx \frac{0.213}{(-(-22.3)/60)/2} \approx$ 1.15 \$/hour. Using $\widehat{\gamma - \tilde{\gamma}} \approx 0.274$ gives $\hat{\gamma} \approx 1.42$ 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 F.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.[18]

The next three rows in Table 5 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 $\hat{\gamma} \approx 1.21$ \$/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 $\hat{\gamma} \approx 2.11$ \$/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:

$$\omega = \frac{d_2^L - d_2^{LC}}{-d_2^{LC}}. \tag{21}$$

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. Table 5 shows that in the restricted model ($\tilde{\rho} = 0$), the limit effect $\hat{\gamma}$ increases from 1.09 to 2.82, while the limit valuation strategy $\hat{\gamma}$ decreases from 1.42 to 0.985.

Finally, we extend the limit effect strategy to allow for individual-specific heterogeneity in γ. To do this, we exploit the fact that we observe each participant's period 2 *limit tightness* H_{i2}, and 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 A8.[19] For each participant, we insert the fitted limit effect $\hat{\tau}_{it}^L = \hat{\tau}_t^L + \hat{\tau}^{HL} H_{i2}$ into equation (15) to infer γ_i. The final row of 5 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

[18] A simple model would predict that the bonus and limit would both cause either an increase or decrease in FITSBY use on other devices, instead of having effects in opposite directions. For these calculations, we take the self-reports at face value.

[19] 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 A8 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.

would expect.

While the estimates differ across rows of Table 5, all but two of the strategies imply temptation γ between about \$1 and \$1.60 per hour. Our primary strategy (the limit effect) is relatively conservative.

7.2 Counterfactuals: Effects of Digital Addiction on Time Use

Methodology

Using these parameter estimates, we predict the effects of changing temptation and habit formation on steady-state time use. Modifying equation (8), we can predict participant i's steady-state FITSBY use at $p = 0$ as a function of the estimated $\left\{\hat{\tilde{\rho}}, \hat{\eta}, \hat{\zeta}\right\}$ from Table 4 and any values of habit formation, temptation, and misprediction parameters $\{\rho, \gamma, \tilde{\gamma}, m_{ss}\}$:

$$\hat{x}_{i,ss}(\rho, \gamma, \tilde{\gamma}, m_{ss}) = \frac{\hat{\alpha}_i + \delta\hat{\tilde{\rho}}\left[\left(\hat{\zeta} - \hat{\eta}\right)m_{ss} - \left(1 + \hat{\lambda}\right)\tilde{\gamma}\right] + \gamma}{-\hat{\eta} - \delta\hat{\tilde{\rho}}(\hat{\zeta} - \hat{\eta}) - \hat{\zeta}\frac{\rho - \delta\hat{\tilde{\rho}}^2}{1-\rho}}. \tag{22}$$

The sample average prediction is denoted $\bar{\hat{x}}_{ss}(\rho, \gamma, \tilde{\gamma}, m_{ss})$.

Since we can't separately identify ξ_i from ϕ, we must hold constant each participant's intercept $\alpha_i := \delta\tilde{\rho}\phi + (1 - \delta\tilde{\rho})\xi_i$ across counterfactuals. Since this intercept contains $\tilde{\rho}$, we can't predict consumption with counterfactual values of $\tilde{\rho}$.

In equation (22), misprediction m_{ss} and perceived future temptation $\tilde{\gamma}$ do not affect steady state consumption when $\hat{\tilde{\rho}} = 0$. This is because consumers simply optimize current flow utility without regard for future behavior if they perceive that current consumption won't affect future consumption. Furthermore, the denominator of equation (22) shrinks toward zero at higher values of $\tilde{\rho}$, magnifying the effects of changes in the numerator. This is because as consumers place increasing weight on how current consumption affects future consumption, they adjust consumption more in response to changes in the future environment.

For both of these reasons, temptation has larger effects on simulated consumption when $\tilde{\rho}$ is large. Indeed, predicted consumption can become unrealistically small (and even negative) in bootstrap draws of the unrestricted model with larger $\tilde{\rho}$. We thus winsorize each individual's predicted consumption at $\hat{x}_{i,ss} \geq 0$, and we think of the restricted model with $\tilde{\rho} = 0$ as our primary specification.

Counterfactual Results

Figure 10 presents point estimates and bootstrapped 95 percent confidence intervals for predicted average FITSBY use at counterfactual parameter values. For this figure, we use our primary estimates of γ and $\tilde{\gamma}$, which are from the limit effect strategy with $\omega = 1$. For each counterfactual, we present predictions from the restricted model ($\tilde{\rho} = 0$) and unrestricted model ($\rho = \hat{\tilde{\rho}}$). As described above, the unrestricted model has skewed confidence intervals due to bootstrap draws with larger $\tilde{\rho}$.

The first "counterfactual" is the baseline at our point estimates: $\hat{x}_{ss}\left(\hat{\rho}, \hat{\gamma}, \hat{\tilde{\gamma}}, \hat{m}^C\right)$. This mechanically matches baseline average FITSBY use of 153 minutes per day. The second counterfactual removes naivete: $\bar{\bar{x}}_{ss}(\hat{\rho},\hat{\gamma},\hat{\gamma},0)$, with $\tilde{\gamma}=\hat{\gamma}$ and $m_{ss}=0$.[20] As described above, naivete has no effect when $\tilde{\rho}=0$. Because naivete and $\hat{\tilde{\rho}}$ are both so small, the point estimate with $\rho=\hat{\tilde{\rho}}$ is very close to the baseline.

The third counterfactual removes temptation: $\bar{\bar{x}}_{ss}(\hat{\rho},0,0,0)$. Relative to baseline, removing temptation reduces predicted FITSBY use by 47.5 minutes per day (31 percent) with $\tilde{\rho}=0$. 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: $\bar{\bar{x}}_{ss}\left(0,\hat{\gamma},\hat{\tilde{\gamma}},\hat{m}^C\right)$ and then $\bar{\bar{x}}_{ss}(0,0,0,0)$. 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 74.7 minutes per day with $\tilde{\rho}=0$. Without habit formation, removing temptation (going from the fourth to the fifth counterfactual) is just the limit treatment effect: with $\rho=\tilde{\rho}=0$, the limit effect strategy gives $\hat{\gamma}=\tau_2^L\hat{\eta}$, and the effect of removing temptation on $\bar{\bar{x}}_{ss}$ is $\frac{\hat{\gamma}}{-\hat{\eta}}$. This reduces predicted consumption by 24.3 minutes per day with $\tilde{\rho}=0$, which is about half the effect of removing temptation with habit formation (47.5 minutes per day). This quantifies how habit formation magnifies the effects of temptation, because current temptation increases current consumption and thus future demand.[21]

Figure 11 presents ten alternative estimates of the effects of temptation on steady-state FITSBY use, i.e. $\bar{\bar{x}}_{ss}\left(\hat{\rho},\hat{\gamma},\hat{\tilde{\gamma}},\hat{m}^C\right)-\bar{\bar{x}}_{ss}(\hat{\rho},0,0,0)$, for the restricted model with $\tilde{\rho}=0$.[22] The first nine estimates are the nine temptation estimation strategies presented in Table 5, so the left-most coefficient is the 47.5 minutes per day reported above for the limit effect strategy with $\omega=1$. The tenth estimate is for limit effect strategy after reweighting the sample to be more representative of U.S. adults on the five demographic characteristics in Table 2. We say "more representative" because if we weight the sample to be fully representative of U.S. adults, the sample weights become too dispersed and the estimates become imprecise.[23]

The key message from Figure 11 is that our primary estimates (from the restricted model in the unweighted sample using the limit effect strategy with $\omega=1$) are the second most conservative. Other than the limit valuation strategy with $\omega=\hat{\omega}$, all other estimates of γ are larger, so the model's predicted effects of temptation on FITSBY use are correspondingly larger. Furthermore, reweighting on observables also increases the predicted effects of temptation. While our sample may still be non-representative on unobservable characteristics, sample selection bias captured by observables causes us to *understate* the effects of digital addiction.

[20] Since Figure 7 shows that participants predicted habit formation fairly accurately, we attribute all of steady-state misprediction m_{ss} to naivete about temptation. Instead attributing misprediction to misperceived habit formation would not substantially affect the results because estimated misprediction $\hat{m}^C$ is so small.

[21] This highlights a 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 vs. self-control tools. Another potential explanation is that motivation to use the limit functionality decays enough that it outweighs the habit stock effect.

[22] Appendix Figure A35 presents parallel estimates for the restricted model.

[23] Appendix Tables A11–A13 present the demographics, moments, and parameter estimates in the weighted sample. Appendix Tables A9 and A10 present the numbers plotted in Figures 10 and 11.

Throughout the paper, we have seen evidence of heterogeneity across people. The heterogeneous limit effect strategy allows us to estimate temptation $\hat{\gamma}_i$ for each participant in the Limit group, which we can then insert into equation (22) to predict the individual-specific effect of temptation on steady-state FITSBY use. Figure 12 presents the distribution of effects across participants. The effect is less than 10 minutes per day for 26 percent of participants, and over 100 minutes per day for 13 percent.

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. As further evidence of perceived self-control problems, participants valued the limit functionality and were willing to pay a “behavior change premium” for the bonus. 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. Put differently, the model predicts that 31 percent of social media use is not what people would choose for themselves in advance. While social media platforms, smartphone makers, and third parties offer some self-control tools, these results suggest additional unmet demand. More broadly, these results suggest that better aligning digital technologies with well-being should be an important goal of users, parents, technology workers, investors, and regulators.

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. 2020a. “The Welfare Effects of Social Media.” *American Economic Review* 110 (3):629–76.

Allcott, Hunt, Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman. 2020b. “Are High-Interest Loans Predatory? Theory And Evidence From Payday Lending.” Working Paper, available at https://sites.google.com/site/allcott/research.

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.

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, Justin Sydnor, and Dmitry Taubinsky. 2019. "How are Preferences For Commitment Revealed?" 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.

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.

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.

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.

———. 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.

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.

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.

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.

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.

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.

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 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	Control group misprediction (minutes/day)	6.13	$[4.52, 7.72]$
$\tilde{\bar{x}}_3^{B+BC}$	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 ($/day)	0.213	[0.187, 0.239]
p^B	Bonus price ($/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. In each bootstrap draw, we winsorize all treatment effect moments at $\{\tau_t^B, \tilde{\tau}_t^B, \tau_t^L, \tilde{\tau}_t^L\} \leq 0$.

Table 4: **Primary Parameter Estimates**

Parameter	Description (units)	(1) Restricted model ($\tau_2^B = 0$, $\tilde{\rho} = 0$)	(2) Unrestricted model ($\tilde{\rho} = \hat{\tilde{\rho}}$)
λ	Habit stock effect on consumption (unitless)	1.15	1.12
		$[0.609, 3.32]$	$[0.572, 3.16]$
ρ	Habit formation (unitless)	0.299	0.302
		$[0.103, 0.493]$	$[0.104, 0.498]$
$\tilde{\rho}$	Perceived habit formation (unitless)	0	0.0492
			$[0, 0.257]$
η	Price coefficient ($-day/hour^2)	-2.68	-2.72
		$[-2.98, -2.43]$	$[-3.03, -2.49]$
ζ	Habit stock effect on marginal utility ($-day/hour^2)	3.08	3.04
		$[1.65, 8.39]$	$[1.55, 8.35]$
$\gamma - \tilde{\gamma}$	Naivete about temptation ($/hour)	0.274	0.278
		$[0.201, 0.349]$	$[0.205, 0.354]$
γ	Temptation ($/hour)	1.09	1.08
	Limit effect	$[0.884, 1.30]$	$[0.878, 1.32]$
$\bar{\alpha}$	Average intercept ($/hour)	2.41	1.86
		$[1.10, 3.61]$	$[-0.139, 3.49]$

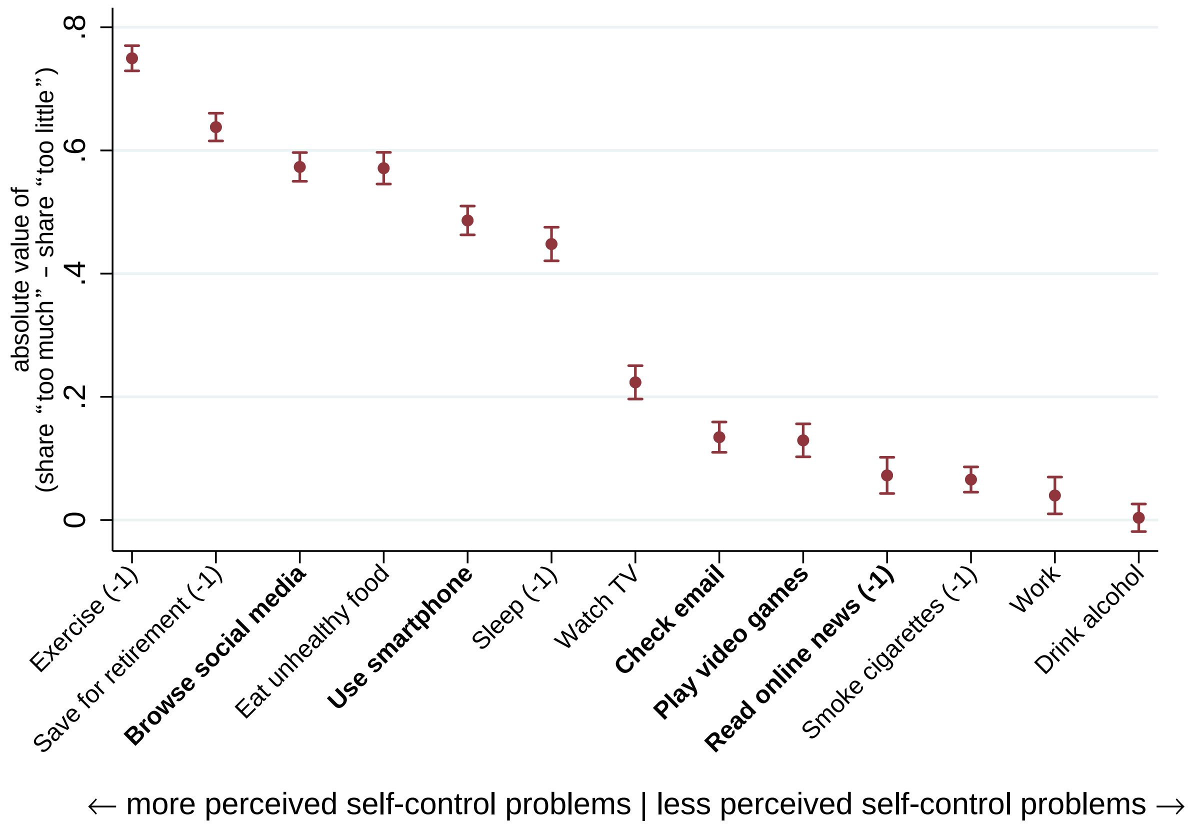
Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals from the estimation strategy described in Section 6.3. In each bootstrap draw, we winsorize parameter estimates according to the model's theoretical limits: $\{\lambda, \rho\} \geq 0$, $0 \leq \tilde{\rho} \leq \rho$, and $\eta \leq 0$. Temptation γ is from the limit effect strategy, using equation (15).

Table 5: **Alternative Temptation Parameter Estimates**

Parameter	Description (units)	(1) Restricted model ($\tau_2^B = 0$, $\tilde{\rho} = 0$)	(2) Unrestricted model ($\tilde{\rho} = \hat{\tilde{\rho}}$)
γ	Temptation ($/hour)		
	Limit effect	1.09	1.08
		$[0.884, 1.30]$	$[0.878, 1.32]$
	Bonus valuation	1.61	1.62
		$[1.29, 1.94]$	$[1.29, 1.94]$
	Limit valuation	1.42	1.42
		$[1.20, 1.77]$	$[1.20, 1.78]$
	Limit effect, multiple-good model	1.31	
		$[1.01, 1.71]$	
	Bonus valuation, multiple-good model	1.21	1.21
		$[0.975, 1.44]$	$[0.979, 1.45]$
	Limit valuation, multiple-good model	2.11	2.11
		$[1.34, 7.18]$	$[1.35, 7.06]$
	Limit effect, $\omega = \hat{\omega}$	2.82	2.87
		$[2.11, 3.92]$	$[2.19, 4.11]$
	Limit valuation, $\omega = \hat{\omega}$	0.985	0.988
		$[0.835, 1.21]$	$[0.841, 1.21]$
$\bar{\gamma}$	Average temptation ($/hour)	1.08	1.06
	Heterogeneous limit effect	$[0.873, 1.29]$	$[0.840, 1.29]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for alternative estimates of temptation γ. γ for the limit effect, bonus valuation, and limit valuation strategies is from equations (15), (17), and (19), respectively, combined with naivete $\gamma - \tilde{\gamma}$ from equation (14). γ for the multiple-good model is from equations (163), (168), and (171) in Appendix F.5; we do not have a limit effect estimate for the unrestricted multiple-good model. $\hat{\omega}$ is from equation (21).

Figure 1: **Online and Offline Temptation**



Notes: This figure presents responses to an online survey in which we asked, "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**

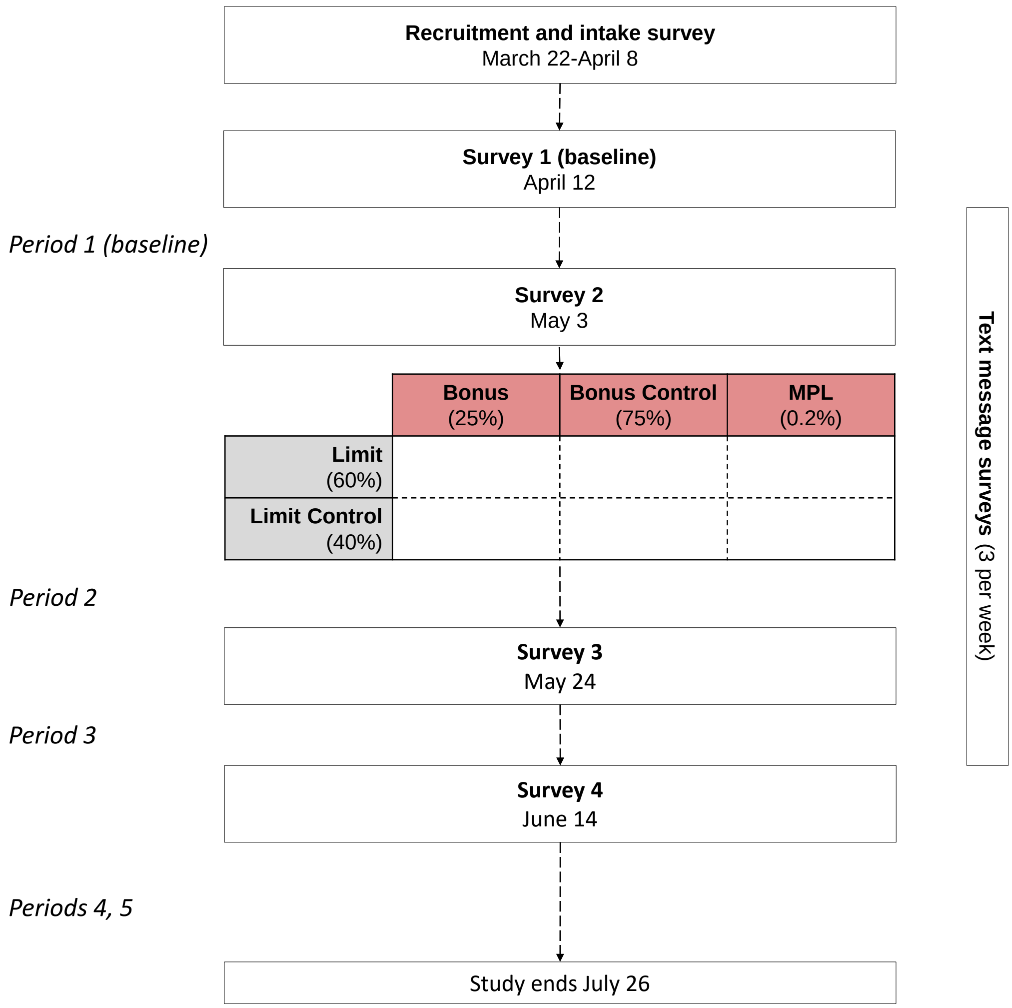
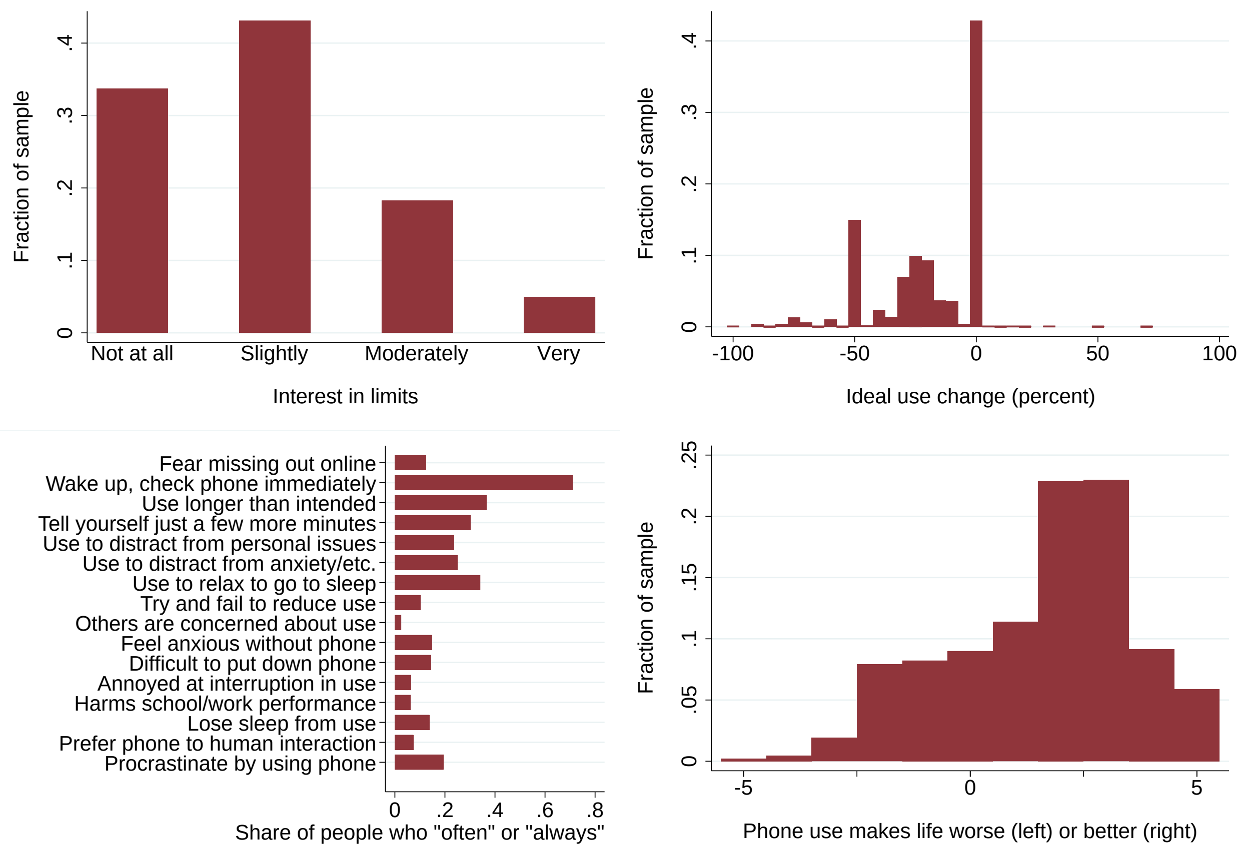
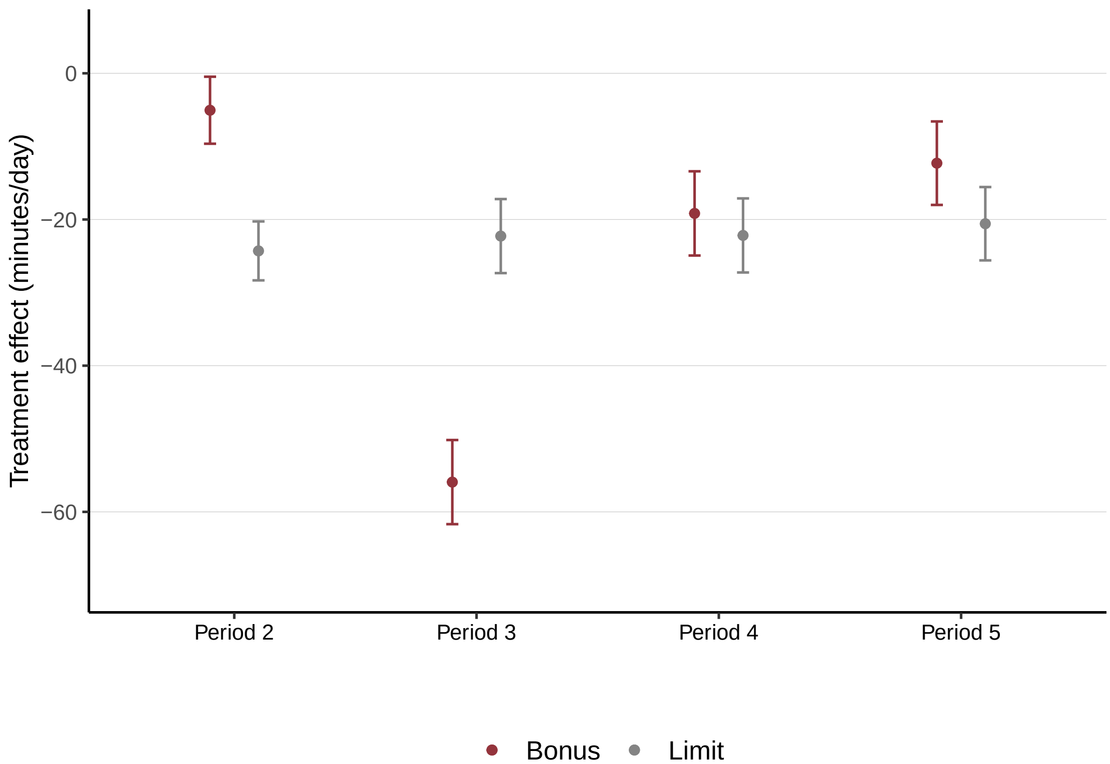


Figure 3: **Baseline Qualitative Evidence of Self-Control Problems**



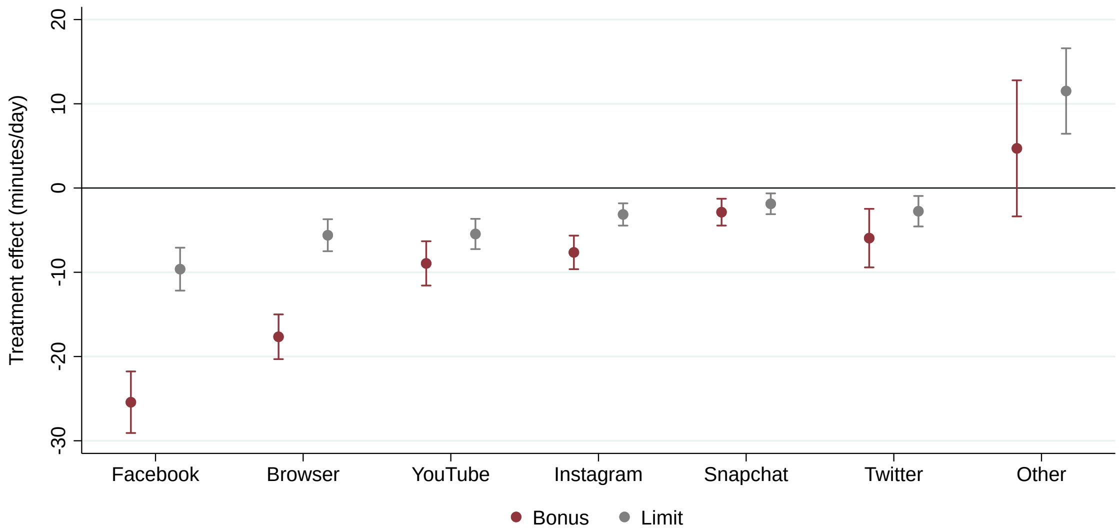
Notes: This figure presents the distributions of four measures of smartphone addiction from the baseline survey. *Interest in limits* is the answer to, "How interested are you to set limits on your phone use?" *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by 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**



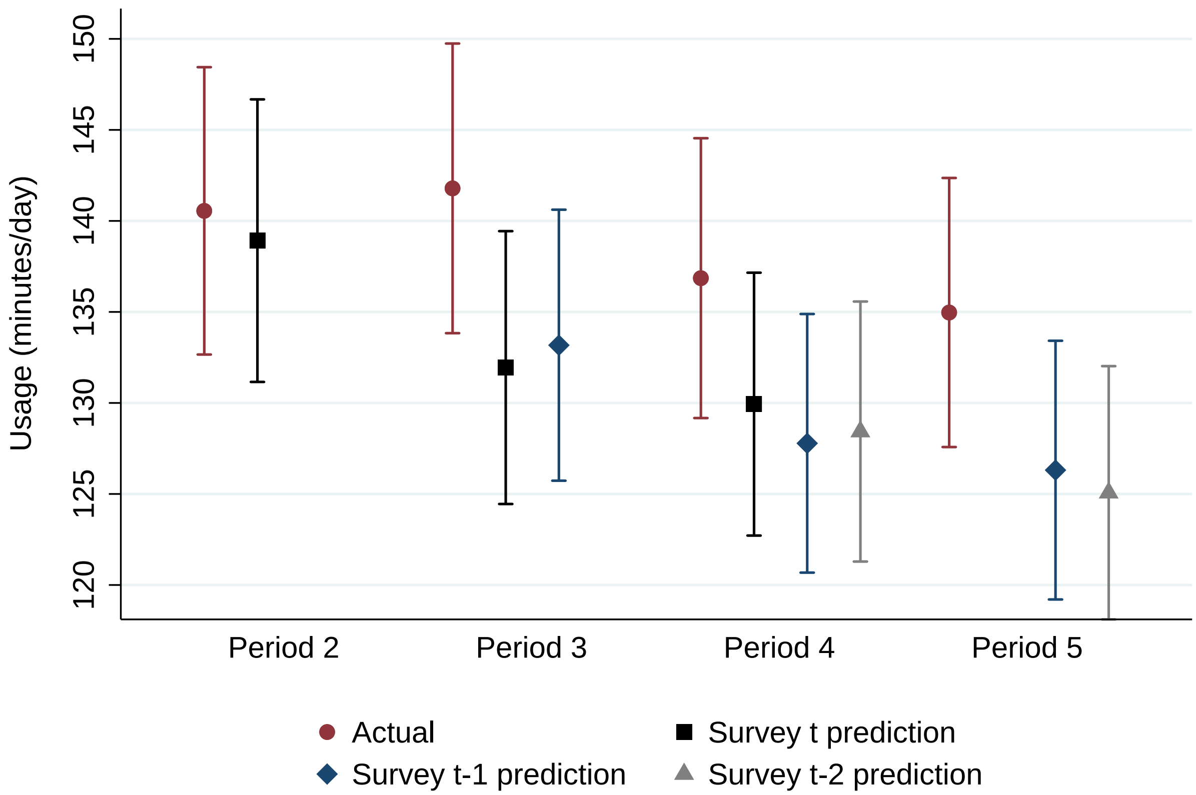
Notes: This figure presents effects of the bonus and limit treatments on FITSBY use using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

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



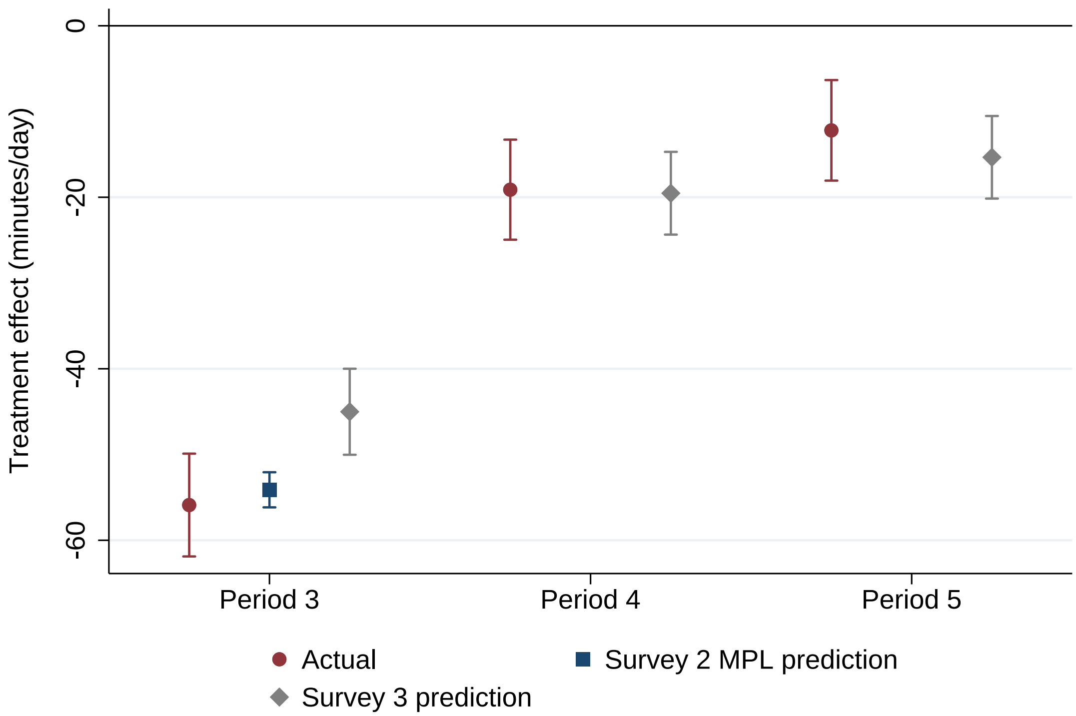
Notes: This figure presents effects of the bonus and limit treatments on smartphone use by app using equation (4). The bonus effects are measured in period 3, while the limit effects are measured in periods 2–5. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and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

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



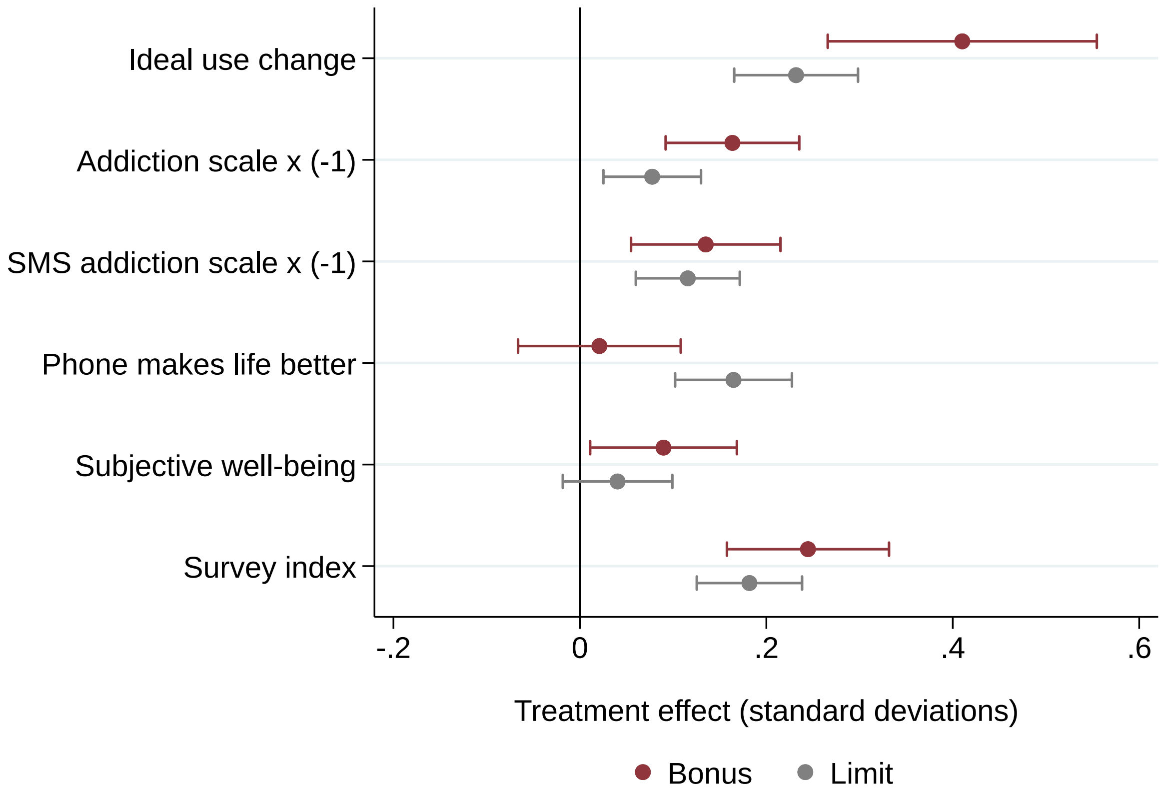
Notes: This figure presents average actual FITSBY use by period and average predicted FITSBY use for that period, for participants in the intersection of the Bonus Control and Limit Control groups. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

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**

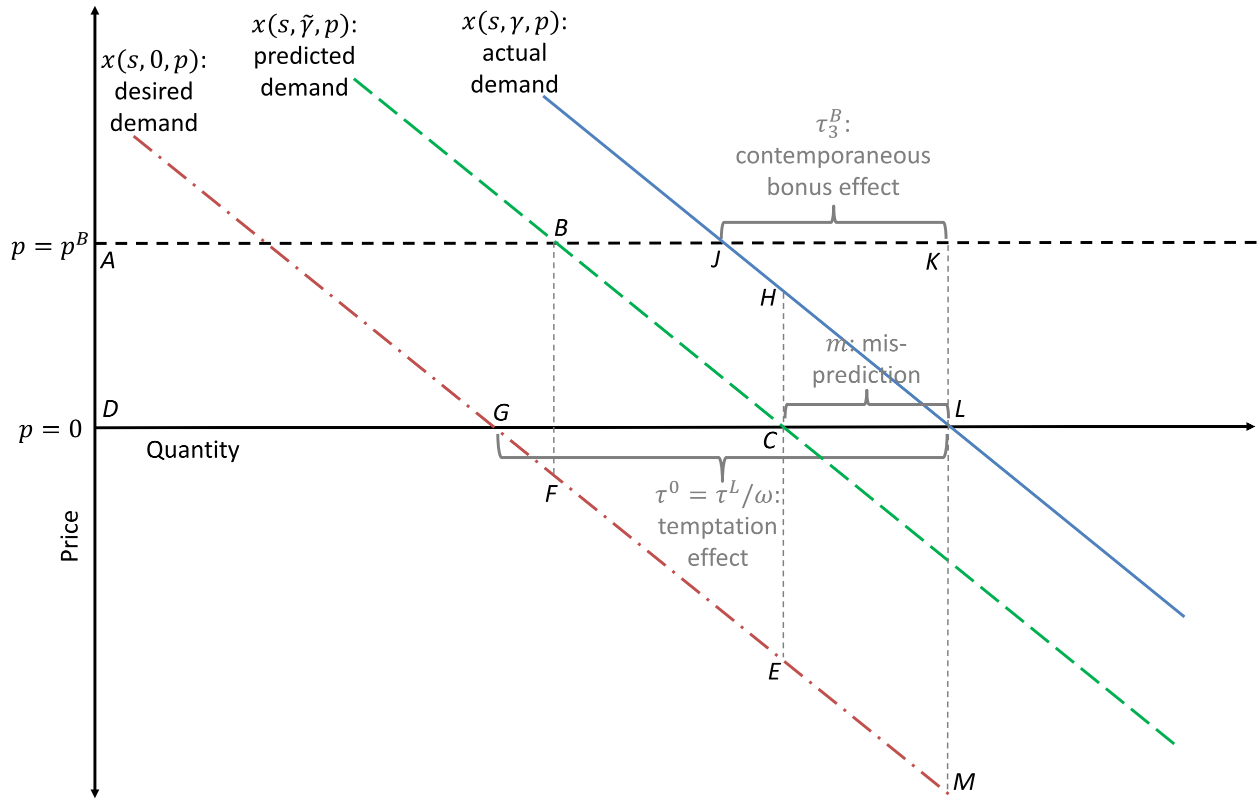
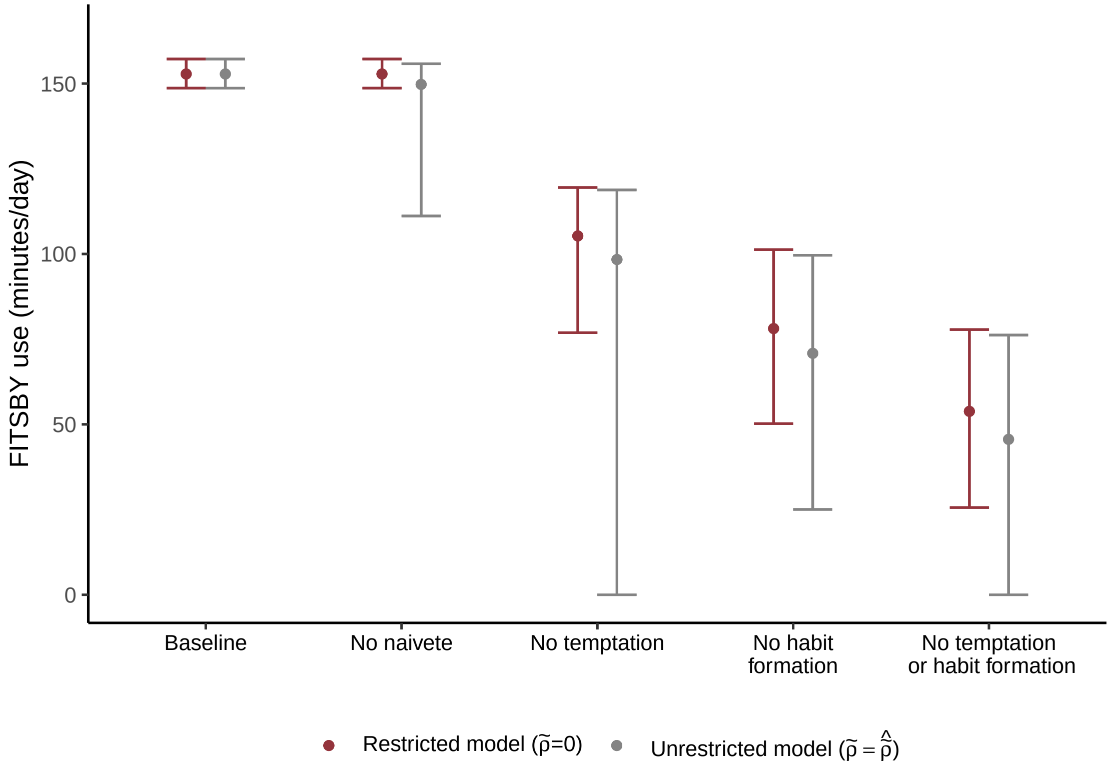
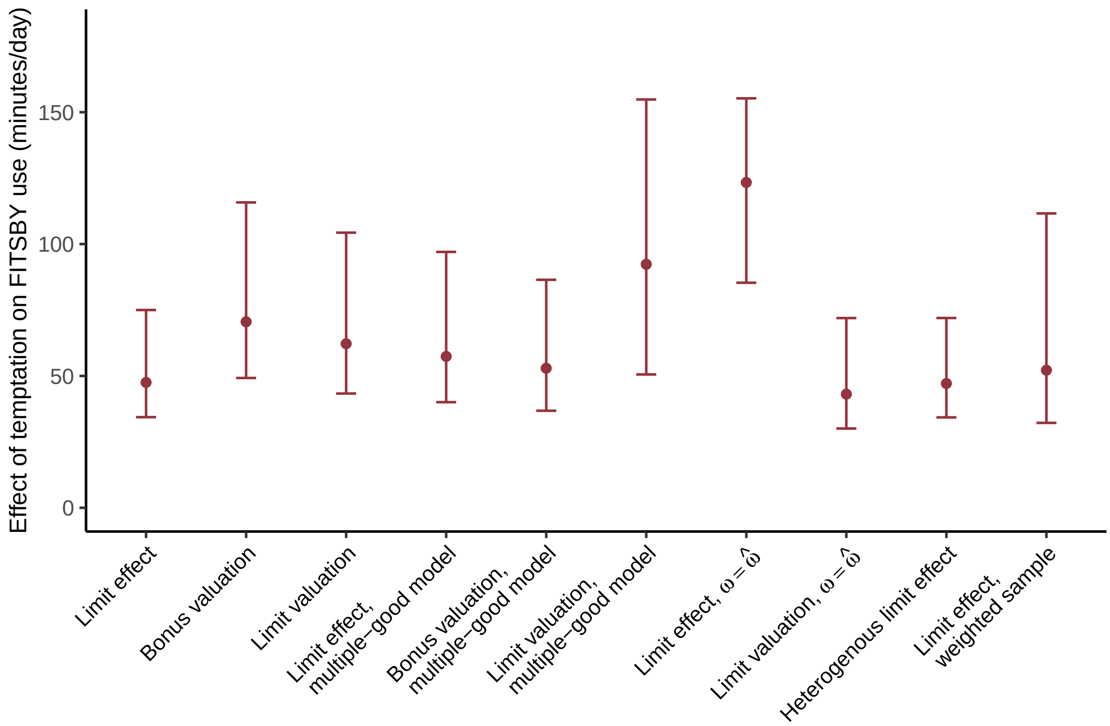


Figure 10: **Effects of Temptation and Habit Formation on FITSBY Use**



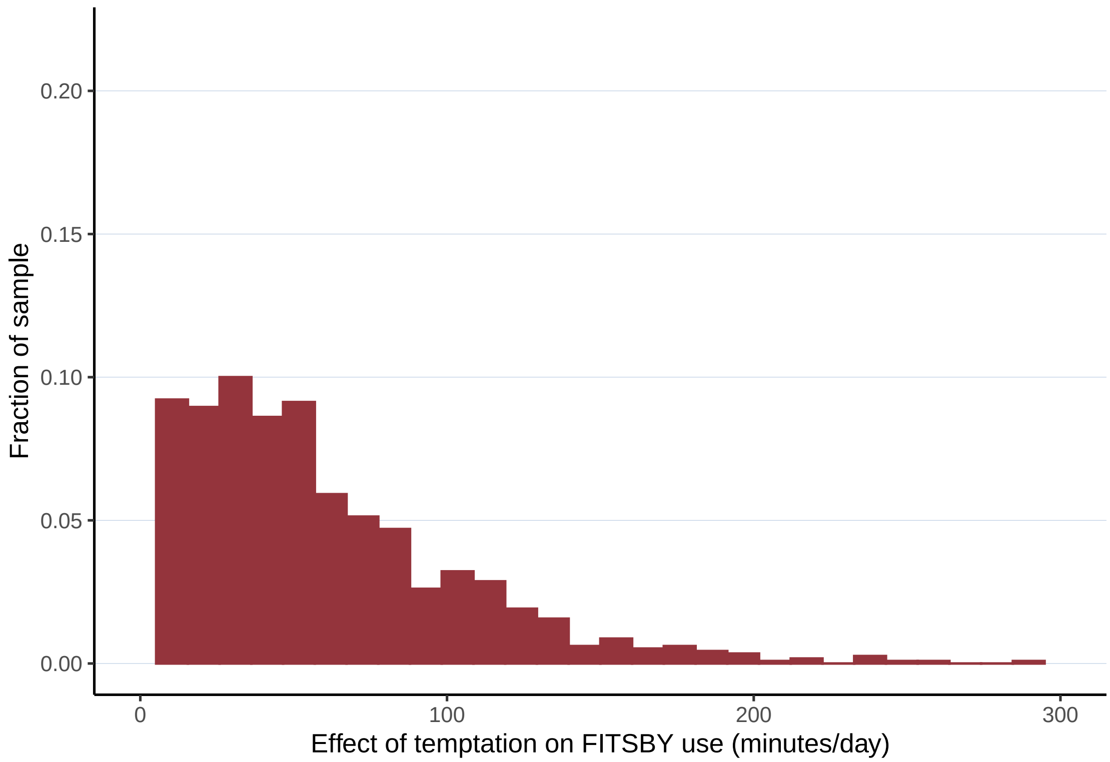
Notes: This figure presents point estimates and bootstrapped 95 percent confidence intervals for predicted steady-state FITSBY use with different parameter assumptions, using equation (22).

Figure 11: **Effects of Temptation on FITSBY Use Under Alternative Assumptions**



Notes: This figure presents point estimates and bootstrapped 95 percent confidence intervals for the effects of temptation on average steady-state FITSBY use in the restricted model ($\tilde{\rho} = 0$), using equation (22). The first nine estimates are for the nine temptation estimation strategies presented in Table 5. 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.

Figure 12: **Distribution of Effects of Temptation on FITSBY Use**



Notes: Using the limit effect strategy, we estimate temptation γ_i for each Limit group participant, which we then insert into equation (22) 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.