Accepted Manuscript

Title: EXPRESS: Leveraging Rational Addiction Theory to Reduce Mobile Usage

Authors: Somasundaram, Jeeva, Zimmermann, Laura, Pham, Quang Duc

DOI: https://doi.org/10.1287/isre.2022.1167

Journal: Journal of Marketing

Citation: Somasundaram, J., Zimmermann, L., & Pham, Q. D. (2025). EXPRESS: Leveraging Rational Addiction Theory to Reduce Mobile Usage. Journal of Marketing, 0(ja)

DOI: 10.1177/00222429251405841

JOURNAL of Marketing

Leveraging Rational Addiction Theory to Reduce Mobile Usage

Journal:	*Journal of Marketing*
Manuscript ID	JM-24-0458.R3
Manuscript Type:	Revised Submission
Research Topics:	Behavioral Decision Theory, Behavioral Economics, Consumer Welfare/Subjective Well being, Public Policy and Marketing
Methods:	Field Experiments, Regression Models, Simulation, Structural Models, Multimethod

Leveraging Rational Addiction Theory to Reduce Mobile Usage

Jeeva Somasundaram, Associate Professor of Decision Sciences, IE Business School, Madrid, Spain, jeeva.somasundaram@ie.edu

Laura Zimmermann, Associate Professor of Marketing, IE Business School, Madrid, Spain, laura.zimmermann@ie.edu

Quang Duc Pham, Assistant Professor of Marketing, CUNEF Universidad, Madrid, Spain, quangduc.pham@cunef.edu

Corresponding author: Jeeva Somasundaram

Author notes: The first and second author contributed equally to the manuscript.

Acknowledgments: We thank the Editors and two anonymous reviewers for very insightful comments, which helped us improve the paper. We are also grateful to Anand Bhatia, Amitav Chakravarti, Pradeep Chintagunta, Enrico Diecidue, Heather Kappes, Ashish Sachdeva, Dayton Steele, Konstantinos Stouras, Luc Wathieu, Bhavani Shankar Uppari, as well as many seminar participants for useful comments on different versions of the paper. We used ChatGPT to refine specific sections of the manuscript's language.

Statements and Declarations

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding Statement

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Ministerio de Ciencia e Innovación, la Agencia y del Fondo Europeo de Desarrollo Regional [Project PID2022-138667OB-I00 funded by MCIN /AEI /10.13039/501100011033 / FEDER, UE], [MCIN /AEI /10.13039/501100011033 / FEDER, UE Grants No. PID2022-138729OA-I00], and [AEI /10.13039/501100011033 Grant No. PID2019-111512RB-I00].

Data Availability Statement

The data that support the findings of this article are publicly available in OSF with the identifier https://doi.org/10.17605/OSF.IO/EBGV9.

Leveraging Rational Addiction Theory to Reduce Mobile Usage

Abstract

The pervasive use of smartphones has raised concerns about their addictive and maladaptive nature. This paper introduces an intervention based on rational addiction theory to cost-effectively nudge consumers to reduce smartphone usage, promoting sustainable digital consumption. We examine whether pre-announcing future targets to reduce smartphone usage influences current consumption and behavioral change. We develop a mathematical model incorporating habit formation, satiation, and projection bias, and test its predictions in three pre-registered randomized control trials using objectively measured smartphone usage. When future incentives and targets are pre-announced, consumers reduce usage pre-emptively compared to their baseline, consistent with rational addiction. This occurs only when participants are given fixed daily reduction targets, not when incentivized proportionally for reductions over time, and seems to reflect forward-looking habit formation, as other explanations (e.g., goal priming or capability testing) were unlikely to drive results. Interestingly, pre-emptive reductions are stronger among heavy users and those with stronger beliefs in meeting their targets. We also find that pre-emptive reductions help consumers meet their targets during the incentivized period and might support post-treatment behavioral sustenance. Our model fitting results reveal considerable heterogeneity and offer insights into how digital detox experiences can be structured to promote sustainable behavior change.

Keywords: smartphone, rational addiction, habits, maladaptive consumption, licensing, projection bias, incentives, goals

Worldwide, there are approximately 7.2 billion smartphone mobile network subscriptions. More than half of the world's population uses social media, typically accessed on one's phone, and an average US teenager spends about 7 hours per day looking at digital media (Gitnux 2025). This widespread and sometimes excessive consumption has led to a heated debate about the addictiveness of digital devices among regulators, tech companies (U.S. Senate Committee on the Judiciary 2024), and researchers concerned about maladaptive smartphone usage (Twenge et al. 2020). It is argued that smartphone usage can turn into an addictive habit with detrimental consequences similar to other vices (Reimann and Jain 2021). Many consumers agree. Half of young adults feel addicted to their phone, and have unsuccessfully tried to limit usage (Ofcom 2020). Increases in anxiety, depression, body image issues, and suicidal thoughts among teenagers have been attributed to social media (Wells et al. 2021), undermining brand image and reputation of major technology firms.

While there is a rich marketing literature on capturing consumers' attention and maximizing online engagement (e.g., Berger, Moe, and Schweidel 2023; Hughes, Swaminathan, and Brooks 2019), far less is known about how to encourage responsible and sustainable technology consumption, considering consumer welfare. We investigate this gap in the smartphone context, building on recent work on digital addiction (Allcott, Gentzkow, and Song 2022) and the theory of rational habit formation (Becker and Murphy 1988).

What is the best way for consumers to reduce smartphone usage, abruptly or gradually? What incentive structure—daily targets or period-based rewards—maximizes the effectiveness of screen time reduction interventions? How can policymakers and technology companies encourage healthier digital behavior while preserving long-term engagement and brand loyalty? Our understanding of these questions has remained limited from a theoretical and practical perspective. One noteworthy exception is Allcott, Gentzkow, and Song (2022) who find limited evidence for rational habit formation and pre-emptive usage reductions in response to future incentives. We extend their work by identifying a practical boundary condition under which consumers can be motivated to anticipate and reduce usage pre-emptively: the granularity of targets.

We test whether the structure of targets and incentives, specifically, how fixed daily targets versus proportional, period-based incentives affect pre-emptive usage reductions. Daily targets may be more effective in prompting pre-emptive reductions because individuals cannot adjust consumption dynamically once the intervention begins. In line with this, we demonstrate that, on average, consumers pre-emptively reduce smartphone usage in anticipation of an intervention with fixed daily targets (e.g., "reduce usage by 90 minutes per day to receive $X") compared to their baseline, but not when incentivized proportionally based on average reductions over a period of time (e.g., "receive $Y for every hour you reduce your average daily screen time below Z hours over a 3-week period"). The pre-emptive reduction allows consumers to gradually habituate to lower consumption levels, facilitating goal achievement during the intervention, and potentially lead to stronger post-treatment sustenance. This contrasts with research on anticipatory licensing, which suggests that knowing about a future deprivation can lead to overconsumption in the present to satiate and compensate for future utility losses (Fishbach and Dhar 2005; Khan and Dhar 2006).

Our empirical studies are grounded in a mathematical model, which studies consumption when both habit forming (Gruber and Köszegi 2001) and satiating (Baucells and Sarin 2007, 2010) forces are at play. In our model, we derive how consumers with varying degree of projection bias (i.e., the extent to which they project current habit stock and satiation levels to the future, ranging from myopic to rationally addicted; Loewenstein, O'Donoghue, and Rabin 2003) behave when expecting to reduce consumption in the future under granular vs. period-based rewards. We then test the model propositions in three pre-registered randomized control trials (RCTs) with objectively measured daily smartphone use.

We also identify sources of heterogeneity. Interestingly, individuals with very high smartphone use, who reported the highest perceived smartphone addiction, benefited most from reducing usage pre-emptively. This is counterintuitive, as one might expect this group to be most resistant to change. These individuals, however, seem to avoid abrupt reductions, preferring slower, more gradual adjustments. Additionally, individuals with strong beliefs in their ability to lower consumption in the future were more likely to proactively reduce usage prior to the intervention. While females reported higher perceived smartphone addiction than males, we do not find moderating effects of gender, or of age on pre-emptive reductions. In an additional survey of consumers' lay beliefs, we find that they prefer a preannouncement phase (i.e., a nine-day lead time), suggesting that having time to adjust gradually increases the appeal of screen time reduction interventions. Finally, we estimate the model parameters based on our empirical data and show that although habit formation is dominant (compared to satiation), there is considerable heterogeneity in projection bias at the individual level.

Theoretically, we contribute to Allcott, Gentzkow and Song (2022) and, more broadly, to Becker and Murphy's (1988) rational addiction theory by identifying the consumption contexts and boundary condition under which consumers are forward-looking in habit formation. We highlight the granular incentive and goal-structure necessary for consumers to be more likely to behave like 'rational addicts.' This extends research on digital addiction, which suggests consumers are generally inattentive to habit formation (Allcott, Gentzkow, and Song 2022). Our findings also differ from studies aimed at increasing the frequency of socially desirable behaviors (e.g., handwashing in rural India; Hussam et al., 2022) by focusing on reducing overconsumption. Finally, our methodology differentiates us from prior research on rational addiction, which relied on non-experimental time series or self-reported consumption data (Baltagi and Griffin 2002; Becker, Grossman, and Murphy 1994; Gruber and Köszegi 2001; Wang 2014; Kwon et al. 2016).

Practically, marketers and policymakers can draw on our theoretical framework and findings to gain insight into the heterogeneity of responses to behavior change interventions, depending on the nature of inter-temporal consumption—habit-forming or satiating—and consumers' projection bias. By identifying these factors, marketers and policymakers could cost-effectively motivate certain segments (e.g., heavy users) to reduce overconsumption by prompting anticipation of future behavior change. For consumers who would like to reduce their own or others' (e.g., children's) technology consumption, our results could offer guidance on how to structure a 'digital detox', potentially leading to more sustainable behavior change and better alignment of technology use with preferences.

The remainder of this manuscript is structured as follows. Based on prior literature, we develop a mathematical model and derive predictions for consumption before, during and after an intervention aimed at reducing consumption depending on 1) whether the good is primarily habit-forming or satiating, 2) whether the consumer is pre-dominantly forward-looking or myopic with respect to habit formation and satiation, and 3) the target and incentive structure. We then present three RCTs testing the model predictions in the smartphone context and use the data to estimate the model parameters. Finally, we discuss theoretical and practical implications.

Theoretical Development

Consider a mobile consumption stream $\boldsymbol{c} = (c_0,\ldots,c_T)$, where c_t indicates mobile consumption in time period $t = 0,\ldots,T$. Consider a consumer choosing between mobile consumption c_t and consumption of another good c'_t in period t, subject to budget constraint $c_t + c'_t = I$. We can think of budget I as the total time available for mobile consumption and alternative activities. The consumer chooses the consumption stream $\boldsymbol{c} = (c_0,\ldots,c_T)$ and $\boldsymbol{c}' = (c'_0,\ldots,c'_T)$ which maximizes total utility over all time periods. The standard economic model that assumes separability over time is the discounted utility (DU) model (Koopmans

1960; Samuelson 1937). The DU is given by $\sum_{t=0,\ldots,T} \delta^t \left(u(c_t) + u(I - c_t)\right)$, where $0 \leq \delta \leq 1$ is the discount rate and u is the utility of consumption.

We consider two factors that affect separability of utility over time: 1) habit formation, where consumption in the previous period increases marginal utility of consumption in the current period (*adjacent complementarity*); and 2) satiation, where consumption in the previous period decreases marginal utility of consumption in the current period (*adjacent substitutability*). In three RCTs, we test if pre-announcing future targets and incentives prompts individuals to pre-emptively reduce smartphone usage and whether such reductions may lead to sustained behavior change post-treatment. We describe the core design next.

Participants were randomly assigned either to a 1) Anticipated Incentive (AI) treatment, 2) Full Incentive (FI) treatment, or 3) Control (C) condition (the second RCT focuses on the AI and C condition only, while the third RCT adds an Anticipated Proportional Incentive (API) treatment). We tracked daily screen time during a baseline, two treatment periods (periods 1 and 2) and a post-treatment period. In the FI treatment, participants were incentivized to reduce daily smartphone usage in both treatment periods. In the AI treatment, participants were only incentivized to reduce usage in period 2 but informed of future targets and incentives at the start of period 1. The C condition received no targets or incentives for reducing mobile usage throughout the study. Below, we derive how adjacent complementarity versus substitutability influences behavior in different periods, depending on consumers' projection bias.

The Effect of Habit Formation

Habit formation implies that greater current consumption increases marginal utility of future consumption, which is known as *adjacent complementarity* (Becker and Murphy 1988; Pollak 1970; Wood and Rünger 2016). If a good is habit-forming, greater consumption of the good in the present (say, one's smartphone) increases demand for future consumption. Becker and Murphy (1988) argue that consumers are forward-looking when forming a habit (i.e., maximizing lifetime utility by taking into consideration future consumption). For instance, a 'rationally addicted' consumer who anticipates a future increase in cigarette prices will start reducing cigarette consumption in the present to reduce the urge to smoke and curb future costs (Gruber and Köszegi 2001). We incorporate the effect of habit formation using the approach by Loewenstein, O'Donoghue and Rabin (2003) and Wathieu (1997). The utility of consumption c_t under habit formation is given by $U = u(c_t,k_t)$, where k_t is the habit stock of mobile usage. Assume that $u' > 0$ and $u'' < 0$ (utility is increasing and concave in consumption). However, the higher the stock of mobile usage, the higher is the marginal utility (or desire for mobile consumption) in the current period, i.e., $\frac{\partial}{\partial k}\left(\frac{\partial \mathrm{U}}{\partial c}\right) > 0$. We assume the stock of mobile usage evolves according to the following equation:

$$k_{t+1} = \gamma(c_t + k_t)$$

γ is the retention rate of the habit stock of mobile usage, where $0 < \gamma < 1$. When $\gamma =$ 0, the model coincides with the discounted utility model. The lower γ, the weaker the influence of past consumption on the stock of accumulated consumption. A similar formulation is used by Gruber and Köszegi (2001) and it is isomorphic to Becker and Murphy (1988). To solve the models, we assume a specific quadratic form for the utility function (Allcott, Gentzkow and Song 2022; Gruber and Köszegi 2001) which allows to capture the essential characteristics:[1]

$$u(c_t,k_t) = -b_1c_t^2 + b_2c_tk_t + b_3c_t \qquad (1)$$

We assume $b_1,b_2,b_3 > 0$ and $c_t < b_3/2b_1$. This ensures: $\frac{\partial \mathrm{U}}{\partial c} > 0$, $\frac{\partial^2 \mathrm{U}}{\partial c^2} < 0$, and $\frac{\partial}{\partial k}\left(\frac{\partial \mathrm{U}}{\partial c}\right) > 0$.

[1] We can also have the direct and quadratic effect of habit stock in the equation (i.e., $u(c_t,k_t) = -b_1c_t^2 + b_2c_tk_t + b_3c_t - b_4k_t - b_5k_t^2$) to capture internality as in Becker and Murphy (1988) and Loewenstein, O'Donoghue, and Rabin (2003) but our results hold even if we include these terms.

There are two habit formation models: (i) the rational addiction model (Becker and Murphy 1988), where consumers are forward-looking and predict future habit stock accurately, and (ii) the myopic habit formation model (Pollak 1970; 1976), where consumers extrapolate current habit stock to the future to calculate the optimal consumption. Loewenstein, O'Donoghue, and Rabin (2003) explore intermediate cases between myopic habit formation and rational addiction by incorporating projection bias (α), which is the extent to which consumers can accurately predict future habit stock when optimizing current consumption. When consumers incorrectly project current habit stock into the future, they exhibit myopic habit formation and projection bias is 1. In contrast, when consumers correctly predict future habit stock, they are rationally addicted and projection bias is zero.

Although we assume that mobile consumption c_t is habit-forming, we assume that the consumption of other goods exhibits no intertemporal complementarities (i.e., habit stock and retention rate are zero in all time periods). A consumer maximizes the total utility over mobile and other consumption given the budget constraint for each period $c_t + {c'}_t = I$. For simplicity, we indicate the utility for other goods with no intertemporal complementarities $u({c'}_t, 0)$ by $u({c'}_t)$.[2] The optimal consumption stream $\boldsymbol{c}_{t'}$ is given by:

For every time period $t' = 0,\ldots,T$,

$$\begin{aligned} \boldsymbol{c}_{t'} = \arg\max_{(c_{t'},\ldots,c_T)} \alpha \times \left(\textstyle\sum_{t=t',\ldots,T} \delta^{t-t'} \left(u(c_t, k_{t'}) + u(I - c_t)\right)\right) \\ + (1-\alpha) \times \left(\textstyle\sum_{t=t',\ldots,T} \delta^{t-t'} \left(u(c_t, k_t) + u(I - c_t)\right)\right) \end{aligned} \quad (2)$$

$$\text{subject to } k_{t+1} = \gamma(k_t + \mathrm{c_t}).$$

When $\alpha = 0$, a consumer exhibits no projection bias and is fully rationally addicted, thus maximizing utility by predicting habit stock k_t correctly in period t (second term of Eq. 2). On the other hand, when $\alpha = 1$, a consumer exhibits full projection bias and is myopic about

[2] The utility from consuming other goods could be represented using a different utility function, without affecting our results.

habit formation, maximizing utility using habit stock in the current period $k_{t'}$ (first term of Eq. 2). When calculating the optimal consumption stream at t', projection bias is not applied to the utility of current consumption ($c_{t'}$), but only to utilities from future consumption. For a detailed discussion, see Loewenstein, O'Donoghue, and Rabin (2003, footnote 14).

We compare our AI and FI treatments to the C condition using the habit formation model in Eq. 2. We consider four periods (periods 0-3): baseline, period 1, period 2, and post-treatment. We assume that consumers derive net utility from the incentives by reducing usage to $((1-\beta)c_0)$, where c_0 is the average baseline usage and β% is the reduction target consumers are incentivized to achieve. During each incentivized period, we assume that participants choose between maintaining optimal consumption based on habit stock (i.e., same as consumption in the C condition) or reducing consumption to the target level. We denote the probability of reducing consumption to $((1-\beta)c_0)$ in period t as p_t. The $p_t \in [0,1]$ is higher, the more attractive the incentive is compared to the loss in utility due to reducing.[3] At the start of period 1, due to randomization, average $c_0^{AI} = c_0^{FI} = c_0^{C}$.

Suppose a consumer is incentivized to reduce mobile usage by β% from their baseline (i.e., reduce usage to $(1-\beta)c_0$) in periods 1 and 2) as in the FI condition. If the incentive is attractive enough compared to the utility loss due to reducing usage, the probability p_t of choosing $(1-\beta)c_0$ becomes higher. For a derivation of optimal consumption in the FI condition under habit formation, see Web Appendix A.

In the AI condition, participants are incentivized in period 2 to reduce usage to $(1-\beta)c_0$. While Proposition 1 predicts anticipatory reductions under non-myopic habit-formation, Proposition 2 highlights the key result: the extent of pre-emptive reductions

[3] The probability can be micro-founded by assuming a modified logit probability $p_t = \frac{2}{1+e^{-\lambda\eta}} - 1$. The probability $p_t \to 0$ when the net utility of reducing usage and receiving incentives (η) is zero, and $p_t \to 1$ when η is extremely positive.

depends on target granularity. We prove the propositions by deriving and comparing optimal consumption levels across conditions (see Web Appendix A for proofs of all Propositions). We assume that net utility from receiving incentives and reducing usage is positive, implying that the probability $p_t > 0$.

PROPOSITION 1. *If participants exhibit habit formation and are not myopic with respect to habit formation ($\alpha < 1$), then AI condition participants have lower usage than C condition participants in period 1.*

Proposition 1 implies that if AI participants are not myopic (i.e., some degree of forward-looking behavior $\alpha < 1$) with respect to habit formation and incentivized sufficiently ($p_t > 0$) to meet the period 2 target (reduce usage to $(1-\beta)c_0$), they will already start reducing their usage in period 1 to lower habit stock, facilitating goal achievement in period 2 and possibly leading to stronger post-treatment sustenance. The extent of any period 1 reduction is inversely proportional to the projection bias (α). In addition, the anticipatory reduction in period 1 should be stronger for individuals with higher absolute reduction targets because meeting such targets is more difficult for them. While all participants had the same relative target β% in the first two RCTs, those with higher baseline usage c_0 faced larger absolute reduction targets (βc_0), leading to stronger pre-emptive reductions (see Web Appendix A for the proof).[4] In contrast, individuals who are myopic with respect to habit formation ($\alpha = 1$) should not reduce period 1 usage compared to the C condition.

Importantly, we analyze how target structure (i.e., granularity) affects pre-emptive reductions. We consider a scenario where period 2 is divided into two sub-periods: 2a and 2b. We compare two target schemes: (i) Sub-period Targets: A target of $(1-\beta)c_0$ is set for each sub-period 2a and 2b separately; (ii) Average Period Target: An average target of $(1-\beta)c_0$

[4] We also show that the extent of pre-emptive reduction increases with baseline c_0. The prediction extends to RCT 3 with a fixed reduction target (i.e., 90 minutes per day).

is set for the entire period 2, that is period 2a and period 2b combined. This helps evaluate how the timing and granularity of targets influence the extent of pre-emptive reductions.

PROPOSITION 2. *If participants exhibit habit formation and are not myopic with respect to habit formation ($\alpha < 1$), then AI condition participants have lower usage in period 1 when given granular sub-period targets compared to an average target for the entire treatment period.*

We prove Proposition 2 in Web Appendix A, with derivations in Web Appendix C. The intuition is as follows: When participants are given an average target for the entire treatment period, they can delay more drastic reductions until later, reducing usage less in period 2a and more in period 2b, to still meet the overall target. This gradual adjustment allows their habit stock to decline slowly, leading to a smaller initial reduction in period 2a. Consequently, they are less motivated to reduce usage pre-emptively in period 1. In contrast, when granular sub-period targets are imposed, it prompts participants to begin reducing usage earlier, in period 1, to prepare for period 2, resulting in larger pre-emptive reductions.

Since participants in the AI and FI condition lower their usage to meet the period 2 target, their habit stock will be lower than in the C condition at the beginning of the post-treatment. As $\frac{\partial}{\partial k}\left(\frac{\partial U}{\partial c}\right) > 0$, participants in the incentivized conditions are expected to maintain a lower post-treatment usage than participants in the C condition.

PROPOSITION 3. *If participants exhibit habit formation, then AI and FI condition participants will have a lower usage than C condition participants in the post-treatment.*

The Effect of Satiation

The AI treatment may backfire if the consumption exhibits the *adjacent substitutability property*—when greater current consumption decreases marginal utility of future consumption. Satiation is a counteracting force to habit formation, as current consumption negatively impacts future consumption (Baucells and Sarin 2007, Iannaccone

1986). Extant physiological and psychological evidence shows that consuming more at a given moment can decrease future consumption for a wide range of products and activities, particularly those of hedonic nature, including foods, fashion, media, and experiences (Galak, Redden, and Kruger 2009; Hetherington, Pirie, and Nabb 2002; Redden and Haws 2013; Small et al. 2001). As Coombs and Avrunin (1977, p. 224) succinctly put it, "Good things satiate." In the domain of technology, satiation has been observed in online and mobile games (Haenlein, Libai, and Muller 2023; Nevskaya and Albuquerque 2019) and is related to the concept of 'social media fatigue' (Bright, Kleiser, and Grau 2015).

A forward-looking consumer, anticipating reduced consumption in period 2, may increase consumption, and thus satiation, in period 1 to compensate for future utility losses. For example, someone planning to start a diet the following week might overindulge beforehand as a form of strategic self-licensing (Fishbach and Dhar 2005; Khan and Dhar 2006). This anticipatory increase in consumption has been dubbed "the last supper effect" (Krpan, Galizzi, and Dolan 2019; Sim, Lee, and Cheon 2018; Urbszat, Herman, and Polivy 2002). It is straightforward to incorporate adjacent substitutability by assuming $b_2 < 0$ in Eq. 1 (we also explore and estimate this model). However, to be consistent with the literature, we incorporate satiation based on Baucells and Sarin (2007, 2010):

$$U(c_t, s_t) = v(c_t + s_t) - v(\mathrm{s_t})$$

where s_t is the satiation level that evolves according to the following equation:

$$\mathrm{s_{t+1}} = \gamma'(s_t + \mathrm{c_t}), \text{ where } 0 < \gamma' < 1.$$

We assume $v' > 0$ and $v'' < 0$, indicating that the higher the satiation level, the lower is the utility of current consumption. We assume a quadratic utility function to solve the model: $v(c_t) = -b'_1 c_t^2 + b'_2 c_t$ with $b'_1, b'_2 > 0$ and $c_t < b'_2/2b'_1$. We use the parameter $0 \leq \alpha_s \leq 1$ to incorporate projection bias in the satiation model. The optimal consumption stream $\boldsymbol{c_{t'}}$ can be calculated similarly under satiation (Web Appendix A: Eq. W4).

Based on this setup, we can compare our AI and FI treatments to the C condition using the satiation model. As before, we assume that participants choose to meet the target and reduce their usage to $(1-\beta)c_0$ with probability p_t depending on the attractiveness of the incentive relative to the utility loss from reducing consumption. Assuming $p_t > 0$, we derive Propositions 4 and 5. Under adjacent substitutability, predictions are reversed. Consumers who are non-myopic with respect to satiation (i.e., a degree of forward-looking behavior with $\alpha_s < 1$) in the AI condition will anticipate a reduction (or deprivation of consumption) in period 2. They will consume more in period 1 to increase satiation and, in turn, experience fewer utility losses when reducing consumption in period 2. In contrast, participants who are myopic with respect to satiation will not change period 1 usage in anticipation of period 2.

PROPOSITION 4. *If participants exhibit satiation and are not myopic with respect to satiation ($\alpha_s < 1$), then AI condition participants have higher usage than C condition participants in period 1.*

Since participants in the FI and AI conditions reduce their usage in period 2, they will have lower satiation levels. As a result, they will increase consumption in the post-treatment.

PROPOSITION 5. *If participants exhibit satiation, then AI and FI condition participants have higher usage than C condition participants in the post-treatment.*

We use Propositions 1-5 to formulate the hypotheses in Table 1 depending on 1) whether habit formation or satiation dominates, 2) whether a consumer is myopic or forward-looking in their habit formation and satiation.[5] We also derive predictions from the DU model as a benchmark in which both habit formation and satiation are absent.

Table 1: Theoretical Predictions for Mobile Usage across Different Conditions and Periods.

		Period 1	Period 2	Post-treatment

[5] A consumer can have more than one of these attitudes. Table 1 predicts the average effect.

Habit formation	Non-Myopic ($\alpha < 1$)	AI < C, FI < C	AI < C, FI < C	AI < C, FI < C
	Myopic ($\alpha = 1$)	AI = C, FI < C	AI < C, FI < C	
Satiation	Non-Myopic ($\alpha_s < 1$)	AI > C, FI < C	AI < C, FI < C	AI > C, FI > C
	Myopic ($\alpha_s = 1$)	AI = C, FI < C	AI < C, FI < C	
Discounted utility		AI = C, FI < C	AI < C, FI < C	AI = FI = C

Methodology and Empirical Results

We conducted three RCTs to test the model predictions. Detailed methodology, experimental instructions, survey items, and additional analyses are available in Web Appendix D. The studies received IRB approval and were pre-registered.[6] Instructions, surveys, data, and R-code for the three RCTs are available at the following OSF link.

RCT 1: Methodology

We recruited 110 international students at a European university (n = 4,400 observations; 74 females; M_{age} = 21.1, SD = 2.25, see Web Appendix D Table W1 for sample characteristics) from different courses by promoting the study via email and during lectures. Participants who completed all tasks received a €15 Amazon voucher and were eligible for a lottery that paid €150. In addition, participants earned a variable payment between €0 and €38 based on target achievement. The rewards were distributed at the end of the study.

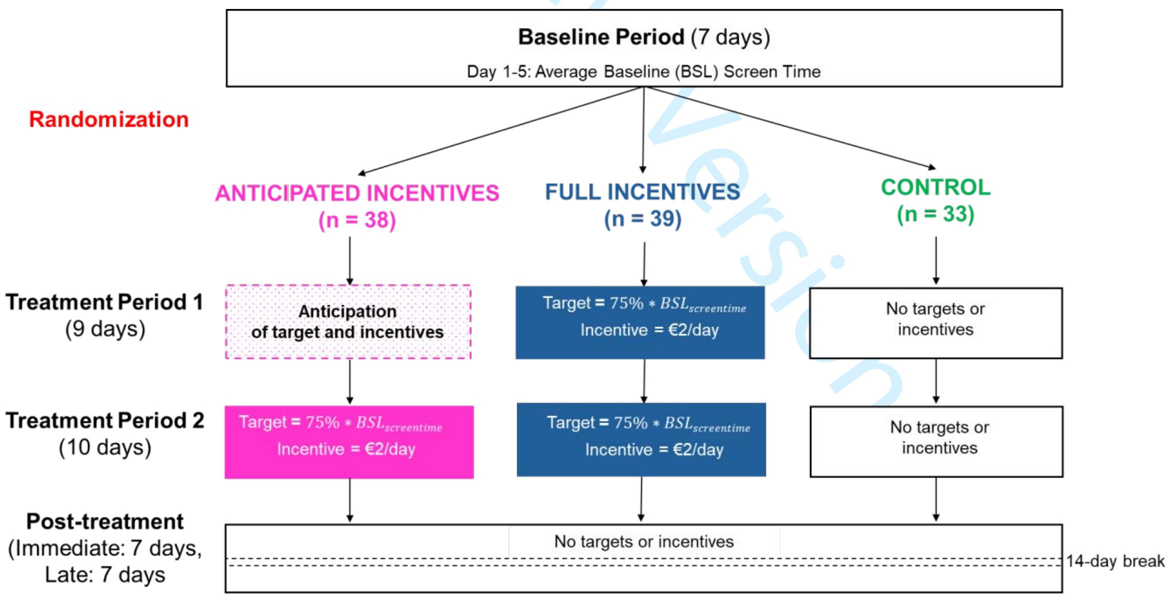
Rather than relying on aggregate-level data (Hussam et al., 2022) or problematic self-reports (Kaye et al., 2020), we objectively measured individual daily screen time over six weeks using screen time applications. Participants had to activate a screen time tracking app on their smartphones. For iPhone users, we used the inbuilt iOS Screen Time app. For

[6] Pre-registration links: RCT 1, RCT 2, RCT 3.

Android users, we used an equivalent app. These apps automatically record mobile screen time without the possibility of excluding certain applications or pausing the recording.[7]

The study was divided into four periods (see Figure 1): baseline (seven days), treatment period 1 (nine days), treatment period 2 (ten days), and post-treatment (14 days). While habit formation timelines vary (Camerer and Li 2021), our 19-day treatment aligns with prior studies (Allcott, Gentzkow, and Song 2022). The treatment periods differed slightly because of constraints of the Screen Time app (screen time had to be reported at weekends and we had to give participants sufficient time to respond). At the end of each week, participants reported their mobile usage for each of the last seven days in a ‘screen time reporting’ survey and uploaded a screenshot of the corresponding usage from the app (see Web Appendix D: Figure W2-A, B). During the baseline period, participants were given neither a screen time target nor an incentive and used their phones as they normally would.

Figure 1: Experimental Design of RCT 1.



Randomization and intervention

After the baseline period, we randomized participants into three conditions:

[7] Whenever a smartphone screen is activated, screen time is automatically recorded (e.g., listening to music while the screen is off, does not count as screen time).

anticipated incentives treatment (AI; N = 38 participants, n = 1,520 observations), full incentives treatment (FI; N = 39, n = 1,560), and control condition (C; N = 33, n = 1,320). We assigned more participants to the treatment conditions (70%) to better detect treatment differences. At baseline, there were no differences across conditions for daily screen time and key observable characteristics (Web Appendix D: Table W1 and Figure W3).

At the start of period 1, all participants were informed of their baseline average screen time and received a handout outlining the benefits of reducing mobile usage. In the treatment conditions, we used the average screen time over the first five days (out of seven) of the baseline period to determine each participant's target.[8] Rather than incentivizing complete abstinence or extremely low usage (Collis and Eggers 2022; Tromholt 2016; Vanman, Baker, and Tobin 2018), we tested a realistic, achievable target of 25% reduction. This target was close to the self-reported ideal reduction of 34% in Allcott, Gentzkow, and Song (2022) and our own pilot study (32%). A key feature of our intervention was the use of *daily* targets to encourage frequent repetition, critical for habit formation (Wood and Rünger 2016), in contrast to Allcott, Gentzkow, and Song (2022), who incentivized participants proportionally based on average reductions over the entire treatment period.

In the treatment conditions, participants first received their individual baseline average screen time (*"During the last 7 days, your average mobile phone usage was X hrs Y minutes per day"*), followed by their personalized reduction target (*"Your target during the NEXT PHASE will be X' hrs Y' mins"*) and the €2 incentive for meeting it (*"During the Z days of the NEXT PHASE, if your daily screen time is less than the target, you will receive €2*

[8] We collected the first screen time data on day 6 and 7 of the baseline period. Hence, we did not have complete data for seven days when randomizing and providing targets on day 8. We collected this data in the subsequent week and used it in all analyses. We also performed the analyses with the 5-day baseline and results remain unchanged.

for that particular day”). They were told that the target was a 25% reduction from their baseline usage, and that targets were unique to each participant.

In the AI condition, participants had no target or incentive in period 1 but were informed at the start of period 1 about the period 2 target and incentive *(“You will have a screen time target IN THE NEXT PHASE, which begins in Z’ days”*). In period 2, they earned €2 each day their screen time was at least 25% below their baseline. In the FI condition, participants received the same daily target and incentive as in the AI condition, but in both periods. They were informed of the target and incentive at the start of each period. In both treatment conditions, participants received two reminder text messages per period about their (future) target. In the C condition, participants were told that there was no specific task at this point and that we were interested in how phone usage fluctuated. They received the same handout on benefits of reducing screen time but were given no targets or incentives.

In the post-treatment, all participants continued reporting their screen time for seven days, followed by a 14-day break with no reporting which was included to test whether the behavior sustained after a month of removing incentives and without surveys. Participants then completed a second post-treatment screen time report where we collected data for another seven days. See Web Appendix D: Figure W4 for the study timeline.

Surveys and attrition

Besides screen time reports, participants completed a baseline and post-treatment survey. The results of self-reported outcomes in RCT 1 and 2 are in Web Appendix D: Tables W4 and W10. Since mobile usage has been linked to lower academic performance (Barwick et al. 2025; Giunchiglia et al. 2018; Zimmermann and Sobolev 2023) we obtained actual GPA at the end of the term. The results are reported in Web Appendix D: Table W6.

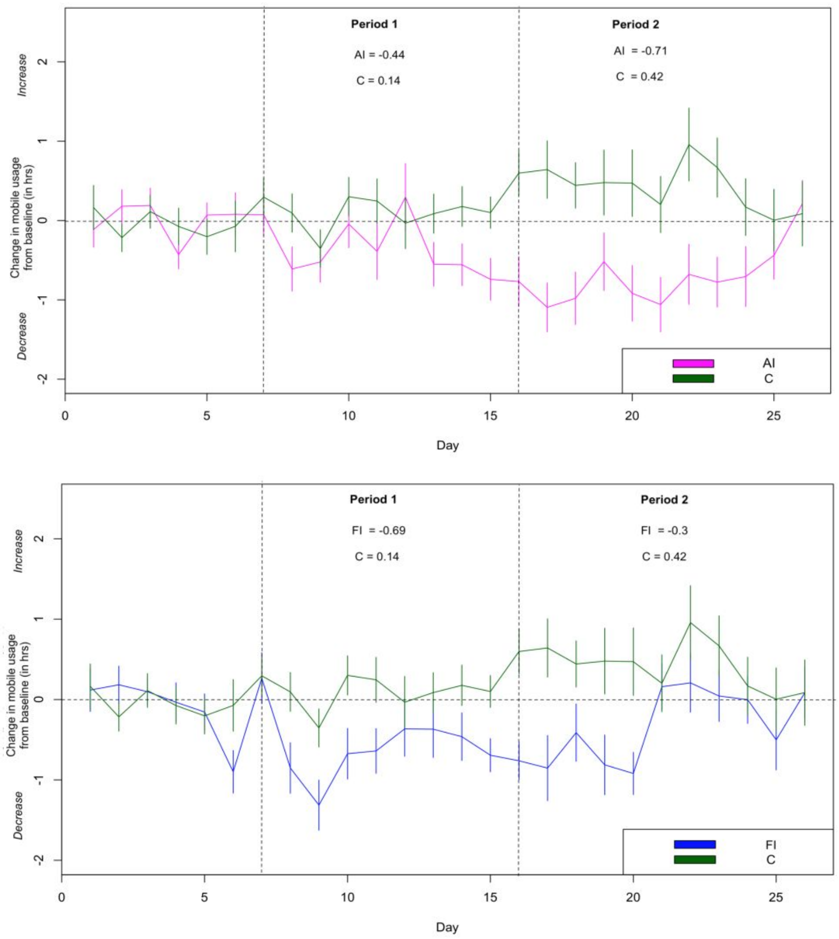
To minimize differential attrition, we closely followed up with participants who missed screen time reports, keeping attrition during treatment below 6% (see Web Appendix

D: Table W2). However, the post-treatment coincided with the onset of COVID-19 lockdowns, leading to a sharp and noisy increase in mobile usage across all conditions. We address this issue separately at the end of the results section.

RCT 1: Results

Before testing the hypotheses, we present the key empirical pattern graphically. Figure 2a (top) compares the average change in mobile usage (from baseline) between the AI and C conditions. Figure 2b (bottom) compares the FI and C conditions. See Web Appendix D: Figure W5a-c for the distribution of individual change in mobile usage across conditions.

Figure 2a–b: AI and FI Conditions Reduce Mobile Usage During Periods 1 and 2.



Notes: Standard error bars are shown.

We test the hypotheses using difference-in-differences ordinary least squares (OLS) regression. We regress mobile usage on (1) the treatment variables (C is the base, AI and FI are dummy variables); (2) the period dummy which captures the change in usage across each period (period 0 or baseline is the base); and (3) the interaction between treatment and period dummies to capture the change in usage from the baseline period in the particular treatment condition. Standard errors are clustered at the subject level to account for potential within-subject correlation. Regression estimates are presented in Table 2. The treatment dummy variables are not significant, indicating no baseline difference in smartphone usage.

Table 2: AI Condition Pre-emptively Reduces Mobile Usage in Period 1.

	Dependent variable = Mobile Usage (hour)			
	Model 1	p-value	Model 2	p-value
AI condition	.362 (.403)	.370	.283 (.389)	.468
FI condition	.423 (.406)	.298	.365 (.417)	.382
Period 1	.122 (.154)	.430	.120 (.154)	.437
Period 2	.443 (.282)	.117	.442 (.282)	.118
AI condition × Period 1	-.552 (.269)	**.041**	-.545 (.267)	**.042**
FI condition × Period 1	-.836 (.251)	**.001**	-.834 (.252)	**.001**
AI condition × Period 2	-1.098 (.393)	**.006**	-1.091 (.392)	**.006**
FI condition × Period 2	-.761 (.367)	**.038**	-.769 (.367)	**.037**
Gender (male)			-.135 (.402)	.738
Age (years)			.081 (.073)	.266
Operating system (iOS)			-.061 (.395)	-.878
Constant	4.529 (.269)	.000	2.958 (1.361)	.030
Observations	2,649		2,649	
R^2	.014		.019	
Adjusted R^2	.011		.015	
Residual Std. Error	2.262 (df = 2,640)		2.258 (df = 2,637)	
F Statistic	4.725 (df = 8; 2,640)		4.677 (df = 11; 2,637)	

Notes: The base category is the C condition in the baseline period. Standard errors in parentheses are robust and clustered at the subject level. P-values < .05 in bold.

We now focus on the treatment × period interaction. In period 1, participants in the FI and AI conditions decreased their usage by .836 hours (p = .001, d = .538) and .552 hours

(p = .041, d = .318), respectively, compared to the C condition. Hence, monetary incentives were effective in reducing usage in the FI condition during period 1. Importantly, participants in the AI condition decreased their usage in period 1, even prior to period 2 when they were actually paid. There was no difference between the FI and AI conditions in period 1 (p = .339), thus anticipated and actual incentives had a similar effect. The pre-emptive reduction in the AI condition was not clustered around days immediately before period 2.

In period 2, participants in the FI and AI conditions decreased usage by .761 hours (p = .038, d = .508) and 1.098 hours (p = .006, d = .365), respectively, compared to the C condition. There was no difference between the FI and AI conditions in period 2 (p = .350). Results are consistent with age, gender, and operating system as controls (Table 2, model 2).

Heavy users show stronger pre-emptive reduction

As predicted by our model, those with larger reduction targets were more likely to pre-emptively reduce their usage in period 1. While all individuals had the same relative target (25% reduction), the absolute reduction (in hours) was greater for those with higher baseline usage. In the AI condition, those with higher absolute targets showed significantly greater reductions in period 1 compared to the baseline (β = -.3, p = .003). In the FI condition, we find no such relationship (p = .17). The results hold controlling for gender, operating system, and age (Web Appendix D: Figures W7-8). This indicates that pre-emptive reductions were highest among heavy users, with potentially the greatest need for reducing usage. It is noteworthy that heavy users responded more strongly to future incentives since it is contrary to the expectation that heavy users would be harder to shift.

Comparing the effect of anticipated versus actual incentives

On average, participants in the AI condition reduced usage by .45 hours in period 1. This is 66% of the .69-hour reduction in the FI condition. To compare AI participants' reduction to similar participants in the FI condition, we employed propensity score matching based on average baseline usage, age, gender, with non-parametric Mahalanobis distance matching. On average, by assigning a participant to the AI instead of FI condition, we achieved 53% of the FI condition's reduction in period 1 without paying incentives.[9]

Post-treatment effect

As a result of the unexpected COVID-19 lockdown, mobile usage in the post-treatment was noisy and surged by ~37%, that is 1.67 hours, across conditions. Due to post-treatment attrition, it was not possible to estimate the treatment effect precisely and the results should be interpreted cautiously. We address this limitation in our second and third RCTs. An OLS regression showed that the FI condition had a marginally lower usage of .864 hours compared to the C condition (p = .052, Web Appendix D: Table W3 and Figure W6), but the AI condition was not different from the C condition in the post-treatment.

Discussion

RCT 1 shows that participants in the AI condition pre-emptively reduced their usage, even prior to the actual incentive period. This reduction was stronger among heavy users (i.e., those with higher baseline usage). In RCT 2, we focused only on the AI and C condition. The goal was threefold: First, we wanted to replicate the findings. Second, we wanted to estimate the post-treatment effect more precisely in the absence of the unforeseeable COVID-19 shock and to have sufficient data for model fitting. Third, we wanted to investigate whether beliefs about future target achievement predict pre-emptive reductions. We conjectured that forward-looking participants who expect to succeed in meeting their targets would proactively reduce usage in the anticipation period to improve their chances of success in period 2.

RCT 2: Methodology

Sixty-eight new students (n = 3,128 observations; 49 females; M_{age} = 25.14,

[9] Genetic and coarsened exact matching led to similar results. The average reduction in the AI condition was 37-45% of the reduction of the FI condition.

SD = 3.91) participated in RCT 2 (Web Appendix D: Table W7 for sample characteristics). We estimated the sample size by assuming 80% power at 5% significance, d = .3 in period 1 (based on AI vs. C in RCT 1) and intra-cluster correlation of .08. The procedures of RCT 2 were identical to RCT 1, barring some key differences: We increased the target to a 30% reduction since we know that participants with higher absolute targets responded more strongly to future incentives. We reduced the anticipation period to eight days and increased the duration of the incentive period to 14 days. Our post-treatment consisted of (i) an immediate phase (9 days) and (ii) a late phase (8 days) with a 35-day winter break between. Other instructions and procedures were identical (see Web Appendix D: "RCT 2" for details).

Surveys and attrition

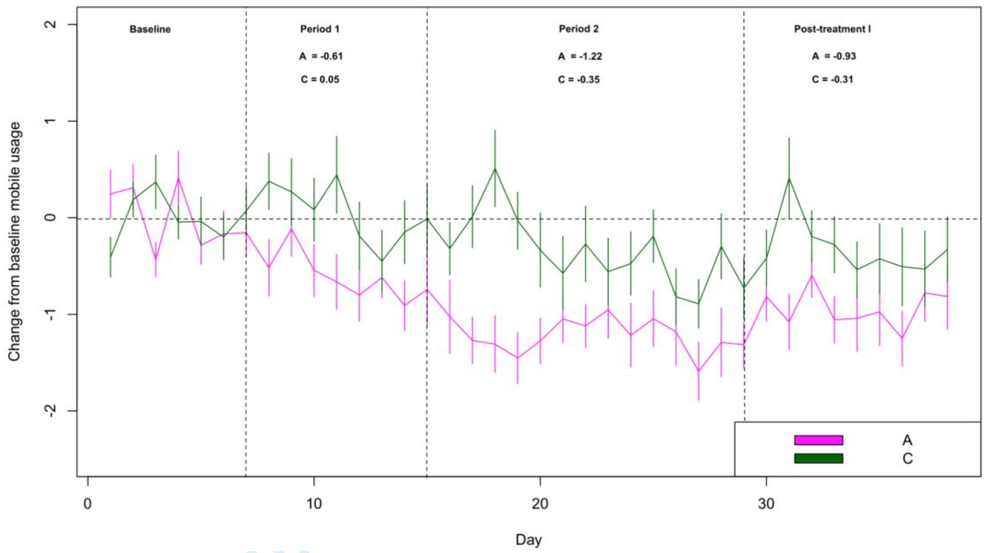
Participants completed a baseline and post-treatment survey in which we repeated some of the RCT 1 measures. Additionally, after period 1, we asked AI participants to explain their reasons for reducing their usage in period 1. We measured beliefs about period 2 target achievement at the beginning and end of period 1 (*"Out of the 14 days in period 2, please predict how many days you will successfully achieve your target screen time",* response scale: 0 – 14 days). Naturally, we elicited these responses only in the AI condition.

Attrition in RCT 2 is reported in Web Appendix D: Table W8. During all periods, attrition was less than ~9% with no differential attrition between conditions, except in the late post-treatment (after the 35-day winter break). While the results suggest consistency in the late post-treatment period, we nevertheless analyze the data separately.

RCT 2: Results

Figure 3 shows the model-free results. We test the hypotheses with difference-in-differences OLS regression (Table 3). The treatment and period dummies are not significant, indicating no baseline difference and no change in smartphone usage across periods.

Figure 3: AI Condition Reduces Usage Before, During and After Being Incentivized.



Notes: Standard error bars are shown.

We now focus on the treatment × period interaction. In period 1, participants in the AI treatment reduced their usage by .789 hours (p = .009, d = .234) compared to the C condition, replicating the pre-emptive reduction in RCT 1. This reduction was stronger among heavy users with higher baseline usage (Web Appendix D: Figures W11 and W12). In period 2, participants in the AI treatment reduced their usage by 1.087 hours per day (p = .002, d = .377) relative to the C condition. In the immediate post-treatment, i.e., for at least nine days after incentives were removed, the AI treatment maintained a lower usage of .766 hours per day (p = .038, d = .279) compared to the C condition. The results are consistent controlling for age, gender, and operating system (see Table 3, model 2).

In the late post-treatment, AI participants continued to show lower usage than those in the C condition. An OLS regression confirmed that this difference was significant ($p < .01$, Web Appendix D: Figure W13 and Table W12). However, given the attrition during this phase, these results should be interpreted with caution. Further robustness checks yielded consistent results (Web Appendix D: "Late post-treatment analysis of screen time").

Next, we provide support for adjacent complementarity and forward-looking behavior using 1) within-subject screen time data and 2) beliefs about target achievement. Our findings indicate that participants exhibited adjacent complementarity, not substitutability, and that they were forward-looking (i.e., non-myopic with respect to habit formation). This is more consistent with predictions of rational addiction models rather than those based on satiation.

Table 3: AI Condition Pre-emptively Reduces and Maintains Lower Mobile Usage.

	Dependent variable = Mobile Usage (hours)			
	Model 1	p-value	Model 2	p-value
AI condition	.207 (.483)	.668	.294 (.466)	.529
Period 1	.101 (.207)	.627	.105 (.206)	.610
Period 2	-.233 (.240)	.334	-.232 (.242)	.338
Immediate Post-treatment (Period 3)	-.280 (.244)	.251	-.284 (.243)	.244
AI condition × Period 1	-.789 (.298)	**.009**	-.773 (.292)	**.009**
AI condition × Period 2	-1.087 (.344)	**.002**	-1.071 (.343)	**.002**
AI condition × Period 3	-.766 (.368)	**.038**	-.738 (.368)	**.045**
Gender (male)			-.589 (.448)	.190
Age (in years)			-.072 (.048)	.140
Operating System (iOS)			.353 (.605)	.561
Constant	5.068 (.331)	.000	6.686 (1.636)	.000
Observations	2,399		2,399	
R^2	.036		.079	
Adjusted R^2	.033		.075	
Residual Std. Error	2.304 (df = 2391)		2.254 (df = 2388)	
F Statistic	12.859 (df = 7; 2,391)		20.472 (df = 10; 2,388)	

Notes: The base category is C condition during the baseline period. Standard errors in parentheses are robust and clustered at the subject level. P-values < .05 in bold.

Within-subject evidence for adjacent complementarity

A key feature of habit formation is adjacent complementarity (i.e., lower usage in period t leads to lower habit stock and therefore lower usage in period $t + 1$). To test adjacent complementarity, we analyze whether participants who frequently reduced usage during periods 1 and 2, sustained lower usage in the post-treatment. If adjacent substitutability plays a role, we should observe the opposite. The change in post-treatment usage with respect to the baseline was the dependent variable. We regressed it on the change in usage in periods 1 and 2 (with respect to baseline) as the independent variable. We cannot run this OLS regression directly as period 2 reductions might be correlated with un-observables and variables that predict post-treatment usage. Therefore, we ran a two-stage least squares estimation using condition as an instrument to predict periods 1-2 reductions. Controls include gender, age, and operating system (see Web Appendix D for the equation).

The condition is a good instrumental variable as it meets two criteria: It predicts each period's reduction and target achievement (relevance) and is not correlated with the error ϵ_i due to randomization and because participants have no task in the post-treatment (exclusion restriction). Results of the two-stage instrumental variable regression are in Table 4.

Table 4: Instrumental Variable Regression Suggests Adjacent Complementarity.

	Dependent variable = Post-treatment usage – Baseline (hrs)			
	(1)		(2)	
	Beta (SE)	p-value	Beta (SE)	p-value
Avg. (P1 usage – Baseline)	.867 (.385)	**.025**		
Avg. (P2 usage – Baseline)			.690 (.200)	**.001**
Gender (male)	.334 (.269)	.215	.257 (.199)	.196
Age	.031 (.035)	.372	.006 (.025)	.809
Operating System (1 = iPhone; 0 = Android)	1.179 (.339)	**.001**	.345 (.326)	.291
Constant	-2.106 (1.181)	.075	-.437 (.965)	.651
Observations	1,068		1,068	
R^2	.079		.288	
Adjusted R^2	.075		.286	
Residual Std. Error (df = 1063)	1.879		1.651	

Notes: Baseline is the average usage during baseline. DV includes immediate and late post-treatment data. Standard errors in brackets are robust and clustered at the subject level. P-values < .05 in bold.

A positive coefficient indicates that a reduction in period 2 (with respect to the baseline) leads to lower post-treatment usage. We find that participants who reduced usage in period 1 (anticipation) and period 2 (incentive phase) tended to have lower post-treatment use, supporting adjacent complementarity (habit formation) rather than adjacent substitutability (satiation). This suggests that mobile usage is predominantly habit-forming, highlighting the potential importance of reductions during anticipation and treatment periods for post-treatment sustenance (see Web Appendix D: Table W13 for additional results).

Belief in target achievement: within-subject evidence for forward-looking behavior

We measured AI participants' belief in their period 2 target achievement both at the start and end of period 1. If participants are forward-looking ($\alpha < 1$) with respect to their habit-formation, those who predict higher target achievement in period 2 (at the start of period 1), should proactively reduce usage in period 1 to lower habit stock and facilitate target achievement in period 2 for the predicted number of days. To test this, we regressed the number of days AI participants met their target in period 1 on their predicted target achievement in period 2 (measured at the start of period 1), controlling for gender, age, and operating system (see Table 5).[10] Results show that belief in period 2 target achievement positively predicts the number of days participants reduced screen time by 30% in period 1 ($p < .001$). In other words, participants who expected to achieve their targets for more days in period 2, were more likely to pre-emptively reduce their usage in period 1. This is consistent with forward-looking (non-myopic) behavior with respect to habit formation ($\alpha < 1$).

Participants' explanations of Period 1 behavior

At the end of RCT 2, we asked AI condition participants if they had reduced their screen time pre-emptively in period 1, and if so, to explain why. A majority (61.76%) reported that they had tried to reduce screen time during period 1 (26.47% indicated no

[10] There was no difference in belief about period 2 target achievement measured at the start vs. end of period 1. Belief about period 2 target achievement measured at the end of period 1 also predicted period 2 target achievement ($p < .001$).

change, 11.76% indicated an increase). Participants provided open-ended explanations of their behavior. A content analysis revealed that 33% of explanations aligned with forward-looking habit formation (verbatim quote: *"I knew it was going to be more difficult to reduce 30% without doing it gradually. Therefore, I used this week to get used to this new usage").* Other explanations were either random or focused on personal goals. Four participants indicated an increase in screen time during period 1, with explanations related to random coincidences (e.g., insomnia, work). None were consistent with satiation. See Web Appendix D: Table W9 for the qualitative analysis.

Table 5: Belief in Period 2 Target Achievement Predicts Period 1 Reduction.

	Dependent variable = Days with ≥ 30% screen time reduction (Period 1)	
	beta (SE)	p-value
Belief in period 2 target achievement at start of period 1	.726 (.235)	**.005**
Gender (male)	2.268 (1.72)	.198
Age	-.156 (.194)	.429
Operating System (1= iOS; 0 = Android)	-1.084 (2.383)	.653
Constant	4.095 (8.681)	.429
Observations	35	
R^2	.273	
Adjusted R^2	.176	
Residual Std. Error (df = 30)	4.558	
F Statistic (df = 4; 30)	2.811	

Notes: p-values < .05 in bold.

To assess lay beliefs about pre-emptive behavior when faced with future incentives and targets, we conducted an additional survey with a separate student sample in which we explained the RCT setup and asked about anticipated behavior. Most respondents expected they would begin reducing usage before the incentive period, consistent with forward-looking habit formation. Participants also expressed a preference for an anticipation period of on average nine days to prepare and gradually change habits, rather than starting an intervention immediately. Details are in Web Appendix D: *"Lay Belief Survey"*.

Discussion

RCT 2 replicates the results that consumers, especially those with excessive screen time, pre-emptively reduce their smartphone usage when anticipating a future reduction target. Participants maintained a lower smartphone usage during the incentivized period, and frequent target achievement (during both anticipation and incentive periods) led to stronger post-treatment sustenance, providing preliminary evidence for adjacent complementarity.

Our findings contrast with those of Allcott, Gentzkow, and Song (2022), who report consumers are inattentive to habit formation. This discrepancy may stem from differences in sample composition. We conducted our RCTs with students of a narrow age range. Prior research suggests that age can matter in studies on screen time reduction (Pedersen et al. 2022; Schmidt-Persson et al. 2024). Cultural context and associated mobile usage patterns could also play a role. To address this, we used a representative sample of US adults in RCT 3, allowing for a direct comparison, and to investigate age as a moderating factor.

A key methodological difference remains, however: we used fixed daily reduction targets (i.e., *"if your daily screen time is less than the target, you will receive €2 for that particular day")*, while Allcott, Gentzkow, and Song (2022) proportionally incentivized the average reduction over the treatment period (i.e., "*you 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"*).[11] Based on Proposition 2, as well as simulations (Web Appendix B and C), we argue that pre-emptive reductions will be stronger when participants must meet fixed daily targets because these create more immediate pressure to adjust behavior early, leaving

[11] FITSBY refers to Facebook, Instagram, Twitter, Snapchat, web browsers, and YouTube. The bonus benchmark [X] refers to a participant's baseline average FITSBY hours per day.

less flexibility to delay reductions. To empirically test this, we added a new condition in RCT 3 where participants were incentivized proportionally for any reduction over the entire treatment period (like Allcott, Gentzkow and Song's (2022) Bonus Treatment). We also included further belief and process measures to explore potential alternative explanations, such as goal priming, practicing, and potential device substitution.

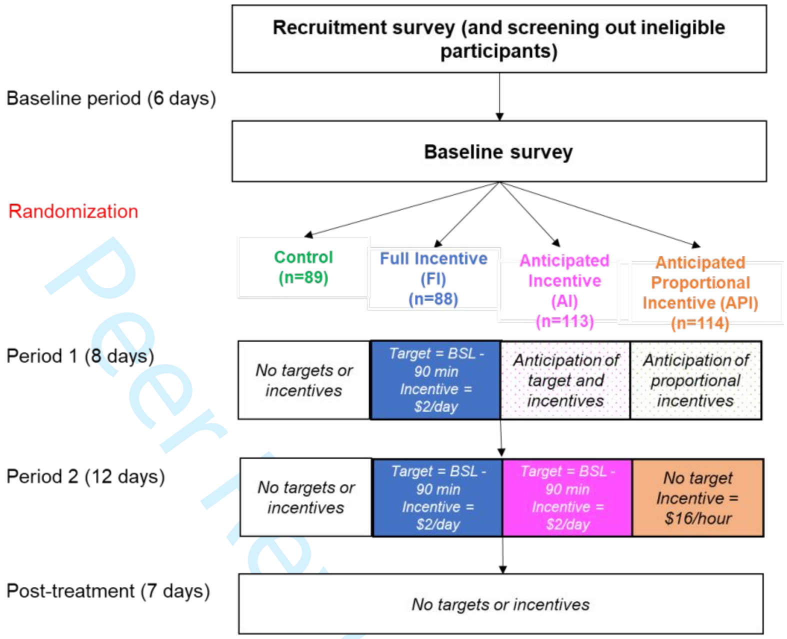
RCT 3: Methodology

We recruited a representative sample (N = 1,047) of US residents (factors: gender, age, ethnicity) using Prolific. Out of those, 455 were excluded based on a screening survey. We screened out individuals (i) without or with more than one smartphone (n = 181), (ii) with neither iOS nor Android operating system (n = 5), (iii) without a phone-embedded screen time app (n = 119), and (iv) with a daily mobile usage of less than three hours (n = 264). We also screened out three individuals who did not commit to completing the study. After these exclusions, 595 participants were invited to complete the baseline survey, of whom 436 responded and were randomized. A total of 404 participants (56% females, M_{age} = 42.75) completed the randomization and period 1 survey. These constitute our analysis sample (total observations n = 13,332). Demographics are shown in Web Appendix D: Table W15. We observed 5.6% attrition of observations across the study. The average attrition did not exceed 11% in any period. Importantly, there was no differential attrition between conditions (Web Appendix D: Table W16). Participants received the standard Prolific payrate for completing each survey, and a $4 bonus for completing all surveys. Participants were also entered in a lottery for $250. In addition, 75% of participants received variable incentives up to $36.

Experimental design and procedure

Figure 4 displays the experimental design of our six-week study. Participants completed a screening survey, and five screen time reporting surveys administered at one-week intervals. As in the previous RCTs, participants reported their daily smartphone usage over the past week and submitted a screenshot of the respective graph from their screen time app. Participants also reported their weekly social media usage.

Figure 4: Experimental design of RCT 3.



We measured the average baseline usage for six days and then randomized participants into four conditions: control, full incentive, anticipated incentive, and anticipated proportional incentive (API) condition, with 22.5%, 22.5%, 27.5% and 27.5% probability, respectively. We stratified based on age, gender, baseline usage, and operating system. We allocated slightly more participants to the AI and API conditions to have sufficient power to detect differences during period 1.[12] All participants were informed about their average daily screen time during the baseline and read a note about benefits of reducing mobile usage. Participants in the C condition were then informed that they had no specific instructions or targets. Participants in the FI condition were told they could earn a screen time bonus of $2 per day during periods 1 and 2 (for 18 days), if they reduced their daily usage by 90 minutes[13]

[12] Assuming d = .2 (from RCT 2, period 1), ICC = .08, 80% power, and α = .05, the required sample was 85 per condition.
[13] We used a fixed 90-minute reduction target (vs. a percentage like 30%) to enable comparability between the AI and API condition and test robustness to target structure. The 90-minute target was chosen for its achievability and consistency with the first two RCTs.

from their baseline, up to a maximum of \$36. In the AI condition, participants were given the same daily target and incentive information as in the FI condition but were told that the bonus period started in period 2, with a maximum payment of \$24:

> *"**Starting in one week (date)** and continuing for the **next 12 days (until date)**, you can earn money for reducing your mobile screen time by **90 minutes** from your baseline, which is your average screen time from the first week of the study (**X hours Y minutes). For** each of the 12 days (dates) that your mobile screen time is **below your target** (TARGET: X' hours Y' minutes) you will receive **\$2**.* […] *You can earn a **maximum amount of \$24** (i.e., \$2 per day for 12 days), if your mobile screen time is below your target each day."*

In the API condition, participants were also told they had been selected for a screen time bonus, starting in period 2. However, they were not given a specific daily target. Instead, they were incentivized at a rate of \$16 per hour for any reductions in Period 2 usage relative to their baseline, up to a maximum of \$24 for reducing usage by 1.5 hours or more. The \$16/hour rate was chosen to match the maximum payment in the AI condition:

> *"**Starting in one week (date)** and continuing for the **next 12 days (until date),** you will earn \$16 for every hour you reduce your **average screen time** relative to your baseline mobile usage of **X hours Y minutes**, up to a maximum payout of \$24."*

We then gave several examples of how much participants could earn for reducing by a certain number of minutes on average (e.g., API condition: "*If you reduce your mobile screen time during the bonus period on average **by 30 minutes** from your baseline, you will earn **\$8**"*). See Web Appendix D, Table W14 for the instruction wording in all conditions. We then asked participants in the API condition by how much they would want to reduce their average mobile usage. The average (M = 96.09, SD = 92.93) was not different from the 90-minute reduction target in the other treatment conditions; t(101 = .663, p < .509, d = .066).

Participants in all treatment conditions answered questions probing their understanding of the instructions (baseline and target usage, bonus start and amount) and rated the attractiveness of the incentive and difficulty of the task. There was no difference in perceived difficulty between the treatment conditions ($F(2, 283) = 1.37$, $p = .256$, $\eta^2 = .009$), but in the API condition the incentive was rated more attractive than in the AI and FI condition ($F(2, 283) = 22.25$, $p < .001$, $\eta^2 = .136$), possibly because they were given an hourly pay rate ($16 vs. $2 per day), a higher numerical value that may have led to a framing effect. After the incentive period, participants entered a one-week post-treatment period. They were told that there were no further incentives and had to acknowledge this.

Survey measures

Our intervention may prime a pre-existing goal among participants to reduce their smartphone usage, potentially leading to a pre-emptive usage reduction. To explore this alternative explanation, we included two goal salience measures (rating and ranking) for all participants, both at baseline and after they received the period 1 instructions. Since both items showed similar results, we only present one of them here (*"How important is the goal of reducing your smartphone usage to you?"* 1 – 7 Likert Scale).

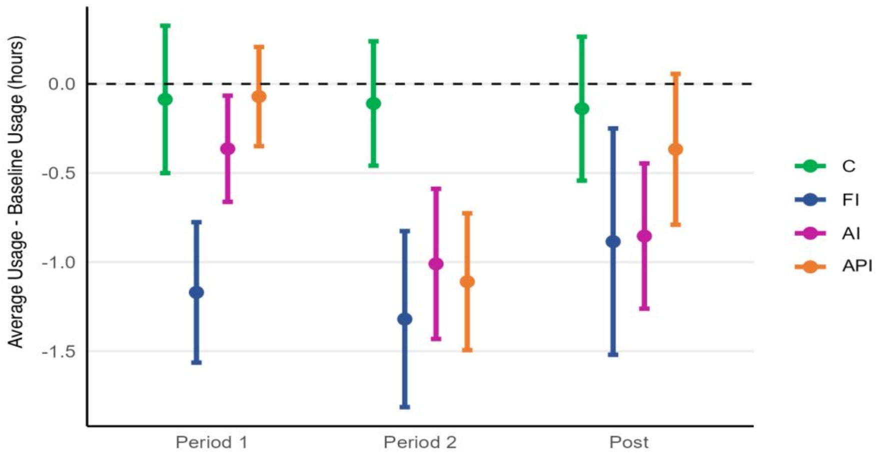
To better understand the motivations behind pre-emptive behavior changes, we included a set of items in survey 4 (administered after period 1) for participants in the AI and API condition. Specifically, participants rated (1 – 5 Likert scale) the extent to which several reasons explained changes in their smartphone usage during period 1. These included: 1) practice (testing whether they were capable of reducing usage before the bonus period), 2) forward-looking habit formation (reducing early to make it easier to sustain lower usage in the bonus period), 3) satiation (increasing usage beforehand to feel less cravings during the bonus period), and 4) myopia (no need to prepare for the bonus period). For exact item wording, see Web Appendix D "*Process measures in the AI and API condition.*"

We also assessed potential substitution effects by asking participants to estimate the percentage of their overall screen time attributed to different devices (smartphone, laptop/PC, TV, tablet, other) before and after the study. For details, see Web Appendix D: *"Device Substitution."* Finally, we included measures of life satisfaction (Cantril, 1965), subjective wellbeing, smartphone addiction (both from Allcott, Gentzkow, and Song (2022)), and social comparison at baseline, after the incentive and post-treatment period.

RCT 3: Results

Baseline usage (p = .857) and demographics did not differ across conditions (Web Appendix D, Table W14). Figure 5 shows the average usage reduction (from baseline) during all periods in each condition (see Web Appendix D: Figure W15 for day-level plots). The 95% CI is based on standard errors clustered at subject level.

Figure 5: Average Usage Reduction from Baseline Across Conditions and Periods.



We test the hypotheses using difference-in-differences OLS regression. We regress mobile usage on (1) the treatment variables (AI is the base, C, API, and FI are dummies); (2) the period dummies which capture change in usage across each period (period 0 or baseline is the base); and (3) the interaction between treatment and period dummies to capture change in usage from baseline in the particular treatment condition. Standard errors are clustered at the subject level to account for potential within-subject correlation. We use the AI condition as base to compare it to other treatment conditions, particularly the API condition. Regression estimates are in Table 2. Model 2 adds controls (age, gender, and operating system).

The treatment dummies are not significant, indicating no baseline differences. The period dummies are significant. Consistent with our hypothesis, the AI condition reduced their usage pre-emptively in period 1 compared to their baseline (p = .017) and continued having a lower usage in period 2 ($p < .001$) and the post-treatment ($p < .001$). In contrast, the API condition did not reduce usage pre-emptively in period 1 ($p = .617$). They reduced their usage during period 2 ($p < .001$), and their post-treatment reduction was marginally significant ($p = .091$). The FI condition significantly reduced their usage during the incentivized periods 1 and 2 as well as in the post-treatment period.

Table 6: AI Condition Reduces Usage in Periods 1, 2 and Post-treatment.

	Dependent Variable: Mobile Usage (hours)			
	Model 1	p-value	Model 2	p-value
API	.250 (.429)	.561	.259 (.427)	.545
Control	.386 (.470)	.412	.401 (.474)	.398
FI	.324 (.493)	.512	.321 (.492)	.515
Period 1	-.364 (.152)	**.017**	-.361 (.153)	**.018**
Period 2	-1.010 (.215)	**.000**	-1.003 (.214)	**.000**
Period 3	-.854 (.208)	**.000**	-.843 (.207)	**.000**
Age			.012 (.011)	.248
Gender (Male)			-.086 (.331)	.795
Operating System (iOS)			-.150 (.355)	.673
API x Period 1	.292 (.208)	.161	.290 (.208)	.164
Control x Period 1	.276(.261)	.290	.264 (.261)	.312
FI x Period 1	-.815 (.252)	**.002**	.814 (.252)	**.002**
API x Period 2	-.100 (.291)	.730	-.107 (-.290)	.712
Control x Period 2	.900 (.279)	**.002**	.894 (.278)	**.002**
FI x Period 2	-.316 (.331)	.340	-.319 (.330)	.334
API x Period 3	.487 (.300)	.105	.474 (.298)	.112
Control x Period 3	.715 (.293)	**.015**	.707 (.292)	**.016**
FI x Period 3	-.032 (.385)	.935	-.040 (.384)	.918
Constant	7.732 (.294)	.000	8.384 (.668)	.000
Observations	12,583		12,583	
R^2	.019		.022	
Adjusted R^2	.018		.020	
Residual Std. Error	3.847 (df = 12567)		3.843 (df = 12564)	
F Statistic	16.562 (df = 15; 12567)		15.470 (df = 18; 12564)	

Notes: p-values < .05 in bold.

While the AI condition pre-emptively reduced their usage in period 1 compared to their baseline, this reduction was not significant compared to the C condition (p = .290, d = .128) which also reduced their usage from baseline, and when compared to the API condition (p = .161, d = .106). The FI condition had a lower usage than all other conditions in period 1 (p's < .01, FI vs. C: d = .395). During period 2, the AI, API and FI conditions significantly reduced their usage compared to the C condition (p's < .01, FI vs. C: d = .416, AI vs. C: d = .350, API vs. C: d = .460). During the post-treatment, the AI condition significantly reduced their usage compared to the C condition (p = .016, d = .274) while the FI condition reduced usage marginally compared to the C condition (p = .052, d = .222). The API condition did not differ from the C condition in the post-treatment (p = .432, d = .175).

We find consistent results with target achievement as a binary dependent variable (i.e., reduction ≥ 90 minutes vs. not). Web Appendix D: Table W17 reports the regression results. The AI condition had higher target achievement in period 1, period 2 and the post-treatment compared to their baseline. In addition, in period 1, the AI condition had a marginally higher target achievement than the C condition (p = .07).

Heavy users showed higher pre-emptive reduction in the AI condition

As in our previous RCTs, heavy users with higher baseline usage showed a stronger pre-emptive reduction in period 1 (b = -.05, p < .001). Heavy users also felt more addicted to their smartphone at baseline (p < .001), with addiction levels generally being higher for females (p = .003) and younger participants (r = -.269 p < .001). However, age, gender, and baseline willingness to reduce did not predict the period 1 reduction in the AI condition (Web Appendix D: Figure W16 and Table W21).

Participants in treatment conditions reduced social media usage

We explored if participants reduced social media usage during the treatment periods, as a pilot study identified this as the app category consumers most wanted to cut back on. At baseline, participants had an average social media usage of approximately 20 hours

(M = 19.85, SD = 17.69) per week, accounting for about 37% of total screen time. There were no baseline differences between conditions. Higher social media usage was associated with younger age, female gender and having an Android phone.

Participants in the treatment conditions reduced their average weekly social media usage by 1.95 hours ($p = .02$) during periods 1 and 2, and 2.26 hours ($p = .019$) during the post-treatment compared to the baseline. There were no differences between the treatment conditions, indicating no differential effect of the AI, FI or API condition on social media reductions. The C condition showed no change in social media usage across periods. We report on social media usage in Web Appendix D: Table W18.

These results were consistent with participants' self-reported social media usage. Those in the treatment conditions reported greater reductions in social media and messaging app usage compared to the C condition ($F(3, 358) = 6.30$, $p < .001$ $\eta^2 = .050$). For further discussion of self-reported social media use (vs. entertainment, productivity, and gaming apps), refer to Web Appendix D: "*Self-reported Social Media Goals and Usage Changes*."

Adjacent complementarity: treatment period reduction predicts post-treatment reduction

As in our previous RCTs, we employed instrumental variable regression to assess whether reductions in periods 1 and 2 predicted post-treatment reductions. Our findings again indicated adjacent complementarity, showing that reductions in both period 1 (marginally at $p = .089$) and period 2 ($p = .006$) were associated with decreased usage in the post-treatment period in the AI condition vis-à-vis C condition (Web Appendix D: Table W19-20). Furthermore, the pre-emptive reduction in period 1 predicted period 2 usage ($p < .001$).

Exploring reasons for pre-emptive reductions

We calculated a goal salience difference score by subtracting the responses given after the treatment instructions from the responses given during the baseline survey (M = 1.44, SD = 1.52). There was a significant difference between conditions

($F(3, 366) = 19.12$, $p < .001$, $\eta^2 = .135$). Post-hoc comparisons with Bonferroni correction showed that goal salience increased more in the treatment conditions than in the C condition (all *p*'s < .001). Moreover, goal salience in the API condition increased more than in the AI condition ($p < .001$). However, despite showing the strongest increase in goal salience, the API condition did not reduce their usage pre-emptively. The goal salience ranking measure led to similar results (Web Appendix D: Table W22-23). Hence, an increase in goal salience cannot fully explain why participants in the AI condition reduced their usage pre-emptively compared to their baseline, while participants in the API condition did not.

After the anticipation period, we surveyed participants in the AI and API conditions with a set of questions aimed at understanding the usage patterns during period 1. Specifically, we asked them to rate several reasons for either maintaining or changing their usage in Period 1. We then examined whether agreement with these reasons predicted Period 1 usage using an OLS regression. Our results revealed that different explanations were associated with a reduction in the AI and API condition (Web Appendix D: Table W24).

In the AI condition, agreement with the statement capturing *forward-looking habit formation* was significantly associated with lower Period 1 usage (b = -.479, $p = .006$), while *practice* was not ($p = .330$). In the API condition, we find the opposite. Higher agreement with the *practice* statement was significantly associated with a lower period 1 usage (b = -.297, $p = .023$), while *forward-looking habit formation* was not ($p = .273$). In both conditions, agreement with the *satiation* explanation was associated with higher period 1 usage, that is less reduction (AI: b = .544, $p = .001$; API: b = .289, $p = .019$). Surprisingly, we also find that in both conditions, individuals who felt no need to prepare beforehand, had a lower period 1 usage (AI: b = -.251, $p = .048$; API: b = -.293, $p = .012$).[14] In summary, different explanations

[14] This might be due to the item wording capturing easiness of reduction, rather than myopia per se: *"I can easily lower my mobile usage to achieve the desired reduction in the bonus period – there was no need to prepare beforehand."*

appear to have motivated pre-emptive reductions in the AI and API conditions.

Substitution with other devices or offline activities?

We asked participants about the percentage of time they spend on different devices (Smartphone, Laptop/PC, TV, Tablet, Other) in the baseline and post-treatment survey. While these results are self-reported, we found no evidence of large-scale substitution. On the contrary, the percentage of screen time attributed to smartphone use showed a slight increase (M_{bsl} = 51.93, SD = 21.92 vs. M_{post} = 53.72, SD = 24.27, t(362) = -2.39, p = .017, d = .126). No significant changes were observed for laptop, TV, or tablet usage, nor did the condition have a significant impact. We also asked participants whether they had spent more or less time on various activities over the past four weeks (1 = a lot less time, 5 = a lot more time). Participants reported spending more time with family and friends (M = 3.23, SD = 0.76, t(361) = 5.87, p < .001, d = .308) and engaging in non-screen activities (M = 3.38, SD = 0.85, t(361) = 8.50, p < .001, d = .447), significantly above the scale midpoint 3 = same amount of time). Overall, substitution seems to be small, and participants reported spending more time socializing or engaging in non-screen activities (Web Appendix D: *"Device Substitution"*). Finally, reductions in period 1 (p = .006) and period 2 (p <.001) significantly predicted a reduction in perceived smartphone addiction, however, they were not associated with changes in life satisfaction or subjective wellbeing (Web Appendix D: Tables W25-28).

Model Fitting Results

To explore the heterogeneity of forward-looking behavior and habit formation at the individual level, we fitted the habit formation and satiation model to the data.

Equations

We solved for the optimal consumption under both habit formation and satiation. The optimal period 1 and post-treatment consumption in the AI condition under habit formation is given by Equations W3, W2 (respectively) and satiation by Equations W6, W5 (respectively) in Web Appendix A. To estimate the parameters of the habit formation and satiation model, we reformulate the equations to match our experimental set-up. We focus on RCTs 2 and 3 for our estimation as we have more robust post-treatment data. Our experimental periods in RCTs 2 and 3 are not of the same length. To maintain consistency with our model, we consider equal-sized time periods of six days for our estimation. From the perspective of our model, RCTs 2 and 3 span five time periods: *t = 0* (baseline), *t = 1* (period 1), *t = 2* (first half of period 2), *t = 3* (second half of period 2), and *t = 4* (post-treatment).[15] In the optimal consumptions equations for habit formation, we indicate $\frac{b_2}{4b_1}$ by H, which represents the strength of habit formation or adjacent complementarity (or how much the stock of accumulated consumption affects current consumption).[16] For identifying γ (habituation retention factor), γ' (satiation retention factor), and α (projection bias), we assume the discount factor $\delta = 1$ (as the period length is only 6 days). But for robustness, we also estimate the parameters assuming $\delta = 0.9$. The reformulated optimal consumption equations in period 1 and post-treatment are as follows. For a participant i in day j of time period t,

Under habit formation:

- Period 1: $c_{1Aij} = \frac{I_i}{2} + Hk_{1Ai} + (1-\alpha)H\gamma c_{2Ai} + (1-\alpha)\gamma^2 Hc_{3Ai} + (1-\alpha)\gamma^3 Hc_{4Ai} + \epsilon_{ij}$ (3)
- Post-treatment: $c_{4Aij} = \frac{I_i}{2} + H \times (\gamma c_{3Ai} + \gamma^2 c_{2Ai} + \gamma^3(c_{1Ai} + k_{1Ai})) + \epsilon_{ij}$ (4)

Under satiation:

- Period 1: $c_{1Aij} = \frac{I_i}{2} - \frac{(s_{1Ai} + (1-\alpha_s)(\delta\gamma' c_{2Ai} + \delta^2\gamma'^2 c_{3Ai} + \delta^3\gamma'^3 c_{4Ai}))}{2} + \epsilon'_{ij}$ (5)
- Post-treatment: $c_{4Aij} = \frac{I_i}{2} - \frac{\gamma' c_{3Ai} + \gamma'^2 c_{2Ai} + \gamma'^3(c_{1Ai} + s_{1Ai})}{2} + \epsilon'_{ij}$ (6)

where c_{Ait} is the average consumption of AI condition participant i in period t. ϵ_{ij} and ϵ'_{ij} are

[15] When there are more than six days in a given period, we use the last six days of period 1, the first twelve days of period 2, and the first six days of post-treatment. However, the results remain unchanged when using alternative days within in the same period.

[16] We cannot separately estimate diminishing marginal utility (b_1) and adjacent complementarity (b_2). We can only estimate them together. When $H > 0$, adjacent complementarity is implied under the assumption that the utility is concave.

zero mean, normally distributed noise, capturing daily deviations from optimal consumption in period 1 and post-treatment due to idiosyncratic factors. To minimize the effect of such random daily variations, we study the effect of habit stock accumulation between adjacent periods rather than adjacent days within a specific period. As the period 0 habit stock is unobserved, we initialize period 1 habit stock based on the observed period 0 (baseline) usage.

We discuss how Equations 3 to 6 help estimate the parameters of the habit formation and satiation model. We assume the budget I to be the maximum usage of a participant over the entire experimental period. For the habit formation model, three parameters must be identified: adjacent complementarity (H),habit retention (γ), and projection bias with respect to habit formation (α). In Equation 4, usage during the second half of the treatment period (c_{3A}) affects post-treatment usage (c_{4A}) through both habit retention (γ) and adjacent complementarity (H). Usage in the first half of the treatment period (c_{2A}) and in the pre-treatment period (c_{1A}) also influences post-treatment usage, but in a nonlinear way—c_{2A} is mediated by H and γ^2, while c_{1A} is mediated by H and γ^3. The degree to which post-treatment usage (c_{4A}) remains low compared to the usage in the second half of the treatment period reflects the stickiness of the treatment effect and identifies the combined role of habit retention and adjacent complementarity (H, γ). Comparing the stickiness of post-treatment usage relative to the second half of the treatment period, with the stickiness relative to the first half of the treatment period and pre-treatment period allows for identification of habit retention γ. Finally, projection bias with respect to habit formation (α) is identified by comparing pre-emptive reductions in usage during period 1 with the reductions observed during the treatment (period 2) and subsequent periods.

For the satiation model, two parameters must be identified: the satiation retention factor (γ') and projection bias with respect to satiation (α_s). The extent to which post-treatment usage increases (c_{4A}) compared to usage during the treatment and pre-treatment periods helps identify the satiation retention factor γ' (Eq. 6). An increase in usage during period 1, driven by anticipation of reduced usage in period 2 (i.e., during treatment), helps identify projection bias with respect to satiation (α_s) (Eq. 5).

We estimate the model with the maximum likelihood technique (Web Appendix E). We minimize the sum of squares between the empirically observed and theoretically predicted consumption for each participant based on Eqs. 3 - 4 and 5 - 6. The observed consumption $\mathrm{c}_{2\mathrm{A}}$ incorporates the implied probability (p_t) of meeting the target and thus the effect of incentives.

Estimation Results

When comparing the fit of both models at the individual level, the habit formation model fits the data better with a lower Akaike Information Criterion (AIC) for 94% of participants in RCT 2 and 91% in RCT 3. In contrast, the satiation model is a better fit for only 6% of participants in RCT 2 and 9% in RCT 3. The estimated habit formation and satiation parameters for each participant in RCTs 2 and 3 are listed in Web Appendix E: Tables W29-30. The average AIC of the habit formation model was significantly lower than of the satiation model in both RCTs ($p < .01$).

While habit formation was dominant, there was considerable heterogeneity in projection bias. The average projection bias under habit formation was .54 in both RCTs. It was significantly higher than zero and lower than one (both p's < .01), suggesting that participants were not myopic with respect to habit formation and displayed considerable forward-looking tendencies. See Table 7 for the parameters.

We further explore heterogeneity in projection bias under habit formation. 37.5% of participants in RCT 2 and 36.4% in RCT 3 had a projection bias below .5, indicating behavior more consistent with rational addiction than myopia. About 52% of participants in both RCTs were weakly forward-looking, while the remaining 12% were purely myopic (α

close to 1) with respect to habit formation (Web Appendix E: Figure W17-18). Estimating the model at the aggregate level yielded similar results (Web Appendix E: Tables W32-33).

Table 7: Individual Level Estimates of the Habit Formation Model (AI Condition).

Habit Formation Model	**RCT 2**		**RCT 3**	
	Mean (SE)	Median	Mean (SE)	Median
H (Adjacent complementarity)	.17 (.034)	.11	.2 (.016)	.16
$\boldsymbol{\gamma}$ (Habituation retention factor)	.41 (.07)	.37	.42 (.03)	.36
$\boldsymbol{\alpha}$ (Projection bias)	.54 (.07)	.78	.54 (.03)	.61
SSE (Sum of square errors)	14.37		45.38	
AIC (Akaike information)	30.11		33.01	

We also test the robustness of our estimates. First, we replicate the results using an alternative assumption of $\delta = 0.9$. While this led to a slightly lower estimate of projection bias (or more forward-looking behavior), the overall results remain consistent (Web Appendix E: Table W34). Additionally, we categorized participants based on how attractive they found the monetary incentive. Among those who rated it ≥ 4 on a 7-point scale (75% of participants, Web Appendix E: Figure W19), indicating high attractiveness to reduce usage, we find consistent results (Web Appendix E: Table W35).

In addition, we tested a different model specification. We built a model that allows for habit formation and satiation depending on the sign of H (assuming that b_2 in Equation 1 can take a positive or negative sign). We assumed both habit formation and satiation to affect total utility in a similar way. With this formulation, participants exhibited habit formation both at the aggregate (i.e., H > 0, $p < .001$) and individual level (H > 0 for 84% of participants; Web Appendix E: Table W36). For model estimates in the API condition, see Web Appendix E: Table W37.

General Discussion

In this research, we theoretically and empirically explored smartphone usage before, during and after a behavior change intervention aimed at reducing consumption. We disentangle the predictions of habit formation and satiation models under varying projection bias. Across three RCTs, we find that, on average, consumers appeared to behave more in line with forward-looking habit formation, lowering smartphone usage pre-emptively relative to their baseline levels, before receiving incentives. While participants in RCTs 1 and 2 reduced usage relative to the control condition, the effect was weaker in RCT 3, where the control condition also reduced their usage during period 1. This may reflect a combination of increased awareness due to self-tracking and receiving information about benefits of reducing screen time. We therefore interpret the pre-emptive reduction in RCT 3 more cautiously.

In general, our results are more consistent with the predictions of rational addiction theory rather than myopic habit formation or satiation (although they do not constitute definite proof). Nearly 35% of participants had a low projection bias (< 50%) and demonstrated behavioral patterns aligned with what we would call "rational addicts." On average, consumers pre-emptively reduced their mobile screen time by approximately half an hour per day, which is more than 30% of the reduction of consumers that were actually paid. These findings suggest that offering incentives during the treatment period, may lead to lower usage in the un-incentivized anticipation and post-treatment periods. Consistent with adjacent complementarity, consumers who reduced their usage more frequently during the anticipation and treatment period, sustain a lower screen time for at least some days after the intervention.

We identified considerable heterogeneity. As predicted by our model, heavy mobile users showed stronger pre-emptive reductions, suggesting that anticipated treatments may be especially effective for those most in need of curbing screen time, since they benefit from additional time to adjust their behavior gradually. This is interesting, as one might instead expect heavy users to be most resistant to change. In addition, consumers with strong beliefs in future goal achievement were more likely to reduce consumption pre-emptively.

We replicate the results in two distinct samples: a young, international student population and a representative sample of US residents. While age and gender did not moderate the effects in our studies, future research could explore how people from different cultural backgrounds with distinct mobile usage patterns (e.g., non-WEIRD samples) or even younger age groups (e.g., children) respond to anticipated treatments.

We contribute to literature that has used monetary incentives to inculcate beneficial habits (Acland and Levy 2015; Charness and Gneezy 2009; Loewenstein, Price, and Volpp 2016;) by demonstrating the conditions under which behavior change can be initiated even before incentives are paid, potentially offering a more cost-effective approach. First, it is crucial that the consumption is more habit-forming than satiating (as with mobile usage), since otherwise, anticipating the intervention could backfire by increasing consumption.

Second, building on Allcott, Gentzkow, and Song (2022), we test a boundary condition that affects the pre-emptive usage reduction. Daily consumption targets (AI condition) seem to be more effective than proportional period-wise incentives (API condition) in prompting pre-emptive reductions because API participants can adjust consumption dynamically during the treatment period, eliminating the need for early reductions. Finally, consumers hold accurate lay beliefs about the benefits of initiating behavior change pre-emptively and prefer a nine-day anticipation period for gradual adjustment. This approach may lead to more sustained behavior change (Somasundaram, Koch, and Lim 2023).

While our findings generally align with rational addiction theory, alternative mechanisms might contribute to the observed reduction. We examined goal priming, and practice effects. Although these are plausible explanations, our data show that pre-emptive reductions were not clustered around the days immediately before the incentive period, and self-reported practice did not predict pre-emptive reductions in the AI condition. Goal salience increased across all treatment conditions, but did not predict pre-emptive usage reductions, suggesting that it alone cannot account for the effect.

Our findings have practical implication. Cost-effective, scalable solutions to excessive mobile consumption are needed since concerns about digital addiction have been growing (Alter 2017; Zimmermann and Somasundaram 2024). As with the general population, 89% of our participants wanted to reduce their mobile usage and 76-87% tried to do so in the past. Given that it would require a significant compensation for the average consumer to forgo a single social media platform (Brynjolfsson, Collis, and Eggers 2019; Mosquera et al. 2020), adding an anticipation period could potentially be an attractive modification of time-limit interventions as it might be less costly and heavy users benefit most. Nevertheless, the feasibility and cost-effectiveness of implementing such interventions remain open questions which future research could build upon.

Policymakers and consumer advocacy groups could take our findings into account when providing recommendations on how to reduce mobile usage. For instance, the American Academy of Pediatrics (2019) has recently announced recommendations to place consistent limits on children's screen time to avoid the risk of sleep problems and negative effects on school performance. While many pieces of advice exist, few provide evidence-based recommendations on how this could be achieved effectively. Parents seeking effective ways to reduce their children's screen time and companies that offer solutions for "digital addiction" could possibly benefit from implementing anticipation periods.

Beyond smartphones, our findings suggest broader applications for addressing overuse and dependency on digital technology (e.g., gaming) and other consumer behaviors, if they are habit forming and consumers are non-myopic with respect to habit formation. For example, financial incentives have been used to reduce smoking, with participants paid for cessation over a three-month period (Volpp et al. 2009). However, consumption patterns often reverted once incentives were removed (Wood and Neal 2016). Incentivizing behavior at a finer temporal granularity (e.g., daily or weekly) may strengthen habits, with recent evidence showing promising results (Kendzor et al., 2024).

We conclude with caveats and avenues for future research. We focused on overall mobile screen time because it did not require the development of a tracking application that participants needed to download. This allowed us to collect data from all smartphone types (it was not restricted solely to Android devices due to privacy restrictions on iOS). Overall screen time is a stable measure of mobile usage, unaffected by trends or changes in specific apps which mitigates potential issues of within-device substitution (Collis and Eggers 2022). Our data showed that participants in the treatment conditions significantly reduced their social media usage over the treatment periods (compared to the C condition)—the app category they ideally wanted to reduce. However, our method did not allow us to conduct analyses of specific app usage (e.g., TikTok) which future research could explore.

When scaling our results to the general population, it is important to note that our sample consisted of individuals with relatively high smartphone usage and high willingness to reduce usage. This sample may be subject to self-selection bias. Although we did not disclose the study's purpose during recruitment, it is possible that individuals who believed they could perform well were more likely to enroll. Similarly, during the treatment periods, participants who felt less confident in their ability may more likely to drop out (although attrition was low and balanced across conditions during the treatment periods). We also emphasize that the observed period 1 and post-treatment effects should be interpreted as suggestive evidence, given the nonsignificant (AI vs. C) contrast in period 1 of RCT 3, potential attrition effects, and external shocks in earlier studies. The persistence of reduced usage, feasibility and cost-effectiveness of our anticipation intervention in broader population remain open questions, which the future research could explore.

While our study focused on incentivizing usage reductions, future research could also explore the possibility of penalizing excessive usage through pricing mechanisms and shed light on differences between willingness to accept for giving up on excessive consumption versus willingness to pay for excessive consumption and pre-emptive behavior change. Future research could further test the interplay between adjacent substitutability and complementarity over different time frames using more nuanced models (e.g., Landry's (2019) model predicts short-term substitutability and long-term complementarity). Our model is based on a finite time horizon, which enables us to derive closed-form solutions for the optimal consumption equations. This offers analytical tractability but comes with a limitation: it overlooks forward-looking behavior in the final period. As a result, it may underestimate habit retention, since individuals anticipating the end of the incentive scheme might increase usage post-treatment because they are aware that no further incentives follow. Finally, while most of our participants only possessed one phone (92-96%) and we excluded participants with more than one phone in RCT 3, our experimental setting did not allow us to measure objectively if consumers substituted their mobile usage with other digital devices, such as tablets or laptops. Our survey results indicated that across-device substitution was unlikely to be the main strategy to reduce smartphone usage. Nevertheless, future research could explore substitution effects further.

References

Acland, D. and Matthew R. Levy (2015), "Naiveté, projection bias, and habit formation in gym attendance," *Management Science*, 61 (1), 146–60.

Allcott, Hunt, Matthew Gentzkow, and Lena Song (2022), "Digital Addiction," *American Economic Review*, 112 (7), 2424–63.

Alter, Adam (2017), *Irresistible: The rise of addictive technology and the business of keeping us hooked*, Penguin.

American Academy of Pediatrics. (2019). *American Academy of Pediatrics Announces New Recommendations for Childrens Media Use*. https://www.aap.org/en-us/about-the-aap/aap-press-room/Pages/American-Academy-of-Pediatrics-Announces-New-Recommendations-for-Childrens-Media-Use.aspx

Baltagi, Badi H. and James M. Griffin (2002), "Rational addiction to alcohol: panel data analysis of liquor consumption," *Health Economics*, 11 (6), 485–91.

Barwick J. Panle, Siyu Chen, Chao Fu, Teng Li (2025), "Digital Distractions with Peer Influence: The Impact of Mobile App Usage on Academic and Labor Market Outcomes," *The Quarterly Journal of Economics*, qjaf048.

Baucells, Manel and Rakesh K. Sarin (2007), "Satiation in discounted utility," *Operations research*, 55 (1), 170–81.

Baucells, Manel and Rakesh K. Sarin (2010), "Predicting utility under satiation and habit formation," *Management Science*, 56 (2), 286–301.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy (1994), "An Empirical Analysis of Cigarette Addiction," *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.

Berger, Jonah, Wendy W. Moe, and David A. Schweidel (2023), "What holds attention? Linguistic drivers of engagement," *Journal of Marketing*, 87 (5), 793–809.

Bright, Laura F., Susan Bardi Kleiser, and Stacy Landreth Grau (2015), "Too much Facebook? An exploratory examination of social media fatigue," *Computers in Human Behavior*, 44, 148–55.

Brynjolfsson, Erik, Avinash Collis, and Felix Eggers (2019), "Using massive online choice experiments to measure changes in well-being," *Proceedings of the National Academy of Sciences*, 116 (15), 7250–55.

Camerer, Colin F. and Xiaomin Li (2021), "Neural autopilot and context-sensitivity of habits," *Current Opinion in Behavioral Sciences*, 41, 185–90.

Cantril, Hadley (1965), *The pattern of human concerns*, New Brunswick: Rutgers University Press.

Charness, Gary and Uri Gneezy (2009), "Incentives to exercise," *Econometrica*, 77 (3), 909–31.

Collis, Avinash and Felix Eggers (2022), "Effects of restricting social media usage on wellbeing and performance: A randomized control trial among students," *Plos one*, 17 (8), e0272416.

Coombs, Clyde H. and George S. Avrunin (1977), "Single-peaked functions and the theory of preference," *Psychological review*, 84 (2), 216.

Fishbach, Ayelet and Ravi Dhar (2005), "Goals as Excuses or Guides: The Liberating Effect of Perceived Goal Progress on Choice," *Journal of Consumer Research*, 32 (3), 370–77.

Galak, Jeff, Joseph P. Redden, and Justin Kruger (2009), "Variety amnesia: Recalling past variety can accelerate recovery from satiation," *Journal of Consumer Research*, 36 (4), 575–84.

Gitnux (2025), "Technology Addiction Statistics," [available at https://gitnux.org/technology-addiction-statistics/].

Giunchiglia, Fausto, Mattia Zeni, Elisa Gobbi, Enrico Bignotti, and Ivano Bison (2018), "Mobile social media usage and academic performance," *Computers in Human Behavior*, 82, 177–85.

Gruber, Jonathan and Botond Köszegi (2001), "Is addiction 'rational'? Theory and evidence," *The Quarterly Journal of Economics*, 116 (4), 1261–1303.

Haenlein, Michael, Barak Libai, and Eitan Muller (2023), "Satiation and cross promotion: Selling and swapping users in mobile games," *International Journal of Research in Marketing*, 40 (2), 342–61.

Hetherington, Marion M., Linda M. Pirie, and Samantha Nabb (2002), "Stimulus satiation: effects of repeated exposure to foods on pleasantness and intake," *Appetite*, 38 (1), 19–28.

Hughes, Christian, Vanitha Swaminathan, and Gillian Brooks (2019), "Driving brand engagement through online social influencers: An empirical investigation of sponsored blogging campaigns," *Journal of Marketing*, 83 (5), 78–96.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol (2022), *Rational Habit Formation: Experimental Evidence from Handwashing in India*, American Economic Journal: Applied Economics, 14(1).

Iannaccone, Laurence R. (1986), "Addiction and satiation," *Economics Letters*, 21 (1), 95–99.

Kaye, Linda, Amy Orben, David A. Ellis, Simon C. Hunter, and Stephen Houghton (2020), "The conceptual and methodological mayhem of 'screen time,'" *International Journal of Environmental Research and Public Health*, 17 (10), 3661.

Kendzor, Darla E., Michael S. Businelle, Summer G. Frank-Pearce, Joseph J. Waring, Sixia Chen, Emily T. Hébert, others, and David W. Wetter (2024), "Financial incentives for smoking cessation among socioeconomically disadvantaged adults: a randomized clinical trial," *JAMA Network Open*, 7 (7), e2418821–e2418821.

Khan, Uzma and Ravi Dhar (2006), "Licensing effect in consumer choice," *Journal of Marketing Research*, 43 (2), 259–66.

Koopmans, Tjalling C. (1960), "Stationary ordinal utility and impatience," *Econometrica: Journal of the Econometric Society*, 287–309.

Krpan, Dario, Matteo M. Galizzi, and Paul Dolan (2019), "Looking at Spillovers in the Mirror: Making a Case for 'Behavioral Spillunders,'" *Frontiers in psychology*, 10, 1142.

Kwon, Hyeokkoo Eric, Hyunji So, Sang Pil Han, and Wonseok Oh (2016), "Excessive dependence on mobile social apps: A rational addiction perspective," *Information Systems Research*, 27 (4), 919–39.

Landry, Peter (2019), "Bad habits and the endogenous timing of urges," *The Review of Economic Studies*, 86 (2), 785–806.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin (2003), "Projection bias in predicting future utility," *the Quarterly Journal of economics*, 118 (4), 1209–48.

Loewenstein, George, Joseph Price, and Kevin Volpp (2016), "Habit formation in children: Evidence from incentives for healthy eating," *Journal of health economics*, 45, 47–54.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie (2020), "The economic effects of Facebook," *Experimental Economics*, 23 (2), 575–602.

Nevskaya, Yulia and Paulo Albuquerque (2019), "How Should Firms Manage Excessive Product Use? A Continuous-Time Demand Model to Test Reward Schedules, Notifications, and Time Limits," *Journal of Marketing Research*, 56 (3), 379–400.

Ofcom. (2020). *Adults' media use and attitudes report 2020.* https://www.ofcom.org.uk/research-and-data/media-literacy-research/adults/adults-media-use-and-attitudes

Pedersen, Jesper, Martin G. B. Rasmussen, Sarah O. Sørensen, Sofie R. Mortensen, Line G. Olesen, Søren Brage, others, and Anders Grøntved (2022), "Effects of limiting digital screen use on well-being, mood, and biomarkers of stress in adults," *Npj mental health research*, 1 (1), 14.

Pollak, Robert A. (1970), "Habit formation and dynamic demand functions," *Journal of political Economy*, 78 (4), 745–63.

Pollak, Robert A. (1976), "Habit formation and long-run utility functions," *Journal of Economic Theory*, 13 (2), 272–97.

Redden, Joseph P. and Kelly L. Haws (2013), "Healthy satiation: The role of decreasing desire in effective self-control," *Journal of Consumer Research*, 39 (5), 1100–1114.

Reimann, Martin and Shailendra P. Jain (2021), "Maladaptive Consumption: Definition, Theoretical Framework, and Research Propositions," *Journal of the Association for Consumer Research*, 6 (3), 307–14.

Samuelson, Paul A. (1937), "A note on measurement of utility," *The review of economic studies*, 4 (2), 155–61.

Schmidt-Persson, Jesper, Martin G. B. Rasmussen, Sarah O. Sørensen, and others (2024), "Screen Media Use and Mental Health of Children and Adolescents: A Secondary Analysis of a Randomized Clinical Trial," *JAMA Network Open*, 7 (7), e2419881.

Sim, Aaron Y., Li Ling Lee, and Bobby K. Cheon (2018), "When exercise does not pay: Counterproductive effects of impending exercise on energy intake among restrained eaters," *Appetite*, 123, 120–27.

Small, Dana M., Robert J. Zatorre, Alain Dagher, Alan C. Evans, and Marilyn Jones-Gotman (2001), "Changes in brain activity related to eating chocolate: from pleasure to aversion," *Brain*, 124 (9), 1720–33.

Somasundaram, Jeeva, Ingrid Koch, and Noah Lim (2023), "Raising the AC Temperature in the Tropics: One degree at a time," *Energy Economics*, 128, 107191.

Tromholt, Morten (2016), "The Facebook experiment: Quitting Facebook leads to higher levels of well-being," *Cyberpsychology, behavior, and social networking*, 19 (11), 661–66.

Twenge, Jean M., Jonathan Haidt, Thomas E. Joiner, and W. Keith Campbell (2020), "Underestimating digital media harm," *Nature Human Behaviour*, 4 (4), 346–48.

Urbszat, Dax, C. Peter Herman, and Janet Polivy (2002), "Eat, drink, and be merry, for tomorrow we diet: Effects of anticipated deprivation on food intake in restrained and unrestrained eaters," *Journal of abnormal psychology*, 111 (2), 396.

U.S. Senate Committee on the Judiciary (2024). Protecting children online. Retrieved from https://www.judiciary.senate.gov/protecting-children-online

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), 496–508.

Volpp, Kevin G., Andrea B. Troxel, Mark V. Pauly, Henry A. Glick, Andrea Puig, David A. Asch, others, and Jill DeGuzman (2009), "A randomized, controlled trial of financial incentives for smoking cessation," *N Engl J Med*, 360, 699–709.

Wang, Yang (2014), "Dynamic Implications of Subjective Expectations: Evidence from Adult Smokers," *American Economic Journal: Applied Economics*, 6 (1), 1–37.

Wathieu, Luc (1997), "Habits and the anomalies in intertemporal choice," *Management Science*, 43 (11), 1552–63.

Wells, Georgia, Jeff Horwitz, and Deepa Seetharaman (2021), "Facebook knows Instagram is toxic for teen girls, company documents show," *The Wall Street Journal*.

Wood, Wendy and David T. Neal (2016), "Healthy through habit: Interventions for initiating & maintaining health behavior change," *Behavioral Science & Policy*, 2 (1), 71–83.

Wood, Wendy and Dennis Rünger (2016), "Psychology of habit," *Annual review of psychology*, 67, 289–314.

Zimmermann, Laura and Michael Sobolev (2023), "Digital strategies for screen time reduction: A randomized field experiment," *Cyberpsychology, Behavior, and Social Networking*, 26 (1), 42–49.

Zimmermann, Laura and Jeeva Somasundaram (2024), "Maladaptive Smartphone Usage," in *Maladaptive Consumer Behavior*, Cham: Springer Nature Switzerland, 103–27.

Web Appendix:

Leveraging Rational Addiction Theory to Reduce Mobile Usage

Jeeva Somasundaram,
Associate Professor of Decision Sciences, IE Business School, Madrid, Spain,
jeeva.somasundaram@ie.edu

Laura Zimmermann,
Associate Professor of Marketing, IE Business School, Madrid, Spain,
laura.zimmermann@ie.edu

Duc Quang Pham,
Assistant Professor of Marketing, CUNEF Universidad, Madrid, Spain,
quangduc.pham@cunef.edu

Table of Contents

These materials have been supplied by the authors to aid in the understanding of their paper.

The AMA is sharing these materials at the request of the authors.

Web Appendix A: Derivation and Proofs of Optimal Consumption Under Habit Formation and Satiation Models

To prove the propositions, we first derive the optimal consumption plan for the different conditions in our core design (i.e., C, FI, and AI condition) for different periods under habit formation and satiation models. For simplicity, we assume all periods have equal length and there are T+1 periods. We indicate the baseline by $t = 0$, period 1 by $t = 1$, period 2 by $t = 2$, and post-treatment by $t = 3$. We derive the optimal consumption below and use it to prove the propositions presented in the following section.

Predictions of the Habit Formation Model with Projection Bias

Consider consumption of goods 1 and 2 in period t indicated by c_t and c'_t, respectively. Consumption of good 1 is habit-forming, while consumption of good 2 does not exhibit intertemporal complementarities. The maximization problem of a consumer is as follows. For every time period $t' = 0, \ldots, T$:

$$\boldsymbol{c}_{t'} = \arg\max_{(c_{t'},\ldots,c_T)} \alpha \times \left(\sum_{t=t',\ldots,T} \delta^{t-t'} \left(u(c_t, k_{t'}) + u(I - c_t) \right) \right)$$
$$+(1-\alpha) \times \left(\sum_{t=t',\ldots,T} \delta^{t-t'} \left(u(c_t, k_t) + u(I - c_t) \right) \right)$$

subject to $k_{t+1} = \gamma(k_t + \mathrm{c}_t)$. Below, we calculate the optimal consumption in the C, FI, and AI condition.

Optimal consumption in the C condition

We indicate the optimal consumption in the C condition for different periods by $c_0, \ldots, c_3$. The optimization equation is as follows:

$$\max_{(c_0,\ldots,c_3)} \sum_{t=0,\ldots,3} \delta^t \left((1-\alpha) u(c_t, k_t) + \alpha u(c_t, k_0) \right) + \delta^t u(I - c_t)$$

subject to $k_{t+1} = \gamma(k_t + \mathrm{c}_t)$.

We represent the first constraint by replacing $k_{t+1} = \gamma(k_t + c_t)$ in the maximization equation. We start by calculating the optimal consumption in the final period. The first order condition (FOC) for optimal consumption in period 3 (c_3) is as follows:

$$u'(c_3, k_3) = u'(I - c_3)$$

We assume a quadratic utility function $u(c_t, k_t) = -b_1 c_t^2 + b_2 c_t k_t + b_3 c_t$ as in Gruber and Koszegi (2001) and Allcott, Gentzkow, and Song (2022).[1] Therefore, the FOC becomes:

$$-2b_1 c_3 + b_2 k_3 + b_3 = -2b_1(I - c_3) + b_3$$

$$c_3 = \frac{2b_1 I + b_2 k_3}{4b_1}$$

The FOC for optimal consumption c_2 is as follows:

$$\delta^2[u'(c_2, k_2) - u'(I - c_2)] + \delta^3 \gamma(1-\alpha)\big[u'\big(c_3, \gamma(k_2 + c_2)\big)\big] = 0$$

$$c_2 = \frac{2b_1 I + b_2 k_2 + (1-\alpha)\gamma\delta b_2 c_3}{4b_1}$$

Similarly, we can obtain optimal consumption c_1 and c_0:

$$c_1 = \frac{2b_1 I + b_2 k_1 + (1-\alpha)(\gamma\delta b_2 c_2 + \gamma^2\delta^2 b_2 c_3)}{4b_1}$$

$$c_0 = \frac{2b_1 I + b_2 k_0 + (1-\alpha)(\gamma\delta b_2 c_1 + \gamma^2\delta^2 b_2 c_2 + \gamma^3\delta^3 b_2 c_3)}{4b_1}$$

Note that $c_3 - c_2 > 0$ when $k_2 < \frac{b_2\gamma c_2 - 0.5\times(1-\alpha)I\delta b_2\gamma - \frac{(1-\alpha)b_2\gamma c_2}{4b1}}{b_2 - b_2\gamma + \frac{b_2^2\gamma^2\delta(1-\alpha)}{4b1}}$ (W1)

Thus, when habit stock is lower than a specific value, we have an increasing consumption pattern (i.e., $c_2 < c_3$). We can similarly demonstrate this for the difference between other consumption levels (k_1 must be lower than a bound for $c_2 - c_1 > 0$, and so on).

[1] We can also have the direct and quadratic effect of habit stock in the equation (i.e., $u(c_t, k_t) = -b_1 c_t^2 + b_2 c_t k_t + b_3 c_t - b_4 k_t - b_5 k_t^2$) to capture internality as in Becker and Murphy (1988) and Loewenstein, O'Donoghue, and Rabin (2003), but this has no effect on our results. Our propositions hold even if we include these terms.

In other words, if habit stock is below a certain level, optimal consumption is increasing. This result is consistent with the literature (Becker and Murphy 1988; see Lemma 1 of Loewenstein, O'Donoghue, and Rabin 2003).

Optimal consumption in the FI condition

Let the optimal consumption of participants in the FI condition be $c_{0F}, \ldots, c_{TF}$. In the optimal consumption equation, participants choose whether to meet the target (i.e., reduce their usage to $(1-\beta)c_{0F}$) or not. Thus, in periods 1 and 2, consumption will either be $(1-\beta)c_0$ or the optimal consumption c_t, determined by the habit stock k_t (which is similar to the C condition). The choice between these two consumption levels is determined by probability p_t which depends on the net utility, which is a function of the incentive's attractiveness, and the difficulty of achieving the reduction given an individual's habit stock k_t.[2]

$$\max_{(c_{0F},\ldots,c_{3F})} u(c_{0F},k_{0F}) + \gamma\big((1-\alpha)u(c_{1F},k_{1F}) + \alpha u(c_{1F},k_{0F})\big)$$
$$+\gamma^2\big((1-\alpha)u(c_{2F},k_{2F}) + \alpha u(c_{2F},k_{0F})\big)+\gamma^3\big((1-\alpha)u(c_{3F},k_{2F}) + \alpha u(c_{3F},k_{0F})\big)$$
$$+\sum_{t=0,\ldots,3} \delta^t\, u(I-c_{tF})$$

The optimal consumption c_{0F} in the FI condition is identical to the optimal consumption c_0 in the C condition during the baseline period, as participants are randomly assigned and receive the FI instructions only at the end of period 0.

The optimal consumption during period 1 is given by $c_{1F} = (1-\beta)c_0$ with probability p_{1F}, and it is c_1 (optimal consumption in control) with probability $1-p_{1F}$. Note that with probability p_{1F}, a participant will not reduce their consumption level c_{1F} below $(1-\beta)c_0$ optimally when the habit stock is not too high (this occurs only when the consumption pattern is decreasing; see discussion below Eq. W1). Thus, the expected consumption in period 1 is:

[2] The probability p_t can be micro-founded by assuming a modified logit probability $p_t = \frac{2}{1+e^{-\lambda\eta}} - 1$. The probability $p_t \to 0$ when net utility of reducing and receiving incentives (η) is zero and $p_t \to 1$ when η is extremely positive.

$$c_{1F} = p_{1F}(1-\beta)c_0 + (1-p_{1F})c_1.$$

The probability p_{1F} is a function of habit stock k_1 and the perceived attractiveness of the incentives to reduce usage. When the incentive is attractive and the required reduction is less painful, then p_{1F} will be higher. Similarly, the optimal consumption during period 2 is given by $c_{2F} = (1-\beta)c_0$ with probability p_{2F}, and by c_2 with probability $1-p_{2F}$. The expected consumption in period 2 is:

$$c_{2F} = p_{2F}(1-\beta)c_0 + (1-p_{2F})c_{2F}.$$

The optimal consumption in the post-treatment is given by:

$$c_{3F} = \frac{2b_1 I + b_2 k_{3F}}{4b_1}$$

FI versus C condition

When the initial habit stock is small, optimal consumption $c_0, \ldots, c_3$ is increasing (see discussion below Eq. W1). When $p_{1F} > 0$ and $p_{2F} > 0$ (when net utility of reducing is positive), this implies that $c_{1F} < c_1$ and $c_{2F} < c_2$. Note that as $c_{1F} < c_1$ and $c_{2F} < c_2$, since $k_{1f} = k_1$ (as the baseline is identical), we get $k_{3F} < k_3$ and thus, $c_{3f} < c_3$.

Optimal consumption in the AI condition

Let the optimal consumption of participants in the AI condition be $c_{0A}, \ldots, c_{3A}$. A participant's optimal consumption is given by:

$$\begin{aligned}\underset{(c_{0A},\ldots,c_{3A})}{\text{Max}} \quad & u(c_{0A},k_{0A}) + (\delta(\alpha u(c_{1A},k_{1F}) + (1-\alpha)u(c_{1A},k_{0A})) + \delta^2(\alpha u(c_{2A},k_{2A}) \\ & + (1-\alpha)u(c_{2A},k_{0A})) + \delta^3(\alpha u(c_{3A},k_{3A}) + (1-\alpha)u(c_{3A},k_{0A})) \\ & + \sum_{t=0,\ldots,3} \delta^t \, u(I - c_{tA})\end{aligned}$$

The optimal consumption c_{0A} in the AI condition is identical to the optimal consumption c_0 in the C condition during the baseline as participants are randomly assigned and receive the AI instructions only at the end of period 0. The optimal consumption during period 3 is given by:

$$c_{3A} = \frac{2b_1 I + b_2 k_{3A}}{4b_1} \qquad \text{(W2)}$$

As in the FI condition, the optimal consumption during period 2 is given by $c_{2A} = (1-\beta)c_0$ with probability p_{2A}, and by c_2 with probability $1 - p_{2A}$. Again, as discussed under the FI condition, with probability p_{1F}, a participant will not reduce their consumption level c_{2A} optimally below $(1-\beta)c_0$ when the habit stock is not too high. The expected optimal consumption in period 2 is:

$$c_{2A} = p_{2A}(1-\beta)c_0 + (1-p_{2A})c_2.$$

The optimal consumption during period $1(c_{1A})$ is:

$$c_{1A} = \frac{2b_1 I + b_2 k_{1A} + (1-\alpha)(\gamma\delta b_2 c_{3A} + \gamma^2\delta^2 b_2 c_{2A})}{4b_1} \qquad \text{(W3)}$$

Substituting $c_{3A} = \frac{2b_1 I + b_2 k_{3A}}{4b_1}$ and $k_{3A} = \gamma c_{2A} + \gamma^2 c_{1A} + \gamma^3 k_{1A}$, we get:

$$c_{1A} = \frac{(8b_1^2 + (1-\alpha)2b_1 b_2 \delta\gamma)I + k_{1A}(4b_1 b_2 + (1-\alpha)\gamma^4\delta b_2) + c_{2A}(1-\alpha)(\gamma^2\delta b_2 + 4b_1 b_2\gamma^2\delta^2)}{16b_1^2 - \gamma^3\delta b_2(1-\alpha)}$$

AI versus C condition

If the habit stock k_1 is small, then optimal consumption in the C condition $c_0, \dots, c_3$ is increasing. This implies that $c_{2A} < c_2$, when $p_{2A} > 0$. The expression for c_1, when the expression for c_3 and k_3 is substituted, is:

$$c_1 = \frac{(8b_1^2 + (1-\alpha)2b_1 b_2 \delta\gamma)I + k_1(4b_1 b_2 + (1-\alpha)\gamma^4\delta b_2) + c_2(1-\alpha)(\gamma^2\delta b_2 + 4b_1 b_2\gamma^2\delta^2)}{16b_1^2 - \gamma^3\delta b_2(1-\alpha)}$$

Comparing the above expression to the expanded expression of c_{1A}, we observe that $k_1 = k_{1A}$ and $c_{2A} < c_2$, when $p_{2A} > 0$. Therefore, $c_{1A} < c_1$.

Pre-emptive reduction conditional on baseline usage

Additionally, we examine how the reduction in the AI condition $c_{0A} - c_{1A}$ changes depending on baseline consumption and the target. Substituting $c_{2A} = p_{2A}(1-\beta)c_0 + (1-p_{2A})c_2$ in the expression above, we obtain:

$$c_{0A} - c_{1A} = c_{0A} - \frac{(8b_1^2 + 2b_1b_2\delta\gamma)I + k_1(4b_1b_2 + (1-\alpha)\gamma^4\delta b_2) + (p_{2A}(1-\beta)c_{0A} + (1-p_{2A})c_2)(1-\alpha)(\gamma^2\delta b_2 + 4b_1b_2\gamma^2\delta^2)}{16b_1^2 - (1-\alpha)\gamma^3\delta b_2}$$

$$c_{0A} - c_{1A} = \frac{c_0\left(16b_1^2 - (1-\alpha)\gamma^3\delta b_2 - (1-\alpha)p_{2A}(1-\beta)(\gamma^2\delta b_2 + 4b_1b_2\gamma^2\delta^2)\right) - (8b_1^2 + 2b_1b_2\delta\gamma)I - k_1(4b_1b_2 + (1-\alpha)\gamma^4\delta b_2)) - (1-p_{2A})c_2(1-\alpha)(\gamma^2\delta b_2 + 4b_1b_2\gamma^2\delta^2)}{16b_1^2 - (1-\alpha)\gamma^3\delta b_2}$$

We observe that $\frac{\partial(c_{0A}-c_{1A})}{\partial\beta} > 0$.

We calculate:

$$\frac{c_{0A} - c_{1A}}{c_{0A}} = \frac{-(8b_1^2 + 2b_1b_2\delta\gamma)I - k_1(4b_1b_2 + (1-\alpha)\gamma^4\delta b_2))}{16b_1^2 - \gamma^3(1-\alpha)\delta b_2} c_0^{-1} + \frac{\left(16b_1^2 - (1-\alpha)\gamma^3\delta b_2 - (1-\alpha)(1-\beta)(\gamma^2\delta b_2 + 4b_1b_2\gamma^2\delta^2)\right)}{16b_1^2 - \gamma^3(1-\alpha)\delta b_2} - \frac{(1-p_{2A})(c_2/c_{0A})(1-\alpha)(\gamma^2\delta b_2 + 4b_1b_2\gamma^2\delta^2)}{16b_1^2 - (1-\alpha)\gamma^3\delta b_2}$$

$\frac{\partial}{\partial c_0}\frac{c_{0A}-c_{1A}}{c_{0A}} > 0$ if $16b_1^2 - \gamma^3(1-\alpha)\delta b_2 > 0$. The expression $16b_1^2 - (1-\alpha)\gamma^3\delta b_2$ is positive, otherwise the optimal consumption c_{1A} would be negative. Therefore, $\frac{\partial}{\partial c_0}\frac{c_{0A}-c_{1A}}{c_{0A}} > 0$. This means that AI participants with larger baseline usage should reduce their usage more during period 1 in anticipation of the target in period 2.

Predictions of Satiation with Projection Bias

We incorporate projection bias into Baucells and Sarin's satiation model (2007, 2010). We formulate the maximization problem as follows. For every time period $t' = 0, \ldots, T$,

$$\max_{(c_{t'},\ldots c_T)} \sum_{t=t',\ldots,T} \begin{matrix} \delta^t \alpha_s \Big(\big(v(c_t + s_{t'}) - v(s_{t'})\big)\Big) + (1-\alpha_s)\Big(\big(v(c_t + s_t) - v(s_t)\big)\Big) \\ + \delta^t v(I - c_t) \end{matrix} \tag{W4}$$

subject to $\mathrm{s}_{t+1} = \gamma'(s_t + \mathrm{c}_t)$. We assume a quadratic utility function for v, that is $v(c_t) = -b'_1 c_t^2 + b'_2 c_t$, with $b'_1, b'_2 > 0$ and $c_t < 2b'_1/b'_2$.

Optimal consumption in the C condition

We begin with the FOC for consumption in period 3 (c_3):

$$\alpha_s\, v'\big((c_3 + s_3)\big) + (1-\alpha_s) v'(c_3 + s_1) = v'(I - c_3)$$

Substituting a quadratic utility, we obtain:

$$c_3 = \frac{I - s_3}{2}$$

The FOC for consumption in period 2 (c_2) is as follows:

$$v'\big((c_2 + s_2)\big) + \alpha_s \gamma' [v'(c_3 + s_3) - v'(s_3)] = v'(I - c_2)$$

$$c_2 = \frac{I - s_2 - (1-\alpha_s)\delta\gamma' c_3}{2}$$

The FOC for consumption in period 1 (c_1) is as follows:

$$v'(c_1 + s_1) + \alpha_s \gamma' [v'(c_2 + s_2) - v'(s_2)] + \alpha_s \gamma'^2 [v'(c_3 + s_3) - v'(s_3)] = v'(I - c_1)$$

$$c_1 = \frac{(I - s_1) - (1-\alpha_s)(\delta\gamma' c_2 + \delta^2 \gamma'^2 c_3)}{2}$$

Optimal consumption in the FI condition

In the optimal consumption equation, participants choose whether to meet the target (i.e., reduce their usage to $(1-\beta)c_{0F}$) or not. Accordingly, in periods 1 and 2, consumption will either be $(1-\beta)c_0$ or the optimal consumption c_t in the C condition, determined by the satiation level s_t (as in the C condition). The choice of the consumption is determined by probability p_t which depends on the net utility, which is a function of the incentive's attractiveness, the difficulty of reducing usage, and the satiation level s_t. The maximization problem is given by:

$$\underset{(c_{t'F},\dots c_{TF})}{\text{Max}} \sum_{t=t',\dots,T} \delta^t \alpha_s \left(\left(v(c_t + s_{t'}) - v(s_{t'})\right)\right) + (1-\alpha_s)\left(\left(v(c_t + s_t) - v(s_t)\right)\right) + \delta^t v(I - c_t),$$

In period 1 and 2, participants choose between c_{tF} and c_t with probability level p_{tF} depending on the satiation level and attractiveness of incentives.

The optimal consumption during period 1 is given by $c_{1F} = (1-\beta)c_0$ with probability p_{1F}, and by c_1 with probability $1 - p_{1F}$. Note that with probability p_{1F}, participants will not optimally reduce their consumption below $(1-\beta)c_0$ when satiation is not sufficiently high. The expected consumption in period 1 is:

$$c_{1F} = p_{1F}(1-\beta)c_0 + (1-p_{1F})c_1.$$

The expected consumption in period 2 is similarly given by:

$$c_{2F} = p_{2F}(1-\beta)c_0 + (1-p_{2F})c_{2F}.$$

The FOC for optimal consumption in period 3 is given by:

$$(1-\alpha_s)v'\left((c_{3F} + s_{3F})\right) + \alpha_s v'(c_{3F} + s_{1F}) = v'(I - c_{3F})$$

$$c_{3F} = \frac{I - s_{3F}}{2}$$

Optimal consumption in the AI condition

The FOC for optimal consumption in period 3 is given by:

$$(1-\alpha_s)v'\left((c_{3A} + s_{3A})\right) + \alpha_s v'(c_{3A} + s_{1A}) = v'(I - c_{3A})$$

$$c_{3A} = \frac{I - s_{3A}}{2} \qquad \text{(W5)}$$

The optimal consumption during period 2 is given by $c_{2A} = (1-\beta)c_0$ with probability p_{2A} and by c_2 with probability $1 - p_{2A}$. The expected consumption in period 2 is:

$$c_{2A} = p_{2A}(1-\beta)c_0 + (1-p_{2A})c_{2A}.$$

As discussed previously, participants will not reduce their consumption level below $(1-\beta)c_0$ when satiation is not sufficiently high.

The FOC for optimal consumption in period 1 is given by:

$$v'(c_{1A} + s_{1A}) + (1-\alpha_s)\gamma'[v'(c_{2A} + s_{2A}) - v'(s_{2A})] + (1 - \alpha_s)\gamma'^2[v'(c_{3A} + s_{3A}) - v'(s_{3A})] = v'(I - c_{1A})$$

$$c_{1A} = \frac{(I - s_{1A}) - (1-\alpha_s)(\delta\gamma' c_{2A} + \delta^2\gamma'^2 c_{3A})}{2} \qquad \text{(W6)}$$

Substituting the value of $c_{3A} = \frac{(I - s_{3A})}{2}$, and $s_{3A} = \gamma'(c_{2A} + \gamma'(c_{1A} + s_{1A}))$ we get:

$$c_{1A} = \frac{2(1 - s_{1A}) - 2(1-\alpha_s)c_{2A}\delta\gamma' - (1-\alpha_s)\delta^2\gamma'^2 + (1-\alpha_s)\delta^2\gamma'^3 c_{2A} + (1-\alpha_s)\delta^2\gamma'^4 s_{1A}}{4 - (1-\alpha_s)\delta^2\gamma'^4}$$

In the above expression, the net effect of c_{2A} is negative i.e., $-2(1-\alpha_s)c_{2A}\delta\gamma' + (1 - \alpha_s)\delta^2\gamma'^3 c_{2A} < 0$.

Proofs of Propositions 1 to 5

Proof of Proposition 1

To prove Proposition 1, we assume that the incentive is attractive such that $p_{2A} > 0$. We now compare the optimal consumptions c_{1A} in the AI and c_1 in the C condition under habit formation. If habit stock k_1 is small, then the optimal consumption in the C condition $c_0, \ldots, c_3$ is increasing (see Eq. W1 and discussion below it). This implies that $c_{2A} < c_2$ as $p_{2A} > 0$. The expression for c_1, when the expression for c_3 and k_3 is substituted, is as follows.

$$c_1 = \frac{(8b_1^2 + (1-\alpha)2b_1b_2\delta\gamma)I + k_1(4b_1b_2 + (1-\alpha)\gamma^4\delta b_2) + c_2(1-\alpha)(\gamma^2\delta b_2 + 4b_1b_2\gamma^2\delta^2)}{16b_1^2 - \gamma^3\delta b_2(1-\alpha)}$$

Comparing the above expression to the expression of c_{1A} below

$$c_{1A} = \frac{(8b_1^2 + (1-\alpha)2b_1b_2\delta\gamma)I + k_{1A}(4b_1b_2 + (1-\alpha)\gamma^4\delta b_2) + c_{2A}(1-\alpha)(\gamma^2\delta b_2 + 4b_1b_2\gamma^2\delta^2)}{16b_1^2 - \gamma^3\delta b_2(1-\alpha)}$$

we can see that $k_1 = k_{1A}$ and $c_{2A} < c_2$. Since $\alpha < 1$, we get $c_{1A} < c_1$. However, when $\alpha = 1$, $c_{1A} = c_1$.

Proof of Proposition 2

To prove Proposition 2, we divide period 2 into two sub-periods: 2a and 2b. We consider four different target structures for participants in AI condition:

i. Constant sub-period targets (CT): Target of $(1-\beta)c_0$ in both period 2a and 2b
ii. Decreasing sub-period targets (DT): Target of $(1-\beta)c_0 - \Delta$ in period 2a and $(1-\beta)c_0 + \Delta$ in period 2b, where $\Delta > 0$ and $(1-\beta)c_0 + \Delta < c_0$
iii. Increasing sub-period targets (IT): Target of $(1-\beta)c_0 + \Delta$ in period 2a and $(1-\beta)c_0 - \Delta$ in period 2b, where $\Delta > 0$ and $(1-\beta)c_0 + \Delta < c_0$
iv. Average period target (PT): An average target of $(1-\beta)c_0$ for the entire period 2, that is period 2a and period 2b combined.

We show in Web Appendix C that when participants are engaging in a habit-forming activity and are forward-looking with respect to habit formation, the DT strategy leads to the highest pre-emptive reductions in period 1, followed by CT, and finally IT: $c_{1A}^{DT} < c_{1A}^{CT} < c_{1A}^{IT}$ (see Web Appendix C for details). When participants are given an average period target (PT), they may attempt to achieve the average target in the following ways: (a) by evenly meeting the average target in both periods 2a and 2b (similar to CT) or, (b) by reducing more in period 2a and less in the period 2b (similar to DT), or (c) by reducing less in period 2a and more in period 2b (similar to IT). We demonstrate in Web Appendix C that under an average period target, the optimal behavior to adopt is to reduce less in period 2a and more in period

2b (similar to IT). Thus, as PT participants will adopt the IT strategy optimally, the pre-emptive reduction with a PT incentive structure is lower than with a CT incentive structure i.e., $c_{1A}^{CT} < \mathrm{c}_{1\mathrm{A}}^{PT}$, thereby proving Proposition 2.

Proof of Proposition 3

To prove Proposition 3, we compare the optimal consumption in period 3 in the FI and AI condition with the C condition under habit formation. The optimal consumption in period 3 in the FI, AI and C conditions is:

$$c_{3F} = \frac{2b_1 I + b_2 k_{3F}}{4b_1} = \frac{2b_1 I + b_2 \gamma(c_{2F} + \gamma(c_{1F} + k_{1F}))}{4b_1}$$

$$c_{3A} = \frac{2b_1 I + b_2 k_{3A}}{4b_1} = \frac{2b_1 I + b_2 \gamma(c_{2A} + \gamma(c_{1A} + k_{1A}))}{4b_1}$$

$$c_3 = \frac{2b_1 I + b_2 k_3}{4b_1} = \frac{2b_1 I + b_2 \gamma(c_2 + \gamma(c_1 + k_1))}{4b_1}$$

If habit stock k_1 is small, the optimal consumption $c_0, \ldots, c_3$ is increasing (see Eq. W1 and discussion below it). When the monetary incentives are attractive and $p_{2A} > 0$, it implies that $c_{1F} < c_1$ and $c_{2F} < c_2$. Similarly, $c_{1A} < c_1$ (see Proposition 1) and $c_{2A} < c_2$. Note that as $k_{1f} = k_{1A} = k_1$ (baseline is identical across conditions), from the above expressions, we get $c_{3f} < c_3$ and $c_{3A} < c_3$.

Proof of Proposition 4

To prove Proposition 4, we assume the incentive is attractive such that $p_{2A} > 0$. Given this assumption, we can derive the optimal consumption under satiation for the C and AI condition for different experimental periods. Under satiation, the optimal consumption in period 1 for the C condition is given by:

$$c_1 = \frac{(I - s_1) - (1 - \alpha_s)(\delta\gamma' c_2 + \delta^2 \gamma'^2 c_3)}{2}$$

Comparing this to the c_{1A} expression in the AI condition:

$$c_{1A} = \frac{(I - s_{1A}) - (1-\alpha_s)(\delta\gamma' c_{2A} + \delta^2\gamma'^2 c_{3A})}{2}$$

Expanding the expressions above by substituting $c_3 = \frac{I-s_3}{2}$, $c_{3A} = \frac{I-s_{3A}}{2}$, $s_3 = \gamma'(c_2 + \gamma'(c_1 + s_1))$, and $s_{3A} = \gamma'(c_{2A} + \gamma'(c_{1A} + s_{1A}))$, we obtain:

$$c_1 = \frac{2(1-s_1) - (1-\alpha_s)\delta^2\gamma'^2 + (1-\alpha_s)\delta^2\gamma'^4 s_1 - 2(1-\alpha_s)c_2\delta\gamma' + (1-\alpha_s)\delta^2\gamma'^3 c_2}{4 - (1-\alpha_s)\delta^2\gamma'^4}$$

$$c_{1A} = \frac{2(1-s_{1A}) - (1-\alpha_s)\delta^2\gamma'^2 + (1-\alpha_s)\delta^2\gamma'^4 s_{1A} - 2(1-\alpha_s)c_{2A}\delta\gamma' + (1-\alpha_s)\delta^2\gamma'^3 c_{2A}}{4 - (1-\alpha_s)\delta^2\gamma'^4}$$

Comparing the above expressions for c_1 and c_{1A}, we observe that the first three terms are identical: $2(1-s_1) - (1-\alpha_s)\delta^2\gamma'^2 + (1-\alpha_s)\delta^2\gamma'^4 s_1 = 2(1-s_{1A}) - (1-\alpha_s)\delta^2\gamma'^2 + (1-\alpha_s)\delta^2\gamma'^4 s_{1A}$ because $s_1 = s_{1A}$. Now comparing the last two terms, as $c_2 > c_{2A}$, we observe that $-2(1-\alpha_s)c_2\delta\gamma' + (1-\alpha_s)\delta^2\gamma'^3 c_2 < -2(1-\alpha_s)c_{2A}\delta\gamma' + (1-\alpha_s)\delta^2\gamma'^3 c_{2A}$. Therefore, $c_{1A} > c_1$ when $\alpha_s < 1$. On the other hand, when $\alpha_s = 1$, $c_{1A} = c_1$.

Proof of Proposition 5

Under satiation, we compare the optimal consumption in period 3 in the FI, AI and C condition. The optimal period 3 consumption is given by:

$$c_{3F} = \frac{I - s_{3F}}{2} = \frac{I - \gamma'(c_{2F} + \gamma'(c_{1F} + s_{1F}))}{2}$$

$$c_{3A} = \frac{I - s_{3A}}{2} = \frac{I - \gamma'(c_{2A} + \gamma'(c_{1A} + s_{1A}))}{2}$$

$$c_3 = \frac{I - s_3}{2} = \frac{I - \gamma'(c_2 + \gamma'(c_1 + s_1))}{2}$$

First, we compare c_3 and c_{3F}. Note that $c_{1F} < c_1$, $c_{2F} < c_2$ as $p_{1F} > 0$ and $p_{2F} > 0$. Therefore, $c_{3F} > c_3$.

We now compare c_3 and c_{3A}. We expand the expressions for c_{3A} and c_3:

$$c_{3A} = \frac{I - \gamma c_{2A} - \gamma^2 \left(\frac{(I - s_{1A}) - (1-\alpha_s)(\delta\gamma' c_{2A} + \delta^2 \gamma'^2 c_{3A})}{2} + s_{1A}\right)}{2}$$

Rewriting the expression, we obtain:

$$c_{3A} = \frac{2I - 2\gamma' c_{2A} - \gamma'^2(1 - s_{1A}) + \gamma'^3(1-\alpha_s)\delta c_{2A} + 2s_{1A}}{4}$$

Similarly,

$$c_3 = \frac{2I - 2\gamma' c_2 - \gamma'^2(1 - s_1) + \gamma'^3(1-\alpha_s)\delta c_2 + 2s_1}{4}$$

We know that $s_1 = s_{1A}$. Also $c_{2A} < c_2$ (as $p_{2A} > 0$), and therefore $-2\gamma' c_{2A} + \gamma'^3(1 - \alpha_s)\delta c_{2A} > -2\gamma' c_2 + \gamma'^3(1-\alpha_s)\delta c_2$. Thus, $c_{3A} > c_3$.

Web Appendix B: Optimal Incentive Strategies for Pre-emptive Usage Reductions

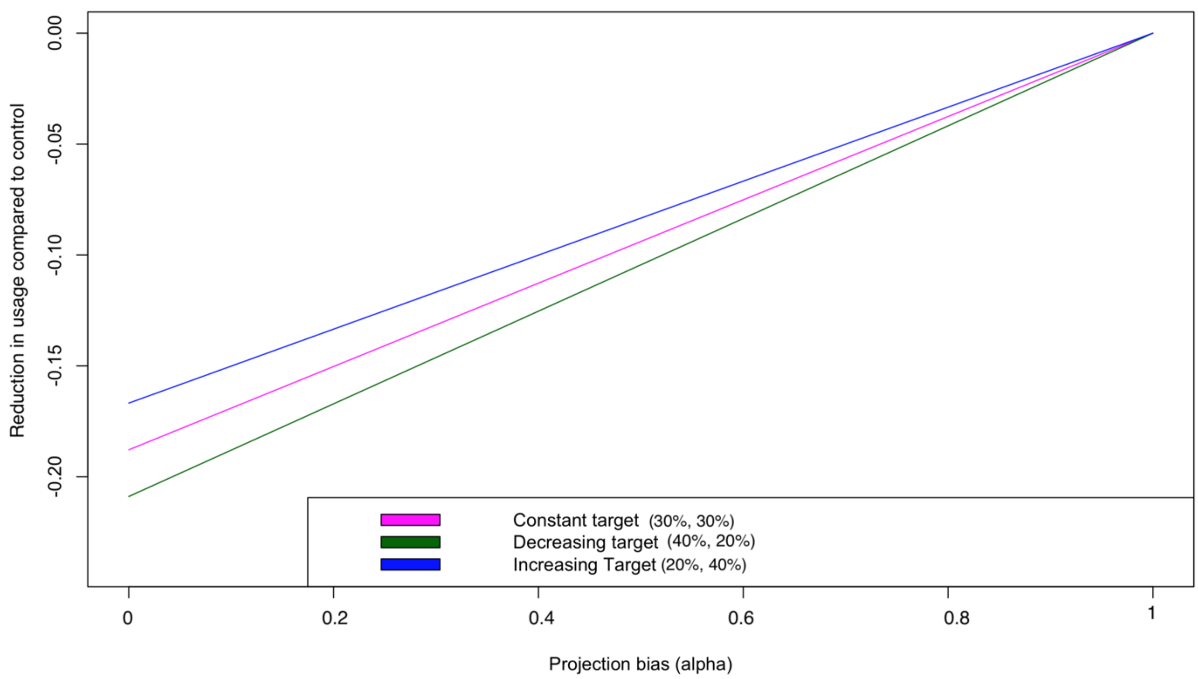
In practice, a variety of strategies could be employed to incentivize usage reductions. In this section, we discuss and simulate the features of incentive schemes that are optimal in stimulating a pre-emptive reduction of usage. We consider several possible incentive designs: Consumers could be incentivized to meet daily targets which are either more difficult or easier to achieve (i.e., high initial reduction targets followed by lower targets or vice versa) rather than maintaining constant daily targets, as in our RCTs. Alternatively, consumers could be incentivized based on their average reduction over the entire treatment period (either through a fixed overall target for the entire treatment period or by incentivizing any usage reduction proportionally, as in Allcott, Gentzkow, and Song (2022) and our RCT 3's API condition). We theorize that daily targets of decreasing difficulty level or constant daily targets are more effective at inducing pre-emptive reductions than daily targets of increasing difficulty level or average-based incentives (for a mathematical discussion of optimal strategies, see Web Appendix C).

For the estimated aggregate-level parameters, by fixing the average reduction target for the treatment period to 30%, we simulated the extent of pre-emptive usage reduction for different incentive designs. For simplicity (as we will assume in our model estimation), we assume the treatment period has two sub-periods. We evaluate the following strategies:

1) Constant target: 30% reduction in both sub-periods;
2) Decreasing target: 40% reduction in sub-period 1 and 20% reduction in sub-period 2;
3) Increasing target: 20% reduction in sub-period 1 and 40% reduction in sub-period 2.

Based on the results, we also discuss the expected anticipatory behavior when no specific target but a proportional incentive for the average reduction during a treatment period is used (as in Allcott, Gentzkow, and Song 2022). In Figure W1, we plot the pre-emptive usage reduction expected for each of the incentive strategies with respect to a control condition without a reduction target for different levels of projection bias (see Web Appendix C for equations). We observe that decreasing targets lead to stronger pre-emptive reductions than constant targets and increasing targets. The differences in pre-emptive reductions between groups are high when projection bias is close to zero (i.e., when participants are rationally addicted) and decrease with increasing projection bias. Without loss of generality, we can extend the findings to a treatment period of many days, in which case the pre-emptive reduction would be stronger across all groups.

Figure W1: Simulation of Pre-emptive Usage Reductions for Different Incentive Strategies.



Notes: Pre-emptive usage reduction is compared to a control group. We assume H = .17, and $\alpha = .64$ for the simulation (based on parameter estimates from RCT 2 model fitting results).

The intuition for the simulation result is as follows. When individuals are confronted with more difficult initial targets, they must reduce usage more during the anticipation period to habituate to a lower usage during the treatment period. On the other hand, when individuals are facing easier initial targets, or when they can choose the reduction amount each day (e.g., when incentivized based on average usage), they are less likely to reduce their usage pre-emptively because they can adapt more slowly to a lower usage during the treatment.

This crucial distinction in incentive design helps explain the disparity between our results and those of Allcott, Gentzkow, and Song (2022). While we find evidence of pre-emptive usage reductions, they find that individuals did not reduce usage in anticipation of an upcoming incentive period. Participants in their study were incentivized proportionally for any reduction in usage averaged over a three-week time period, whereas our participants were required to meet a specific daily reduction target.[3] As a result, participants in Allcott, Gentzkow, and Song (2022) might have chosen to gradually reduce their usage throughout the treatment period (similar to the simulated increasing-target group) to meet their desired average. Since participants were paid proportionally for any level of reduction from their baseline, the pre-emptive reduction effect is expected to be smaller.

We conclude that to foster a substantial pre-emptive reduction in domains characterized by forward-looking habit formation (with a lower projection bias), the implementation of daily targets, especially with ambitious initial targets, is essential. Daily targets, where each day serves as a unique unit for achievement, are more effective in stimulating pre-emptive reductions compared to incentives based on long-term averages.

[3] Consider a person with a baseline usage of four hours. In our study, the person is asked to reduce usage by 1 hour every day during the treatment period to receive the incentive for the day. Allcott, Gentzkow, and Song (2022) incentivized the average usage reduction over the entire treatment period at a rate of $50 per hour. The target was not proportional to the baseline usage. Somebody with a 30-minute average reduction over the treatment period can still receive $25 and somebody even with a 6-minute reduction can receive $5.

Web Appendix C: Mathematical Discussion of Different Incentive Strategies

To discuss the pre-emptive reduction under different incentive strategies in the AI condition, we divide period 2 into two sub-periods: 2a and 2b. Since our estimation showed consumers are forward-looking with respect to their habit formation, we study the predictions of the habit formation model (with $\alpha < 1$) for four different target-incentive strategies described below:

i. Constant sub-period targets (CT): Target of $(1-\beta)c_0$ in both period 2a and 2b
ii. Decreasing sub-period targets (DT): Target of $(1-\beta)c_0 - \Delta$ in period 2a and $(1-\beta)c_0 + \Delta$ in period 2b, where $\Delta > 0$ and $(1-\beta)c_0 + \Delta < c_0$
iii. Increasing sub-period targets (IDT): Target of $(1-\beta)c_0 + \Delta$ in period 2a and $(1-\beta)c_0 - \Delta$ in period 2b, where $\Delta > 0$ and $(1-\beta)c_0 + \Delta < c_0$
iv. Average period target (PT): An average target of $(1-\beta)c_0$ for the entire period 2, that is period 2a and period 2b combined.

We discuss the predictions of the habit formation model with projection bias. For simplicity, we refer to consumption in the AI condition in different periods as $c_{0A} = c_0, c_{1A}, c_{2aA}, c_{2bA}$, and c_{3A} and Control condition in different periods as c_0, c_1, c_{2a}, c_{2b}, and c_3. Optimal consumption in period 1 is given by:

$$c_{1A} = \frac{2b_1 I + b_2 k_1 + b_2(1-\alpha)(\gamma\delta c_{2aA} + \gamma^2\delta^2 c_{2bA} + \gamma^3\delta^3 c_{3A})}{4b_1}$$

We now compare the different target-incentive structures. We assume habit stock is not too high and that optimal consumption without constraints is increasing. Considering a probability p_{tA} of meeting the target in AI condition and c_{2a} and c_{2b} being the consumption in the C condition in period 2a and period 2b, the expected consumption in AI condition in periods 2a and 2b under each target-incentive structure will be as follows:

$$c_{2aA}^{CT} = p_{2aA}^{CT}(1-\beta)c_0 + (1-p_{2aA}^{CT})c_{2a}$$

$$c_{2bA}^{CT} = p_{2bA}^{CT}(1-\beta)c_0 + (1-p_{2bA}^{CT})c_{2b}$$

$$c_{2aA}^{DT} = p_{2aA}^{DT}((1-\beta)c_0 - \Delta) + (1-p_{2aA}^{DT})c_{2a}$$

$$c_{2bA}^{DT} = p_{2bA}^{DT}((1-\beta)c_0 + \Delta) + (1-p_{2bA}^{DT})c_{2b}$$

$$c_{2aA}^{IT} = p_{2aA}^{IT}\big((1-\beta)c_0 + \Delta\big) + (1-p_{2aA}^{IT})c_{2a}$$

$$c_{2bA}^{IT} = p_{2bA}^{IT}\big((1-\beta)c_0 - \Delta\big) + (1-p_{2bA}^{IT})c_{2b}$$

Given this, the optimal consumption in period 1 under each target-incentive structure is given by:

$$c_{1A}^{CT} = \frac{2b_1 I + b_2 k_{1A} + b_2(1-\alpha)(\gamma\delta(p_{2aA}^{CT}(1-\beta)c_0 + (1-p_{2aA}^{CT})c_{2a}) + \gamma^2\delta^2(p_{2bA}^{CT}(1-\beta)c_0 + (1-p_{2bA}^{CT})c_{2b}) + \gamma^3\delta^3 c_{3A})}{4b_1}$$

$$c_1^{DT} = \frac{2b_1 I + b_2 k_{1A} + b_2(1-\alpha)(\gamma\delta(p_{2aA}^{DT}((1-\beta)c_0 - \Delta) + (1-p_{2aA}^{DT})c_{2a}) + \gamma^2\delta^2(p_{2bA}^{DT}((1-\beta)c_0 + \Delta) + (1-p_{2bA}^{DT})c_{2b}) + \gamma^3\delta^3 c_{3A})}{4b_1}$$

$$c_1^{IT} = \frac{2b_1 I + b_2 k_{1A} + b_2(1-\alpha)(\gamma\delta(p_{2aA}^{IT}((1-\beta)c_0 + \Delta) + (1-p_{2a}^{IT})c_{2aA}) + \gamma^2\delta^2(p_{2b}^{IT}((1-\beta)c_0 - \Delta) + (1-p_{2b}^{IT})c_{2bA}) + \gamma^3\delta^3 c_{3A})}{4b_1}$$

Consumption $c_{3\mathrm{A}}$ in the expressions above can vary depending on $k_{3\mathrm{A}}$ and therefore depending on the strategy. However, the effect is insignificant, as it is scaled by $\gamma^3\delta^3$ and γ is not very high (γ = 0.47) in our estimation.

Typically, in period 2a, it is expected that $p_{2aA}^{IT} > p_{2aA}^{CT} > p_{2aA}^{DT}$ and in period 2a, it is expected that $p_{2bA}^{DT} > p_{2bA}^{CT} > p_{2bA}^{IT}$ because of the size of the reduction. If we assume that the monetary incentives are attractive (as we find in our studies),$p_{2aA} > 0$ and $p_{2bA} > 0$ across all conditions. In that case, when comparing the other terms, as $\alpha < 1$, it is clear that $c_{1A}^{DT} < c_{1A}^{CT} < c_{1A}^{IT}$ as $\gamma, \delta < 1$. Thus, the DT strategy leads to the highest pre-emptive reduction, followed by CT, and finally IT.

We now examine the impact of an average period target (strategy iv) on the pre-emptive reduction. This strategy entails giving participants an average target of $(1-\beta)c_0$ for the entire treatment period (PT). For simplicity, we assume that the monetary incentives are attractive, and participants want to reach $(1-\beta)c_0$ (our results do not change even if we assume a probability p_t of meeting the target). Given this setup, participants may attempt to achieve the average target in the following ways:

(a) by evenly meeting the average target in both periods 2a and 2b (similar to CT),

(b) by reducing more in period 2a and less in period 2b (similar to DT),

(c) by reducing less in period 2a and more in period 2b (similar to IT).

We demonstrate below that, under an average period target, the optimal behavior to adopt is to reduce less in period 2a and more in period 2b (similar to IT). To show this, we first compare the utility losses incurred when a participant with an average period target adopts the IT strategy versus when they adopt a DT strategy:

$$
\begin{aligned}
IT - DT = {} & u\big((1-\beta)c_0 + \Delta, k_{2aA}\big) - u\big((1-\beta)c_0 - \Delta, k_{2aA}\big) \\
& + u\big((1-\beta)c_0 - \Delta, \gamma(k_{2aA} + (1-\beta)c_0 + \Delta)\big) \\
& - u\big((1-\beta)c_0 + \Delta, \gamma(k_{2aA} + (1-\beta)c_0 - \Delta)\big)
\end{aligned}
$$

We expand using the quadratic form for the utility function and solve it to get:

$$IT - DT = 2b_2 k_{2aA}\Delta(1-\gamma) > 0$$

Thus, when a participant is given an average period target, they will adopt a IT strategy rather than a DT strategy. In other words, they will try to reduce their usage less initially, postponing larger reductions to later in the treatment period.

We now compare the utility loss if a participant with an average period target adopts CT versus a IT strategy:

$$IT - CT = u\big((1-\beta)c_0 + \Delta, k_{2aA}\big) - u\big((1-\beta)c_0, k_{2aA}\big)$$
$$+ u\big((1-\beta)c_0 - \Delta, \gamma(k_{2aA} + (1-\beta)c_0 + \Delta)\big)$$
$$- \ u\big((1-\beta)c_0, \gamma(k_{2aA} + (1-\beta)c_0)\big)$$

We expand using the quadratic form for the utility function and solve it to get:

$$IT - CT = k_{2aA} - \Delta\gamma - \gamma k_{2aA}$$

The above expression is greater than zero because if $IT - CT < 0$, then expanding the above term $\Delta > (1-\gamma)\big(c_{1A} + \gamma(c_0 + k_{0A})\big)$. Taking the upper bound $\Delta = \beta c_0$, this would imply $(\beta - \gamma)c_0 > (1-\gamma)(c_{1A} + \gamma k_{0A})$. This is not possible as our reduction $\beta \leq 0.3$. Therefore, $IT - CT > 0$.

Thus, when participants are given an average reduction target over the entire period, they will try to adopt an approach like IT. They will reduce their usage less in period 2a, allowing their habit stock to lower gradually, and aim for greater reductions in period 2b to meet the average target. As a result, the initial reduction in period 2a is smaller, which in turn implies that the pre-emptive reduction in period 1 is smaller under a PT strategy compared to DT or CT.

This underscores the importance of daily targets, particularly decreasing daily targets, for stimulating pre-emptive reductions. The bonus treatment in Allcott, Gentzkow, and Song (2022) can be viewed as a variation of the PT strategy. In their study, participants were eligible for a maximum payment of $150 if they reduced their average usage by at least 1.5 hours over three weeks. Participants were paid proportional to their average reduction, even if they failed to reduce their average usage by 1.5 hours. Given that participants were rewarded for any reduction from baseline, rather than for meeting specific daily goals, the pre-emptive reduction is expected to be smaller.

Web Appendix D: Methodology & Additional Analyses

RCT 1

Screen time data collection and verification checks

Participants had to activate a screen time tracking application on their smartphone (iPhone: 'Apple Screen Time', Android: 'Screen Time - Restrain yourself & parent control'). To receive the reward, they were not allowed to deactivate the app throughout the study.[4] Screen time was recorded automatically by the app whenever the mobile screen was activated (e.g., listening to music while the screen is "off" does not count as screen time; using a navigation app on the other hand counts as screen time).

Participants had to complete six weekly screen time reports one week apart. Every Saturday, participants received a 'screen time reporting' survey which asked them to manually enter their daily mobile screen time for each of the previous seven days, as elicited via the app (Figure W2-a). Participants also uploaded a screenshot of their usage from the app (Figure W2-b) for verification. We verified participants' self-reported screen time compared to each screenshot. To ensure screen time had been reported accurately, the screenshot had to show the participant's name consistently across all weekly reports (blacked out in Figure W2-a, b). The screenshot had to show the 'Last 7 Days' bar chart and the weekly average screen time (e.g., 1h 46m per day). In case of discrepancy, participants were contacted via email and text message for clarification. Daily screen time constituted our main dependent variable.

Incentives

Participants who completed all tasks were eligible to receive a €15 Amazon voucher and a variable payment between €0 and €38 based on performance. Additionally, 10% of

[4] In case participants deactivated the screen time app, the entire history of usage was deleted automatically, allowing us to identify and exclude participants who deactivated their app.

participants who completed all study requirements were eligible to win a lottery of €150. The rewards were calculated and distributed at the end of the study after all data collection.

Figure W2-A: Bars with Daily Screen Time.

Figure W2-B: Typical Screenshot Submitted.

Baseline characteristics after randomization

In the baseline survey, we asked questions about demographics and mobile usage, such as the estimated daily mobile screen time, the number of smartphones owned, monthly data allowance etc. All variables along with the relevant statistics are presented in Table W1. The sample was balanced across conditions on these characteristics. Importantly, there was no difference in baseline screen time across conditions, see Figure W3.

Figure W3: Baseline Screen Time Across Conditions (Hours Per Day).

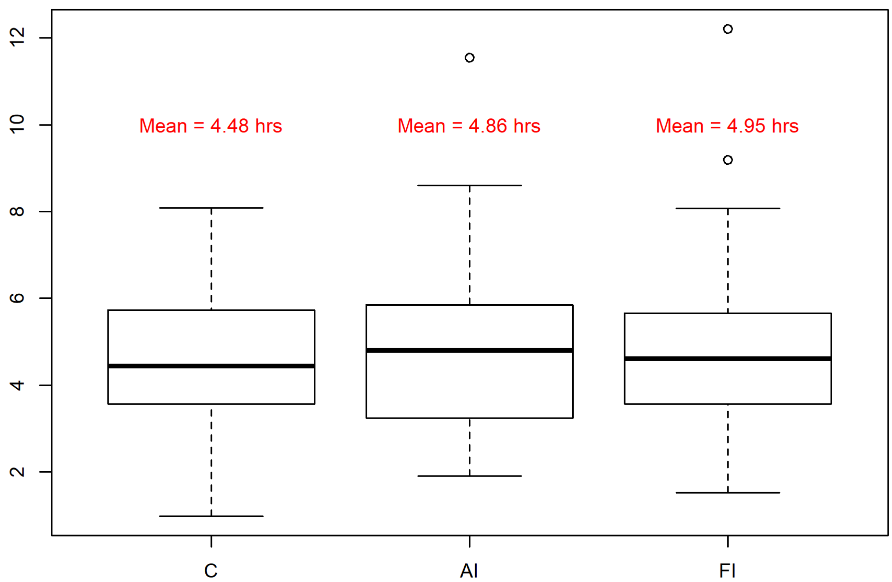
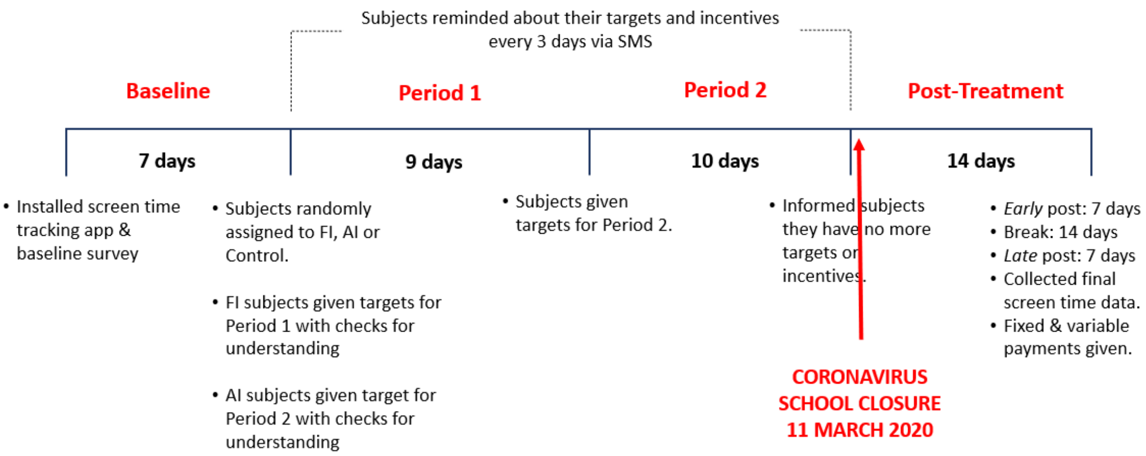


Table W1: Baseline Characteristics of Participants Across Conditions.

	Full sample	Control	FI	AI	Test statistic
Demographics					
Age	21.11 (2.24)	21.15 (2.23)	21.10 (2.19)	21.10 (2.36)	$F(2, 107) = .01$, $p = .994$
Female	67.27%	66.67%	69.23%	65.79%	$\chi^2(2) = .11$, $p = .946$
Years living in country of university	5.35 (7.85)	5.76 (7.83)	6.71 (8.69)	3.61 (6.78)	$F(2, 107) = 1.58$, $p = .210$
Bachelor	65.45%	72.73%	64.10%	60.53%	$\chi^2(2) = 1.21$, $p = .546$
Located on main campus	69.09%	57.58%	76.92%	71.05%	$\chi^2(2) = 3.23$, $p = .198$
Smartphone variables					
iOS (Apple iPhone)	86.36%	84.85%	84.62%	89.47%	$\chi^2(2) = .47$, $p = .788$
Owner of two phones	3.77%	0%	7.69%	2.78%	$\chi^2(2) = 2.96$, $p = .227$
Post-paid phone plan	56.36%	57.58%	56.41%	55.26%	$\chi^2(4) = .09$, $p = .999$
Monthly data allowance	3.44 (.89)	3.44 (.84)	3.51 (.94)	3.37 (.91)	$F(2, 98) = .22$, $p = .799$
Estimated percentage of screen time spent on mobile	60.72 (15.48)	64.09 (14.49)	59.87 (16.03)	58.68 (15.66)	$F(2, 107) = 1.17$, $p = .313$
Estimated percentage of mobile screen time spent on social media	55.52 (17.60)	52.75 (16.47)	55.17 (16.77)	58.28 (19.35)	$F(2, 107) = .88$, $p = .417$
Willingness to reduce mobile usage	4.9 (1.47)	5.18 (1.64)	4.82 (1.44)	4.82 (1.34)	$F(2, 107) = .89$, $p = .415$
Has previously attempted to reduce mobile usage	76.36%	78.79%	74.36%	76.32%	$\chi^2(2) = .19$, $p = .907$

Timeline and attrition

Figure W4: Study Timeline.



See Table W2 for the attrition during the study across conditions. N denotes the number of participants who reported data out of 110 participants. Attrition was low during the treatment periods, but increased during the post-treatment periods, especially the late post-treatment, due to the unexpected arrival of the COVID-19 pandemic and the subsequent university closure. We provide further analyses of the post-treatment in a separate section.

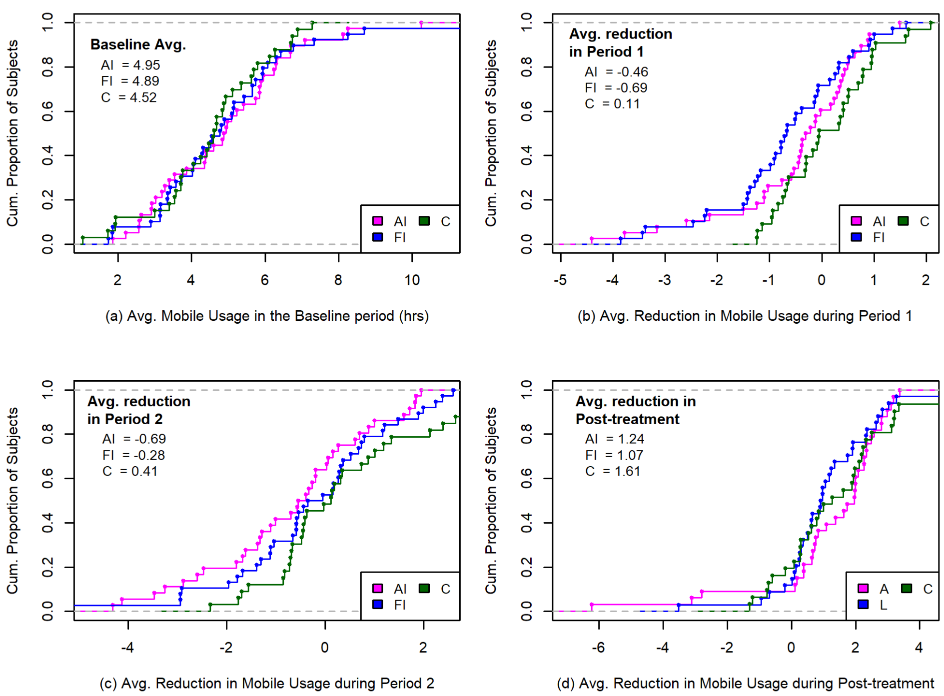
Table W2: Proportion of Participants Reporting Mobile Usage in Different Survey Rounds.

Screen Time Reports	**N**	**Total Proportion (N/110)**	**Control**	**AI**	**FI**
Completed baseline	110	1.00	1.00	1.00	1.00
Completed report 1 usage (period 1)	109	.99	0.97	1.00	1.00
Completed report 2 usage (period 1 & 2)	109	.99	.97	1.00	1.00
Completed report 3 usage (period 2)	105	.95	1.00	.95	.92
Completed report 4 usage (period 2 & post-treatment	103	.94	1.00	.92	.90
Completed report 5 usage (immediate post-treatment)	95	.86	.85	.89	.85
Completed report 6 usage (late post-treatment)	55	.50	.36	.66	.46

Distribution of change in mobile usage for each participant across conditions

Figures W5a-d plot the cumulative distribution functions (CDFs) of average usage with each dot representing one participant. In the baseline (Figure W5a), the CDFs overlap, indicating the average baseline usage was similar across conditions. Figures W5b-d plot the *change* in average usage relative to the baseline and reveal the proportion of participants that reduced usage. In period 1 (Figure W5b), the CDFs for FI and AI consistently lie above the C condition, indicating a higher proportion of FI and AI participants decreased their usage compared to the C condition. In period 2 (Figure W5c), the CDFs for FI and AI participants consistently lie above the C condition, indicating a higher proportion decreased their usage compared to the C condition. In the post-treatment (Figure W5d), the CDFs of FI, AI, and C conditions shifted to the right because participants increased usage due to the COVID-19 pandemic. We still notice that the FI condition lies above the C condition, indicating participants in the FI condition sustained a lower usage than the C condition.

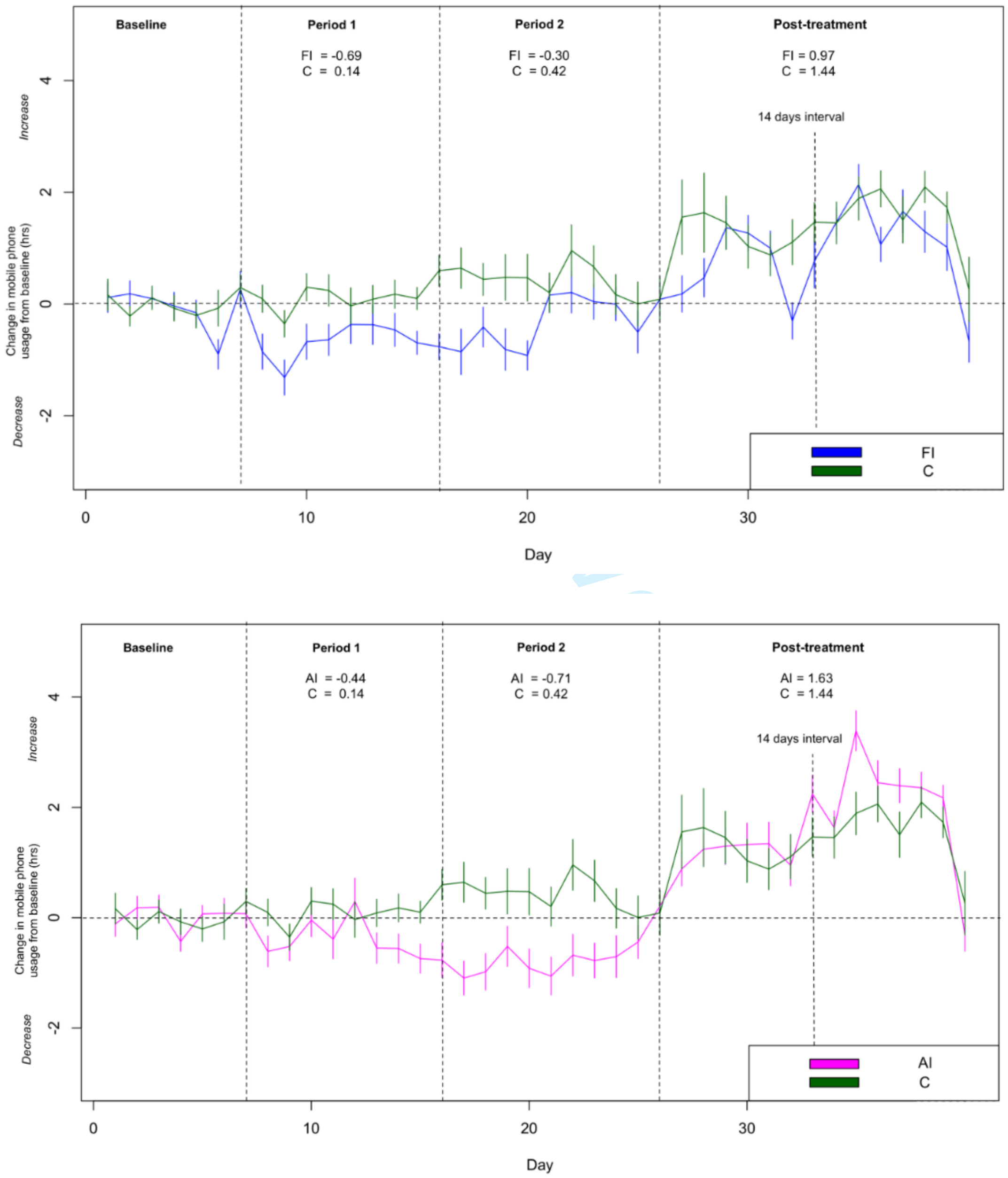
Figure W5a-d: Cumulative Distribution Functions: a. Usage in baseline. b. Usage reduction from baseline in period 1. c. Usage reduction from baseline in period 2. d. Usage reduction from baseline in post-treatment.



Average change in mobile usage across conditions

Figure W6a (top) compares the average change in mobile usage (from baseline) between the FI and C condition. Figure W6b (bottom) compares the average change in mobile usage (from baseline) between the AI and C condition. There was no difference in baseline mobile usage across conditions (p = .56).

Figure W6a–b: Average Change in Daily Mobile Usage from Baseline Across Conditions.

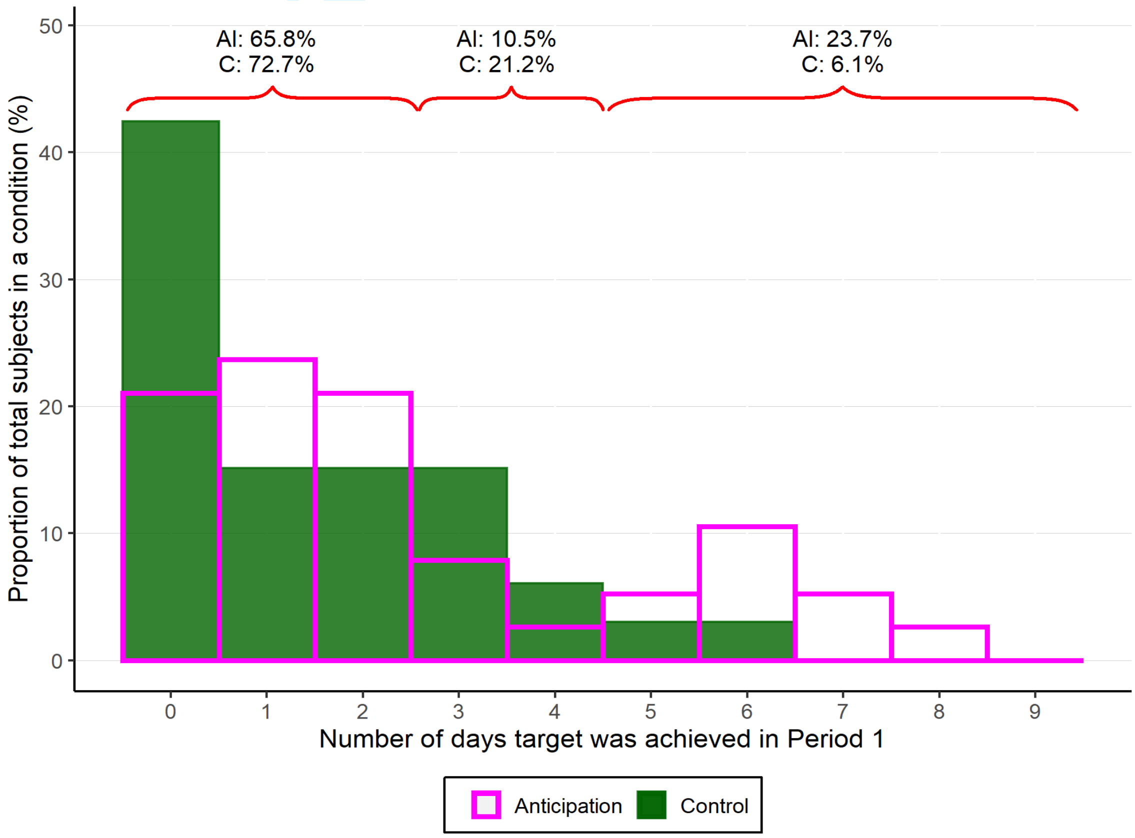


Notes: Standard error bars are shown.

Which participants responded to the AI treatment?

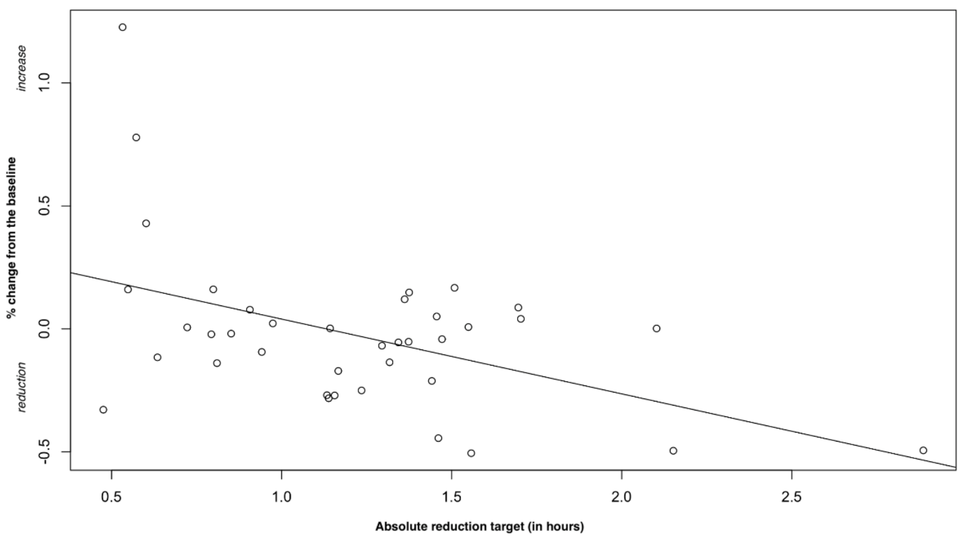
In Figure W7, we observe that an equal proportion of participants in the AI and C condition (AI: 66%, C: 73%) reduced their mobile usage by at least 25% between zero and two days. However, a significantly higher proportion of participants in the AI condition (23.7%) reduced their mobile usage by at least 25% between five to nine days compared to the C condition (6.1%, $p = .03$). Frequent reduction was achieved by 23.7% of participants who met their target for at least five days.

Figure W7: Target Achievement Days in the AI and C Condition in Period 1.



Next, we identify the participant group who responded more strongly to the anticipation treatment. In Figure W8, we plot the target (in hours) on the x-axis and the number of days the target was achieved on the y-axis. While all individuals were given the same relative target in terms of percentage (i.e., a 25% reduction), the absolute value of this reduction (in hours) was higher for individuals with higher baseline usage. We find a downward sloping relationship in the AI condition. Figure W8 shows that the absolute reduction required to meet the target significantly predicts the percentage of usage reduction in the AI condition (β = -.3, p = .003), but not in the FI condition (p = .17). The results do not change when we add controls (gender, operating system, and age). This indicates that the pre-emptive usage reduction in the AI condition was stronger for participants with higher (absolute) reduction targets (i.e., those with higher baseline smartphone usage). Results are equivalent when we focus on the number of days in period 1 participants reduced their usage by ≥ 25% as dependent variable. The absolute reduction target positively and significantly predicts the number of days the target was achieved in period 1 (β = 1.61, p = .03). In the FI condition we find no relationship between target magnitude and target achievement (p = .91).

Figure W8: Scatterplot of AI Participants with Reduction Target in Hours on the X-Axis, % Usage Reduction and Period 1 Target Achievement on the Y-Axis.



Post-treatment analysis of screen time

In the post-treatment period, the FI condition had a marginally lower usage by.864 hours compared to the C condition (p = .052) and .763 hours compared to the AI condition (p = .061, see Table W3 for the OLS regression). Thus, participants in the FI condition decreased their usage in periods 1 - 2 and maintained a lower usage than the C condition even without incentives. Participants in the AI condition did not sustain a lower usage during the post-treatment period. However, due to post-treatment attrition, it is not possible to estimate the treatment effect precisely. Hence, these results should be interpreted cautiously.

Table W3: Difference-in-Difference OLS Estimation of Post-treatment Usage.

	Dependent Variable = Mobile Usage (hour)
AI condition	.362 (.403)
FI condition	.423 (.406)
Period 1	.122 (.154)
Period 2	.443 (.282)
Post-treatment Period (Period 3)	1.675*** (.349)
AI condition × Period 1	-.552** (.269)
FI condition × Period 1	-.836*** (.251)
AI condition × Period 2	-1.098*** (.393)
FI condition × Period 2	-.761** (.367)
AI condition × Period 3	-.101 (.462)
FI condition × Period 3	-.864* (.443)
Constant	4.529*** (.269)
Observations	3,569
R^2	.086
Adjusted R^2	.083
Residual Std. Error (df =3557)	2.367
F Statistic (df = 11; 3557)	30.460***

Notes: *p = .10; **p < .05; ***p < .01. Standard errors (clustered at the subject level) are shown in parentheses. Reference categories are C condition and baseline period.

Baseline vs. post-treatment comparison of additional survey measures

Reducing mobile phone usage might benefit a person's level of productivity, focus and smartphone-related anxiety (Ward et al. 2017). Thus, we measured these variables before and after the treatment period. In addition, we elicited participants' time preference using two intertemporal choice items (Loewenstein and Prelec 1992) before and after the treatment period. We used two items from the Cognitive Reflection Test (Frederick 2005) to measure any change in cognitive reflection. We measured cognitive reflection because prior research (Frederick 2005) has found that individuals with higher CRT scores exhibit lower present bias and fewer self-control problems. Based on this, we explored whether CRT might be associated with more forward-looking behavior and stronger responses to incentives. Table W4 tabulates all self-reported items measured at baseline and/or the post-treatment period.

Toward the end of period 2 (first week of March 2020), COVID-19 cases appeared in the university and started receiving considerable news coverage. We expected this to affect the consumption of social media and other news sources via mobile, so we asked participants how concerned they were about the COVID-19 pandemic and whether they thought it had increased their phone usage. On average, participants were quite concerned (M = 4.19, SD = 1.94 on a seven-point scale) about COVID-19. Interestingly, controlling for baseline mobile usage, participants in the FI condition were less worried about the COVID-19 pandemic compared to the AI (p = .023) and C condition (p = .065). Thus, the FI treatment—through reducing mobile usage—may have reduced concern about COVID-19.

In the FI condition, participants predicted their success in achieving the target for period 1 and 2 ("Predicted"). They also predicted the success of a randomly selected person with the same instructions ("Predicted Others"). In the AI condition, participants only completed this prediction task for period 2. Table W5 shows the results. Participants tended to be overconfident in both conditions and periods. In the FI condition, participants overestimated their target achievement in Period 1 ($t(37) = 5.27$, $p < .001$) and in Period 2 ($t(31) = 3.40$, $p = .001$) in comparison to the actual group mean. Similarly, in the AI condition, participants overestimated their target achievement compared to the group mean ($t(33) = 3.60$, $p < .001$).

Table W4: Self-Reported Measures Across Conditions and Survey Rounds.

	Baseline					Post-treatment				
Item	**All**	**C**	**AI**	**FI**	**Test statistic**	**All**	**C**	**AI**	**FI**	**Test statistic**
Productivity:										
Do you think your mobile phone usage hinders your productivity? (1=not at all, 7=very much)	5.08 (1.45)	5.27 (1.33)	4.71 (1.54)	5.28 (1.45)	$F(2, 107) = 1.91, p = .153$	4.40 (1.74)	4.39 (1.91)	4.28 (1.85)	4.54 (1.48)	$F(2, 100) = .19, p = .828$
How focused are you these days considering your work? (1=not at all, 7=extremely focused)	4.44 (1.19)	4.64 (1.11)	4.29 (1.29)	4.41 (1.16)	$F(2, 107) = .76, p = .471$	4.28 (1.31)	4.30 (1.26)	4.17 (1.34)	4.37 (1.35)	$F(2, 100) = .21, p = .813$
Nomophobia: How anxious would you feel if you forgot to bring your mobile phone for a day? (1=not anxious at all, 7=very anxious)	4.47 (1.72)	4.73 (1.81)	4.37 (1.70)	4.36 (1.69)	$F(2, 107) = .51, p = .602$	4.34 (1.78)	4.63 (1.88)	4.28 (1.56)	4.14 (1.91)	$F(2, 100) = .68, p = .510$
Inter-temporal choices: Which option do you prefer?										
Item 1: A: 10€ today B:11€ in one week										
Item 2: A: 10€ in 4 weeks B:11€ in 5 weeks										
Impatient (% always option A)		24.24	44.74	35.90			30.30	42.86	37.14	
Patient (% always option B)		45.45	23.68	30.77			39.39	20.00	45.71	
Present-biased (% option A in Item 1, option B in item 2)		9.09	18.42	15.38	$\chi^2(6) = 6.49, p = .371$		9.09	20.00	5.71	$\chi^2(6) = 8.65, p = .194$
Anti-present-biased (% option B in Item 1, option A in item 2)		21.21	13.16	17.95			21.21	17.14	11.43	
Cognitive Reflection Test:										
Racquet and tennis ball problem (% correct)	43.64	45.45	44.74	41.03	$\chi^2(2) = .17, p = .918$					
Lily pad problem (% correct)						66.02	63.64	71.43	62.86	$\chi^2(2) = .69, p = .706$
COVID-19 related items:										
How concerned are you about the Corona virus outbreak? (1=not at all, 7=very concerned)						4.19 (1.94)	4.39 (1.86)	4.62 (2.00)	3.57 (1.83)	$F(2, 100) = 2.96, p = .056$
Do you think the Corona virus outbreak increased your phone usage? (1=not at all, 7=very much)						4.05 (2.34)	3.93 (2.37)	4.34 (2.47)	3.88 (2.21)	$F(2, 100) = .39, p = .677$

Table W5: Actual and Predicted Target Achievements.

Condition	Period 1				Period 2			
	N	Actual	Predicted	Predicted Others	N	Actual	Predicted	Predicted Others
FI	38	3.87 (2.75)	5.8 (2.29)	4.9 (1.54)	32	4.22 (3.57)	5.5 (2.54)	4.9 (1.93)
AI	-	-	-	-	34	4.24 (3.46)	6.35 (2.41)	5.06 (2.07)

Notes: Standard deviation in brackets.

Target achievement and grade point average

We found no significant difference in grade point average (GPA) between conditions. But target achievement in both periods was positively associated with GPA. Every additional day of target achievement (i.e., reduced usage by at least 25%) in both periods was associated with an increase in GPA by .035 (p = .03) on a 10-point scale.[5] While this evidence is correlational, it may suggest that students who reduced their usage (and achieved targets) obtained a higher GPA. Of course, it is entirely plausible that high-GPA students responded more strongly to the treatment. Table W6 shows the regression results.

Table W6: OLS Estimation of Association Between Targets Achievement and GPA.

	Dependent Variable = GPA
# Days target achieved in Period 1 & 2	.035** (.017)
Master	-.708*** (.170)
Weekday baseline usage	-.114** (.046)
Constant	8.788*** (.230)
Observations	104
R^2	.224
Adjusted R^2	.201
Residual Std. Error	.827 (df = 100)
F Statistic	9.610*** (df = 3; 100)

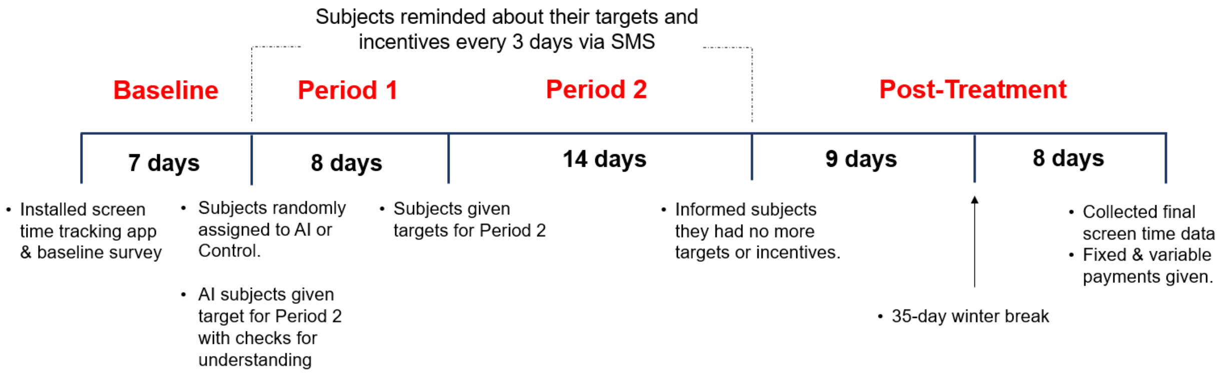
Notes: *p < .1; **p < .05; ***p < .01. Standard errors (clustered at the subject level) are shown in parentheses. Reference category is Bachelor students.

[5] In RCT 2, as participants were from different programs, we did not perform a similar analysis.

RCT 2

We recruited sixty-eight participants (49 females; M_{age} = 25.14, SD = 3.91) from the same European university but using a different cohort. Participants received the same instructions about incentives and activation of the screen time app as in RCT 1. Participants completed six weekly screen time reports and additional surveys delivered via email. The rewards were calculated and distributed at the end of the study after all data collection. The study lasted for six weeks (Nov 2020 - Jan 2021, see timeline in Figure W9).

Figure W9: Study timeline.



Changes to experimental design compared to RCT 1

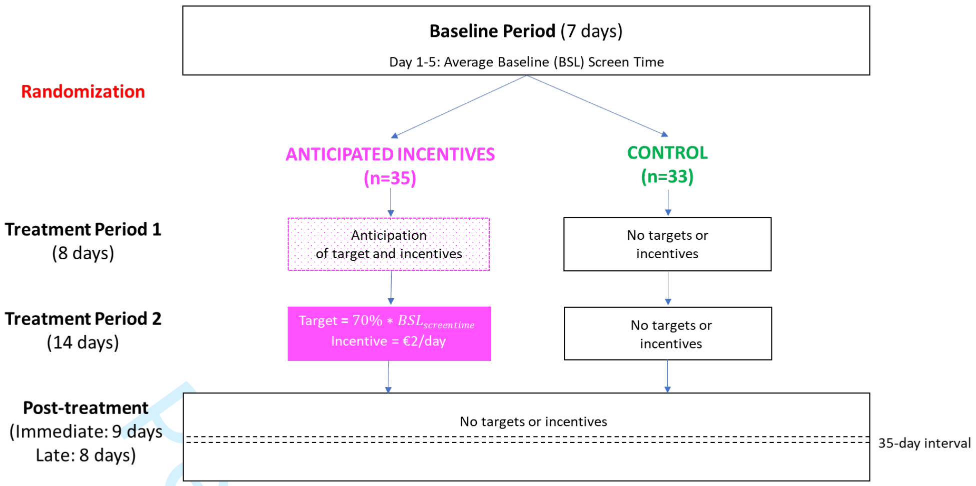
The procedure was identical to RCT 1 apart from the following changes. The aim was to strengthen the anticipatory usage reduction in period 1 and to support habit formation in period 2 for a lasting post-treatment reduction. See Figure W10 for the experimental design.

1) **Randomization**: Instead of three conditions, we focused only on the AI condition and compared it to the C condition. Hence, after the 7-day baseline period, participants were randomly assigned to either the AI (N = 35) or C condition (N = 33).

2) **Target**: The target in the AI condition was a 30% reduction from the average baseline usage and therefore slightly higher than in RCT 1.

3) **Duration of periods**: We reduced the duration of period 1 (anticipation period) to eight days and extended the duration of period 2 (incentive period) to 14 days.

Figure W10: Experimental Design of RCT 2.



There were no differences between conditions at baseline on key observable characteristics, including the average baseline screen time (see Table W7).

Table W7: Baseline Characteristics of Participants Across Conditions.

	Full sample	**Control**	**AI**	**Test statistic**
Age	25.14 (3.91)	24.78 (2.90)	25.48 (4.69)	$t(66) = .73, p = .466$
Female	72.06%	72.73%	71.43%	$\chi 2(1) = .014, p = .905$
iOS (Apple iPhone)	86.76%	90.91%	82.86%	$\chi 2(1) = .95, p = .327$
Owner of two phones	7.35%	6.06%	8.57%	$\chi 2(1) = .15, p = .692$
Post-paid phone plan	57.35%	48.48%	65.71%	$\chi 2(2) = 3.53, p = .171$
Estimated % of screen time on phone	52.86 (19.78)	54.84 (20.40)	51.00 (19.28)	$t(66) = -.79, p = .426$
Estimated % social media of mobile screen time	52.35 (22.01)	57.12 (21.28)	47.85 (22.03)	$t(66) = -1.76, p = .082$
Willingness to reduce mobile usage	5.11 (1.40)	5.24 (1.45)	5.00 (1.37)	$t(66) = -.70, p = .482$
Has attempted to reduce mobile usage	83.82%	84.85%	82.86%	$\chi 2(1) = .04, p = .824$
Average baseline screen time	5.22 (2.17)	5.11 (2.08)	5.34 (2.27)	$t(66) = .43, p = .667$

Attrition

See Table W8 for the attrition in RCT 2. N denotes the number of participants out of 68 who reported their usage for a particular period. Attrition was low during the treatment periods and the immediate post-treatment, but it increased during the late post-treatment. As there was differential attrition between the AI and C condition in the late post-treatment, we provide additional analysis in a separate section below.

Table W8: Proportion of Participants Reporting their Mobile Usage in Different Rounds.

Screen time reports	**N**	**Total Proportion (N/110)**	**Control**	**AI**
Completed baseline survey	68	1.00	1.00	1.00
Completed report 1 usage (baseline)	68	1.00	1.00	1.00
Completed report 2 usage (baseline & period 1)	67	.99	1.00	.97
Completed report 3 usage (period 1 & period 2)	66	.97	1.00	.94
Completed report 4 usage (period 2)	64	.94	.97	.91
Completed report 5 usage (period 2 & post-treatment)	64	.94	1.00	.89
Completed report 6 usage (immediate post-treatment)	62	.91	.97	.86
Completed report 7 usage (late post-treatment)	49	.72	.82	.63

Which participants responded to the AI treatment?

In Figure W11, we plot the proportion of participants who reduced their usage by 30% or more for different number of days in both the AI and C condition. As in RCT 1, we find that 60% of participants reduced their usage by 30% or more on at least 2 days (out of 8 days) in the AI condition compared to 36% in the C condition (p = .05).

We replicate the finding from RCT 1 that participants with higher baseline usage and therefore higher absolute targets respond more strongly to the anticipation treatment. Figure

W12 plots AI participants' reduction target in hours on the x-axis and the percentage usage change in period 1 (from baseline) on the y-axis. The reduction target significantly predicts the percentage usage reduction in the AI condition (b = -.12, p = .03).

Figure W11: Target Achievement Number of Days in the AI and C Condition in Period 1.

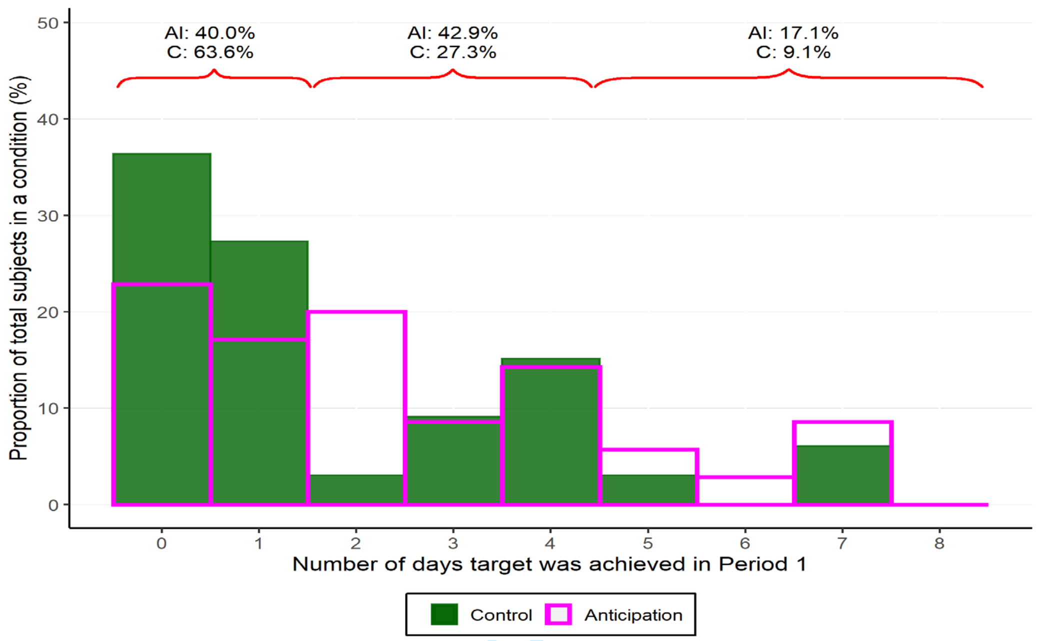
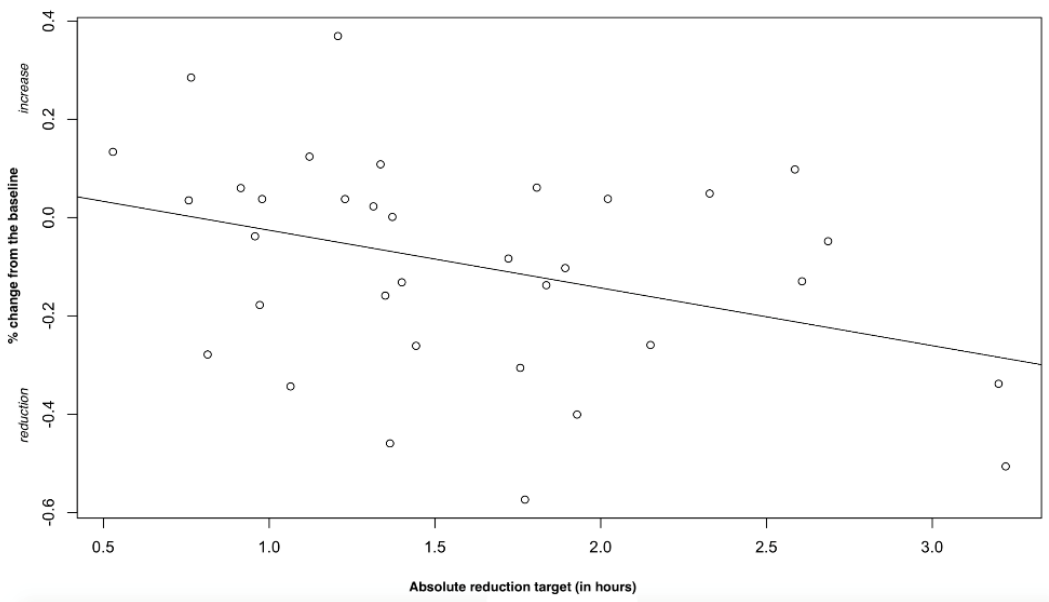


Figure W12: Scatterplot of AI Participants with Reduction Target in Hours the X-axis, Change in Period 1 % Usage from the Baseline on the Y-axis.



Qualitative analysis of reasons for reducing screen time in period 1

After period 1, we asked AI participants whether they had tried to change their mobile screen time during the previous (anticipation) period. The majority (61.76%) reported that they had tried to reduce their screen time during period 1. 26.47% indicated they had not changed their screen time, 11.76% indicated they had increased their screen time. Those who indicated they had reduced their screen time were asked to provide an explanation in a text box. A content analysis of the responses revealed that one third of the explanations were directly related to the anticipation of targets and incentives in period 2 (see Table W9 for verbatim quotes). Other responses were related to random coincidences (23%), personal goal achievement (38%) or expectations of positive outcomes (14%).

Table W9: Explanations for Reducing Screen Time During Period 1.

Theme	Example quotes
Anticipation of period 2 (33%)	*"I knew it was going to be more difficult to reduce 30% without doing it gradually. Therefore, I used this week to get use to this new usage."* *"I tried to change my screen time a bit to practice for my target this upcoming week. I didn't quite do it, but thought it would be good to start mentally preparing."* *"I tried to decrease my screen time to better prepare for this next phase."*
Personal goal achievement (38%)	*"I wanted to focus on only one thing at a time and avoid multitasking. Also to challenge myself and actually achieve my goals."* *"Because I felt I was using my phone too much and it was distracting me from my duties."* *"Personal goal to not spend too much time on my phone"*
Expectation of positive outcomes (14%)	*"I feel more productive if I reduce my screen time."* *"To increase productivity"* *"Because I needed to stay focused on my masters"*
Random coincidences (23%)	*"I was sick and tried to sleep more."* *"Consciously, I didn't really try to reduce it but I think that it was greatly affected by the workload I was given for the week and therefore, I had less free time to be on my phone."* *"I didn't actively try to, it just happened to be less."*

Table W10: Baseline vs. Post-treatment Comparison of Additional Survey Measures.

Item	Baseline				Post-treatment			
	All	C	AI	Test statistic	All	C	AI	Test statistic
Attention and Procrastination								
Do you think your mobile phone distracts you from more important things? (1=not at all, 7=very much)	5.32 (1.48)	5.57 (1.48)	5.08 (1.46)	$t(66) = 1.37$, $p = .174$	5.01 (1.58)	5.30 (1.55)	4.72 (1.58)	$t(64) = 1.49$, $p = .140$
How often do you have difficulty keeping your attention when you are doing boring or repetitive work? (1=never, 7=very often)	5.60 (1.31)	5.85 (1.03)	5.37 (1.54)	$t(66) = 1.49$, $p = .140$	5.24 (1.14)	5.36 (1.14)	5.24 (1.17)	$t(64) = .43$, $p = .672$
When you have a task that requires a lot of thought, how often do you avoid or delay getting started? (1=never, 7=very often)	4.92 (1.63)	5.09 (1.70)	4.74 (1.59)	$t(66) = .87$, $p = .387$	4.88 (1.33)	5.24 (1.35)	4.52 (1.25)	$t(64) = 2.27$, $p = .027$
Nomophobia								
How anxious would you feel if you forgot to bring your mobile phone for a day? (1=not anxious at all, 7=very anxious)	4.94 (1.85)	5.21 (1.79)	4.69 (1.90)	$t(66) = 1.17$, $p = .246$	4.34 (1.78)	4.73 (1.70)	4.70 (1.53)	$t(64) = .07$, $p = .939$
Fear of missing out								
When I have a good time, it is important for me to share the details online. (1=not at all true of me, 7=extremely true of me)	3.07 (1.61)	3.36 (1.64)	2.83 (1.58)	$t(66) =1.37$, $p = .175$	2.79 (1.62)	2.85 (1.60)	2.79 (1.67)	$t(64) = .15$, $p = .881$
Self-control								
I am good at resisting temptation (1=not at all, 7=very much)	4.08 (1.31)	4.18 (1.29)	3.97 (1.36)	$t(66) = .65$, $p = .515$	4.23 (1.44)	4.12 (1.65)	4.33 (1.24)	$t(64) = -.59$, $p = .558$
I have a hard time breaking bad habits (1=not at all, 7=very much)	4.36 (1.38)	4.48 (1.25)	4.20 (1.49)	$t(66) = .85$, $p = .398$	4.29 (1.41)	4.39 (1.32)	4.21 (1.53)	$t(64) = .52$, $p = .608$
I am able to work effectively toward long-term goals. (1=not at all, 7=very much)	4.92 (1.37)	4.91 (1.51)	4.97 (1.27)	$t(66) = .19$, $p = .854$	4.85 (1.38)	4.67 (1.53)	5.06 (1.22)	$t(64) = 1.15$, $p = .253$
COVID Concern								
How concerned are you about the Covid-19 situation? (1=not at all concerned, 7=very concerned)					4.80 (1.34)	4.75 (1.39)	4.84 (1.32)	$t(64) = .27$, $p = .786$

In the AI condition, participants predicted the number of days (out of 14) they expected to meet the usage reduction target at the beginning of the anticipation period, and again at the beginning of the incentive period. Table W11 shows the results.

Table W11: Actual and Predicted Target Achievements in the AI condition.

Actual target achievement	Predicted at start of anticipation period	Predicted at start of incentive period
7.02 (5.02)	8.22 (3.36)	7.97 (4.18)

Notes: Standard deviation in brackets.

Late post-treatment analysis of screen time

The late post-treatment results are reported in Figure W13 and Table W12. AI condition participants used their phone on average 1.1 hours less per day than in the C condition. Although the OLS regression supports this, we have ~30% missing observations in the late post-treatment and differential attrition (41% missing observations in the AI condition vs. 18% in the C condition (p = .006)). We conducted additional analysis controlling for attrition using Heckman selection. Even after correcting for selection, the AI condition had a lower usage in the late post-treatment. Nevertheless, since we have differential attrition, it is difficult to conclude that the AI condition had a lower usage than the C condition in the late post-treatment. Therefore, these results should be interpreted cautiously.

Figure W13: Average Change in Mobile Usage (in Hours) in the Late Post-Treatment Compared to Baseline.

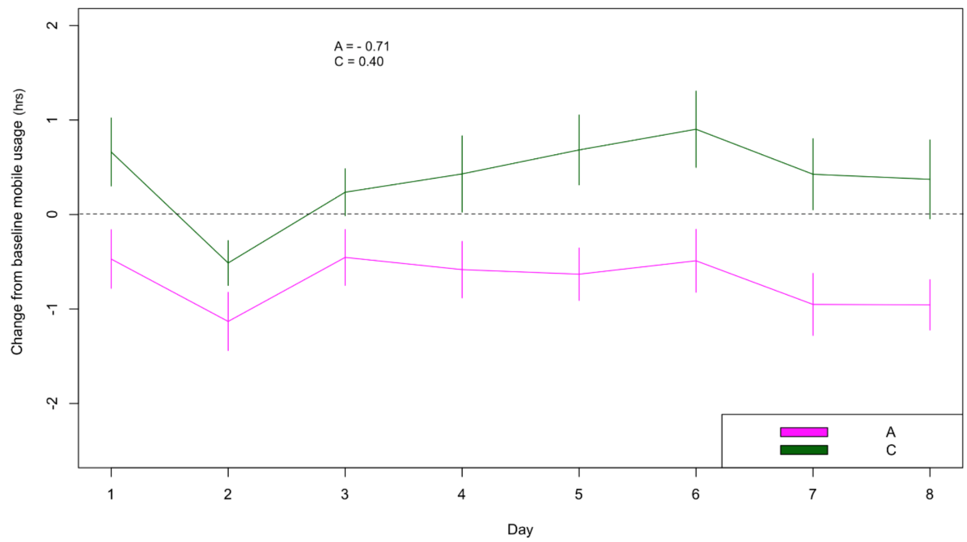


Table W12: AI Condition Maintains Lower Mobile Usage in the Late Post-Treatment.

	Dependent Variable = **Late post-treatment usage – Baseline usage (hrs)**
AI condition	-1.109*** (.405)
Constant	.400 (.263)
Observations	380
R^2	.073
Adjusted R^2	.071
Residual Std. Error	1.935 (df = 378)
F Statistic	29.872*** (df = 1; 378)

Notes: * $p < .10$; ** $p < .05$; *** $p < .01$. Standard errors (clustered at the subject level) are shown in parentheses.

Adjacent complementarity - instrumental variable regression results

Formula for the instrumental variable regression:

$$Post\ change_{it} = \alpha + \beta_1\left(Avg.\ Period\ \ 2\ change\hat{\ }_i\right) + \sum_{j=2} \beta_j\ \ (Other\ controls) + \epsilon_i$$

$$Avg. Period\ x\ change_i = \alpha_1 + \beta'(condition_i) + \sum_{j=2} \beta'_j\ \ (Other\ controls) + \epsilon'_i$$

where $Post\ change_{it} = immediate\ post\ treatment\ usage_{it} - baseline_i$

and $Avg. Period\ x\ change_i = Avg\ (Period\ x\ usage_{it} - baseline_i)$, where $x = 1{,}2$.

Table W13: Instrumental Variable Regression with Change in Period 1 and 2 Usage (from Baseline) Predicting Change in Post-Treatment Usage.

	Dependent variable = Immediate post-treatment usage – baseline
Avg. (P1 & P2 usage – Baseline)	.603** (.268)
Gender	.191 (.223)
Age	-.023 (.029)
Operating system (1 = iOS; 0 = Android)	.356 (.447)
Constant	2.093 (1.45)
Observations	564
R^2	.475
Adjusted R^2	.471
Residual Std. Error (df = 1063)	1.465

Notes: * *p* < .10; ** *p* < .05; *** *p* < .01. Baseline is the average usage during baseline.

Standard errors (in parentheses) are robust and clustered at the subject level.

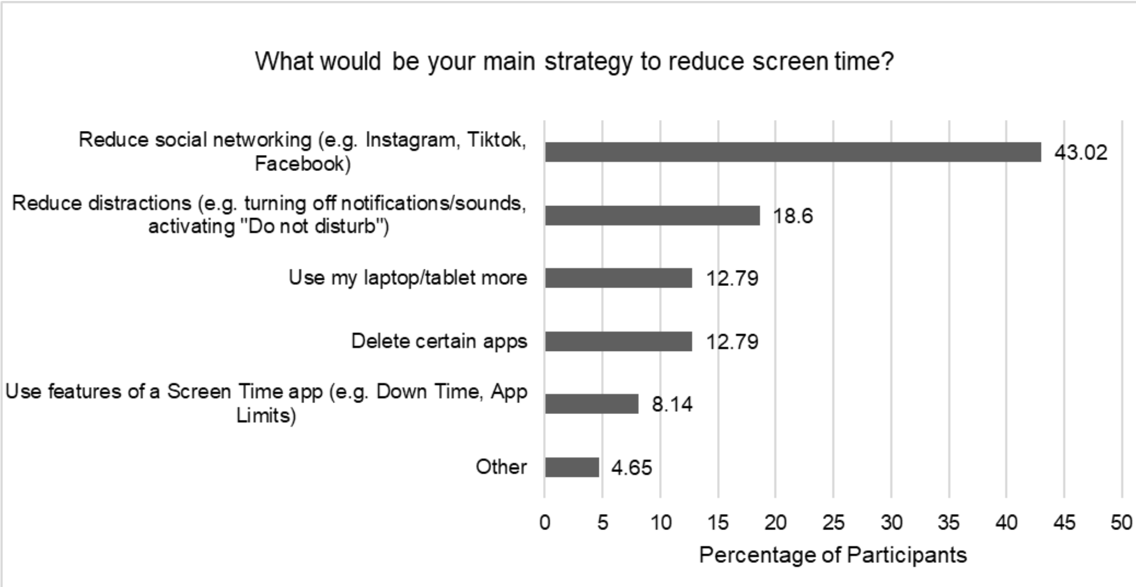
Lay belief survey with an independent sample

We conducted an online survey with an independent sample from the same student population (N = 86, 87% female, M_{age} = 20.09, SD = .93) to understand (i) whether participants hold accurate lay beliefs about forward-looking habit formation, (ii) their preferred length of an anticipation period and (iii) potential substitution behavior. Participants first read a description of the experimental setup of the AI condition in RCT 2. They then predicted how they would behave during period 1 (option 1: *I will use my mobile phone as much as I can now, because I will use it less during the incentivized period* (forward-looking satiation); option 2: *I will already start reducing my usage now to make it easier to meet the target in the incentivized period* (forward-looking habit formation)*;* option 3: *I will use my mobile phone just like before, because the incentives only start in 10 days from now* (myopic)).

More than half of participants selected the option indicating forward-looking habit formation (59.3%) while only a small minority selected forward-looking satiation (4.65%). The remaining 36.05% chose the option indicating myopia. We also asked participants to indicate whether they would like to change the lead time before period 1 (scale: 0-30 days). The ideal lead time participants indicated was on average 8.86 days (SD = 7.32). These results indicate that individuals have a lay belief about forward-looking habit formation and prefer having an anticipation period to prepare and gradually change habits.

We also asked participants what strategies they would employ to reduce their mobile usage (Figure W14). The largest proportion of participants chose reducing social networking (43%), followed by reducing distractions from the mobile (18.65%). Only 12.8% indicated they would substitute mobile usage with a laptop or tablet. Hence, while there might be some across-device substitution, we believe it is unlikely to be the main strategy.

Figure W14: Reducing Social Networking Is Main Strategy to Reduce Screen Time.



RCT 3

Table W14: RCT 3 Period 1 Instructions in the FI, AI And API Condition.

FI Condition	AI Condition	API Condition
Congratulations! You have been selected to receive a **Screen Time Bonus.** **Starting from Wednesday (DATE),** and continuing for the **next 18 days (until DATE)**, you can earn money for reducing your mobile screen time by **90 minutes** from your baseline, which is your average screen time from the first week of the study (**X hours Y minutes**). For each of the 18 days (from DATE - DATE) that your mobile screen time is **below your target** (TARGET: **X' hrs Y' minutes**) you will receive **$2.** **Your mobile screen time during the first week of the study (or baseline) was <u>X hours Y minutes</u>.** **Your target screen time, starting Wednesday, is <u>X' hours Y' minutes</u> (90 minutes lower than your baseline).** You can earn a **maximum amount of $36** (i.e.,$2 per day for 18 days), if your mobile screen time is below your	Congratulations! You have been selected to receive a **Screen Time Bonus.** HOWEVER, the Screen Time Bonus **starts in one week from now.** So for the time being, there is no specific task. **Starting in one week (DATE),** and continuing for the **next 12 days (until DATE)**, you can earn money for reducing your mobile screen time by **90 minutes** from your baseline, which is your average screen time from the first week of the study (**X hours Y minutes**). For each of the 12 days (from DATE - DATE) that your mobile screen time is **below your target** (TARGET: **X' hours Y' minutes** you will receive **$2.** **Your mobile screen time during the first week of the study (or baseline) was <u>X hours Y minutes.</u>** **Your target screen time, <u>starting in one week</u>, is <u>X' hours Y' minutes</u> (90 minutes lower than your baseline).** You can earn a **maximum amount of $24** (i.e.,$2 per day for 12	Congratulations! You have been selected to receive a **Screen Time Bonus.** HOWEVER, the Screen Time Bonus **starts in one week from now.** So for the time being, there is no specific task. **Starting in one week (DATE),** and continuing for the **next 12 days (until DATE)**, you will earn $16 for every hour you reduce your **average screen time** relative to your baseline mobile usage of **X hours Y minutes**, up to a maximum payout of $24. **There is no specific target for you**—your earnings depend on how much you reduce your screen time. The more you reduce, the more you earn (up to a maximum of $24). You can choose to reduce your usage **by any amount.** For example: • If you reduce your mobile screen time during the bonus period on average **by 30 minutes** from your baseline, you will earn **$8**. • If you reduce your mobile screen time during the bonus period on

target each day. For example: • If on any given day your screen time is **only 30 minutes lower** than your baseline, you will **not receive** a bonus for that day. • However, if you reduce your screen time by the full **90 minutes or more**, you will earn **$2 for that day**. **Summary**: • Baseline: X hours Y minutes • Target: X' hours Y' minutes (90 minutes lower than baseline) • Bonus Period: DATES • Earnings: $2 per day if your mobile screen time is 90 minutes below your baseline, up to $36 total.	days), if your mobile screen time is below your target each day. For example: • If on any given day your screen time is **only 30 minutes lower** than your baseline, you will **not receive** a bonus for that day. • However, if you reduce your screen time by the full **90 minutes or more**, you will earn **$2 for that day**. **Summary**: • Baseline: X hours Y minutes • Target: X' hours Y' minutes (90 minutes lower than baseline) • Bonus Period: DATES • Earnings: $2 per day if your mobile screen time is 90 minutes below your baseline, up to $24 total. **Note: This target does NOT apply IMMEDIATELY but starts in the next phase, one week from now.** **We are letting you know about your target in advance so you can prepare for it.**	average **by 15 minutes** from your baseline, you will earn **$4.** • However, if your average screen time during the bonus period **exceeds your baseline usage,** you will not earn any bonus ($0). **Important**: You reward is based on your **average screen time over the entire 12-day period**, not on individual days. If your screen time is higher on some days, you can make up for it by reducing it more on other days. **Summary**: • Baseline: X hours X minutes • Bonus Period: DATES • Earnings: $16 for every hour reduced, up to a maximum of $24. **Note: This screentime bonus does NOT apply IMMEDIATELY but starts in the next phase, one week from now.** **You don't have a specific target, but we are letting you know about the screen time bonus in advance so you can prepare for it.**

Sample description and attrition

Table W15: Demographics and Smartphone Variables Across Conditions.

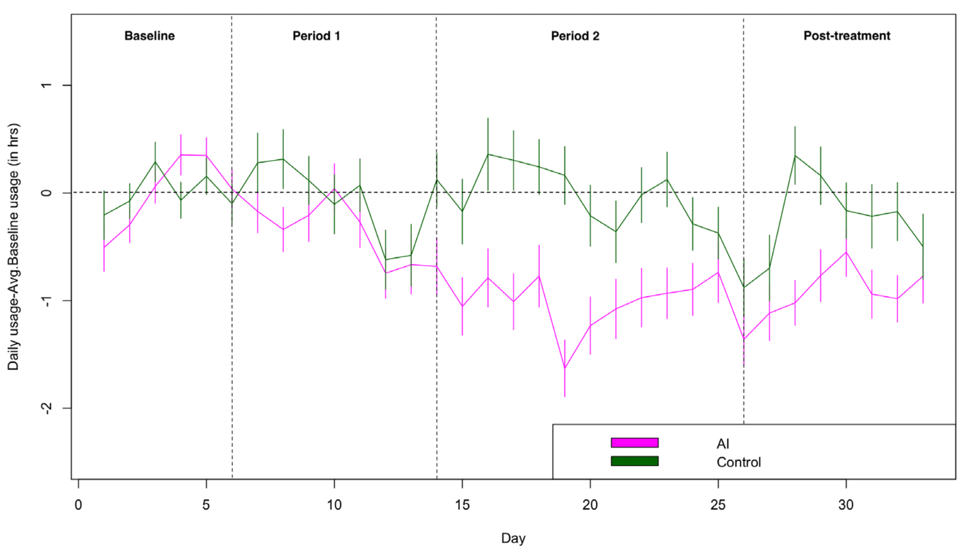
	Full sample	Control	FI	AI	API	Test statistic
Demographics						
Age	42.75 (14.49)	44.27 (13.89)	41.82 (15.75)	41.96 (14.69)	43.07 (13.80)	$F(3, 400) = .57$, $p = .632$
Gender						
Male	42.33%	38.20%	45.45%	41.59%	43.86%	$\chi^2(6) = 1.70$, $p = .945$
Female	55.45%	59.55%	53.41%	55.75%	53.51%	
Non-binary	2.23%	2.25%	1.14%	2.65%	2.63%	
Educational level						
High School or less	11.14%	13.48%	11.36%	14.16%	6.14%	$\chi^2(12) = 9.95$, $p = .620$
Some college	18.81%	22.47%	18.18%	17.70%	17.54%	
Associate's degree	13.12%	14.61%	11.36%	11.50%	14.91%	
Bachelor's degree	38.12%	35.96%	34.09%	38.05%	42.98%	
Graduate degree	18.81%	13.48%	25.00%	18.58%	18.42%	
Income level						
Low (< $30,000)	21.53%	16.85%	25.00%	23.01%	21.05%	$\chi^2(12) = 7.57$, $p = .818$
Lower-middle ($30,000-$49,999)	21.29%	22.47%	21.59%	23.01%	18.42%	
Middle ($50,000-$74,999)	22.52%	21.35%	19.32%	24.78%	23.68%	
Upper-middle ($75,000-$124,999)	21.53%	24.72%	25.00%	15.04%	22.81%	
High (> $125,000)	12.12%	14.61%	9.09%	14.16%	14.04%	
Race/ethnicity						
White/Caucasian	72.03%	76.40%	76.14%	67.26%	70.18%	$\chi^2(9) = 8.98$, $p = .439$
Black/African American	11.14%	13.48%	7.95%	9.73%	13.16%	
Asian	6.19%	5.62%	4.55%	7.96%	6.14%	
Other	10.64%	4.49%	11.36%	15.04%	10.53%	
Nr. of people above 18 years in same household	2.18 (1.08)	2.14 (.95)	2.07 (1.11)	2.33 (1.13)	2.15 (1.09)	$F(3, 399) = 1.08$, $p = .359$
Smartphone variables						
iOS (Apple iPhone)	63.61%	59.55%	64.77%	66.37%	63.16%	$\chi^2(3) = 1.07$, $p = .785$
Tablet regular usage	36.62%	40.45%	31.82%	42.48%	31.58%	$\chi^2(3) = 4.35$, $p = .226$
Estimated % of screen time spent on smartphone	51.94 (21.93)	49.22 (21.61)	51.29 (21.34)	54.21 (21.61)	52.32 (22.95)	$F(3, 400) = .89$, $p = .444$

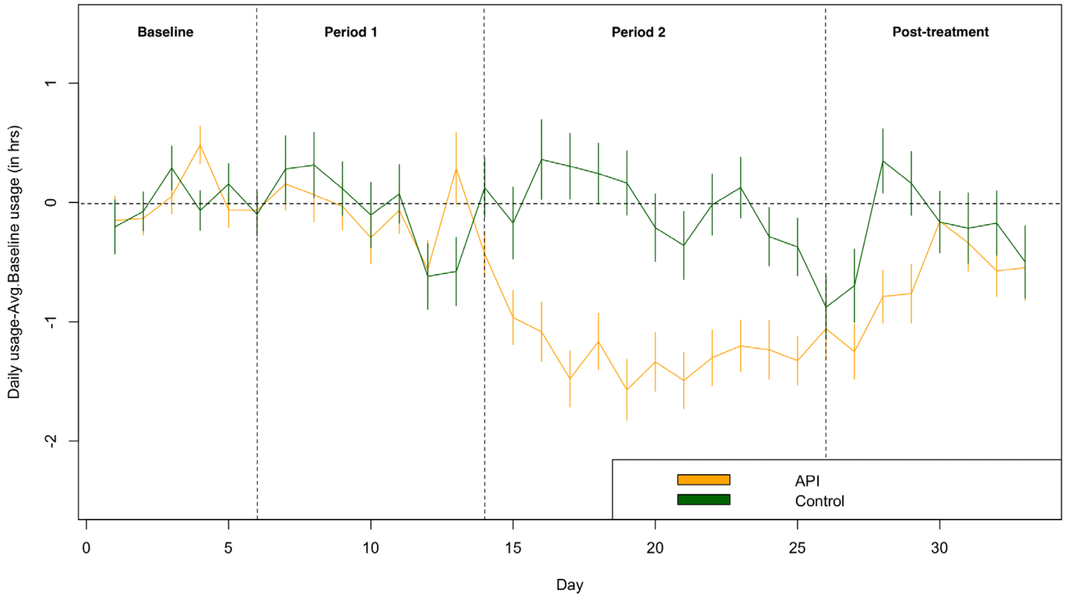
Ideal smartphone usage reduction (scale: 0-100%)	36.36 (21.91)	36.35 (25.15)	33.61 (18.35)	40.52 (23.63)	33.48 (18.88)	F(3, 201) = 1.30, *p* = .276
Willingness to reduce mobile usage (1-7 scale)	4.28 (1.50)	4.32 (1.62)	4.16 (1.56)	4.32 (1.46)	4.31 (1.41)	F(3, 400) = .24, *p* = .867
Previously attempts to reduce mobile usage (% yes)	55.45%	59.55%	55.68%	60.18%	47.37%	$\chi^2(3) = 4.64$, *p* = .200

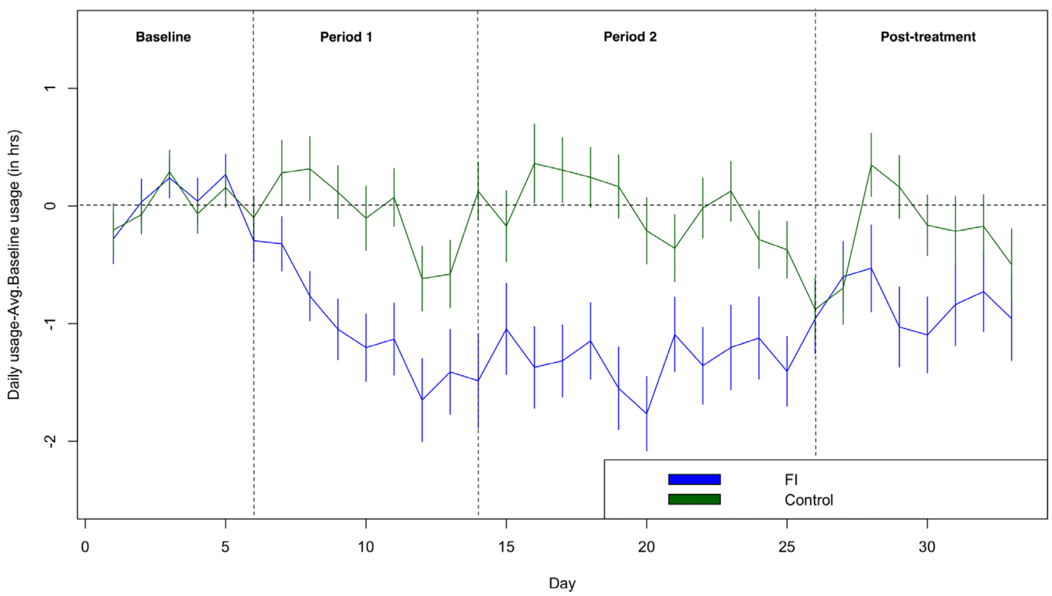
Table W16: Percentage of Missing Observations Across Conditions and Periods.

Condition	**Period 1**	**Period 2**	**Post-treatment**
Control	7.5%	2.6%	10.5%
AI	2.6%	10%	12.3%
API	2.3%	7.7%	13.5%
FI	2.5%	4.3%	6.7%

Figure W15 (a) – (c): Day-level Mobile Usage Reduction Compared to Control Condition During Period 1, 2 and Post-treatment in the AI (a), API (b), and FI (c) Condition.




Baseline
Period 1
Period 2
Post-treatment
Daily usage-Avg.Baseline usage (in hrs)
1
0
-1
-2
API
Control
0
5
10
15
20
25
30
Day


Baseline
Period 1
Period 2
Post-treatment
Daily usage-Avg.Baseline usage (in hrs)
1
0
-1
-2
FI
Control
0
5
10
15
20
25
30
Day

Goal achievement (90 minutes lower than baseline usage) as dependent variable

Table W17: OLS regression Predicting Goal Achievement Across Conditions.

	Dependent Variable: Goal Achievement (1/0)	
	beta	p-value
API	-.015 (.019)	.428
C	-.009 (.021)	.665
FI	.004 (.021)	.844
Period 1	.105(.028)***	.000
Period 2	.273(.037)***	.000
Period 3	.169(.033)***	.000
Age	-.0001(.001)	.909
Gender (Male)	.013(.021)	.541
iPhone	-.015(.021)	.471
API x Period 1	-.012(.036)	.742
C x Period 1	-.066(.037)*	.07
FI x Period 1	.153(.044)***	.000
API x Period 2	.011(.049)	.011
C x Period 2	-.22(.044)***	.000
FI x Period 2	-.012(.052)	.815
API x Period 3	-.023(.044)	.598
C x Period 3	-.109(.04)***	.009
FI x Period 3	.006(.05)	.911
Constant	.178(.038)***	.000
Observations	13,332	
R^2	.058	
Adjusted R^2	.056	
Residual Std. Error	.448 (df = 13313)	

Social Media Usage

Table W18: Average Social Media Usage (Hours Per Week) Per Period in Control and Treatment Conditions.

	Control Condition	**Treatment Conditions**	**FI**	**AI**	**AC**
Baseline	20.9 (17.13)	19.5 (17.86)	18.92 (18.71)	20.69 (17.16)	18.89 (17.96)
Period 1	19.1 (13.23)	19.36 (17.71)	17.14 (16.55)	21.31 (17.04)	19.15 (19.12)
Period 2	21 (15.05)	16.7 (15.23)	17.25 (15.36)	17.75 (15.59)	15.07 (14.74)
Post-treatment	18.6 (14.7)	16.41 (15.92)	17.01 (17.97)	16.4 (13.43)	15.88 (16.75)

Self-reported Social Media Goals and Usage Changes

In the baseline survey, participants reported their ideal usage change for different app categories: social media and messaging, entertainment apps, productivity apps, and gaming apps. There were no significant differences between conditions at baseline (all p's = ns). Participants reported wanting to reduce usage of social media, entertainment, and gaming apps (all p's < .001), ratings were significantly below the scale midpoint ("same as I do now"). There was no significant desire to change usage of productivity apps. Additionally, participants expressed stronger desire to reduce social media compared to entertainment and gaming apps (p's < .001).

In the endline survey, participants self-reported how much they had changed their usage in each app category over the past four weeks. Those in the treatment conditions reported reducing social media and messaging app usage more than those in the C condition ($F(3, 358) = 6.30$, $p < .001$). Post-hoc comparisons with Bonferroni correction showed significant differences between the C and treatment conditions (C vs. FI: $p = .022$, C vs. AI:

$p = .053$, C vs. API: $p < .001$). There were no significant differences between conditions for entertainment, gaming, or productivity apps.

Comparing ideal usage change (at baseline) with self-reported usage change (at endline) revealed that participants did not reduce their usage of social media, entertainment, or gaming apps as much as they had initially intended (all p's < .001). In contrast, there was no significant difference between ideal and actual change for productivity apps ($p = .424$).

Participants estimated the percentage of their overall mobile screen time attributed to different app categories (social media & messaging, entertainment, productivity, utility, and others), both before the study and at the end. We calculated a difference score and found a significant effect of condition for the social media category ($F(3, 358) = 4.96$, $p = .002$), but no significant differences for other app categories. While participants in the C condition increased their percentage of social media usage (C: 40.4% vs. 42.5%, $p = .046$), participants in the FI and API condition reduced it (FI: 40.5% vs. 37%, $p = .009$, API: 35.8% vs. 33.8%, $p = .039$). There was no significant change in the AI condition.

In summary, these findings suggest that social media was the primary category consumers wanted to cut back on, and while the intervention facilitated a reduction more than in other app categories, participants still fell short of their initial (possibly over-ambitious) goals to reduce social media.

Adjacent Complementarity: Results of Instrumental Variable Regression

Table W19: Period 1 Reduction Marginally Predicts Post-Treatment Usage Reduction.

	Dependent Variable: Post-treatment usage reduction		
	beta	Std. error	p-value
Intercept	.234	.387	.546
Period 1	.474	.378	.089
Age	-.009	.006	.190
Gender (Male)	-.272	.195	.164
iPhone	-.319	.207	.124
Residual Std. Error	1.75 (df = 326)		
Multiple R^2	.276		
R^2 adjusted	.267		
Wald Test	1.801 (df = 4, 326), p = .128		

Table W20: Period 2 Reduction Significantly Predicts Post-Treatment Usage Reduction.

	Dependent Variable: Post-treatment usage reduction		
	beta	Std. error	p-value
Intercept	.322	.342	.347
Period 2	.506**	.184	.006
Age	-.006	.006	.269
Gender (Male)	-.206	.170	.227
iPhone	-.170	.173	.328
Residual std. error	1.482 (df = 323)		
Multiple R^2	.479		
R^2 adjusted	.473		
Wald Test	3.286 (df = 4, 323), p = .012		

Heterogeneity in Period 1 Reduction (Period 1 usage – Baseline)

Note that participants with higher baseline usage would also have a higher absolute target.

Figure W16: Absolute Target Predicts Period 1 Usage Reduction in the AI Condition.

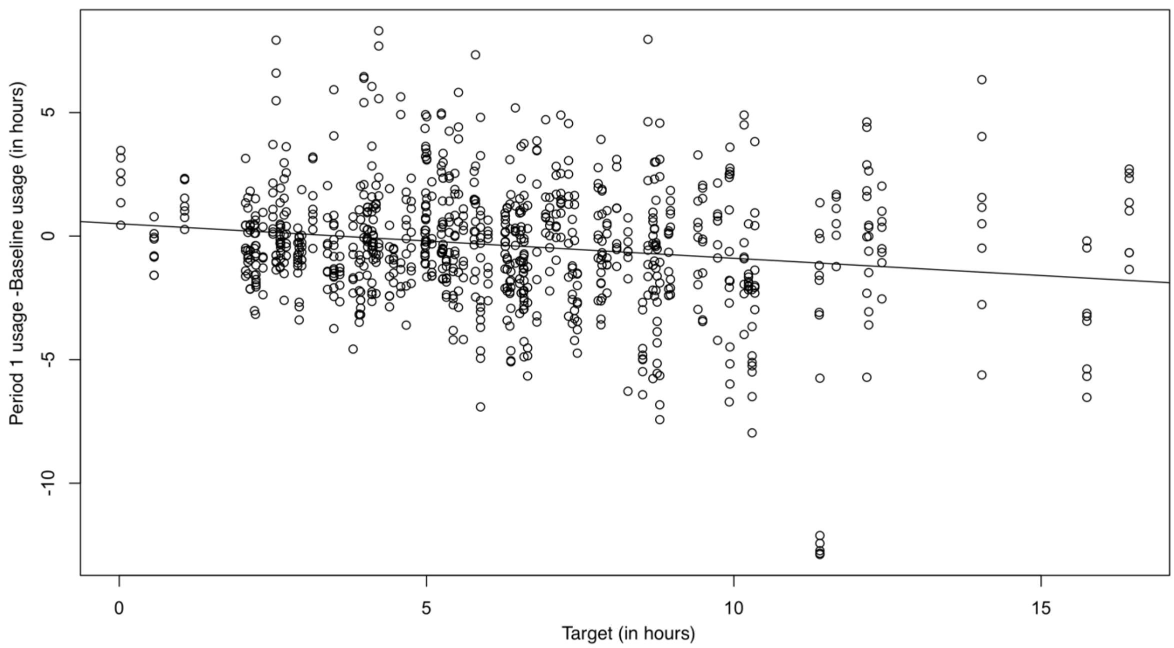


Table W21: AI Condition Participants with Higher Target (and Baseline Usage) Reduce Usage More in Period 1.

	Dependent Variable: Period 1 Usage - Baseline Usage (hours)		
	Beta	Std. Error	p-value
Target in hours	-.127***	.026	.000
Age	.0005	.006	.937
Male	-.004	.171	.982
iPhone	.039	.178	.826
Constant	.357	.358	.321
Observations	365		
R2	.063		
Adjusted R2	.053		
Residual Std. Error	1.614 (df = 360)		
F statistic	6.086*** (df = 4; 360)		

Goal salience measures

Two questions tapping into goal salience were asked in the baseline survey and after participants received the instructions for period 1 in all conditions. Specifically, we asked participants the following question: "*How important is the goal of reducing your smartphone usage to you?*" 1 = not at all important, 7 = extremely important. The results predicting the difference score (subtracting the baseline goal salience from the goal salience measured after the treatment instructions were given) with the conditions and control variables are presented in Table W22. In all treatment conditions, the goal of reducing one's smartphone usage became more important compared to the C condition.

Table W22: Goal Salience Increases in All Treatment Conditions Compared to the Control.

	Dependent Variable: Goal Salience (Treatment – Baseline)		
	beta	Std. error	p-value
Intercept	.474	.304	.120
API	1.558***	.209	.000
AI	.826***	.209	.000
FI	1.224***	.221	.000
Age	.006	.005	.216
Gender (Male)	-.031	.149	.832
iPhone	-.335*	.156	.032
Residual Std. Error	1.417 (df = 363)		
R^2	.151		
Adjusted R^2	.137		
F-statistic	F (6, 363) = 10.79, $p < .001$		

Participants also ranked five goals that people commonly pursue (exercise more, eat healthily, save for retirement, drink less alcohol, reduce phone usage) in order of personal importance. A lower number indicates a higher rank of the goal 'reduce phone usage', and greater importance. We again created a difference score for the 'reduce phone usage' goal subtracting the baseline goal ranking from the goal ranking after receiving the treatment instructions to capture the change in goal ranking. This difference score was negative (M = -.367, SD = 1.056), and significantly different from zero (t (369) = -6.691, *p* < .001), indicating that, across all conditions, the goal of reducing one's smartphone usage significantly increased in the importance ranking from the baseline.

We used regression analysis to predict the difference in goal rank using the conditions, as well as controls (see table W23). We find that, compared to the C condition, the FI condition's goal rank increased (*p* = .011). The API condition's goal rank marginally increased compared to C condition (*p* = .050). The AI condition's goal rank did not increase compared to the C condition (*p* = .413). There was no difference between the AI and API condition's goal rank increase (*p* = .228). Thus, the increase in goal ranking of 'reduce phone usage' cannot fully explain why we find a pre-emptive reduction in the AI condition but not in the API condition in period 1. Overall, these results indicate that goal salience is not a key mechanism explaining participants' pre-emptive usage reduction in the AI condition.

Table W23: Goal Ranking of 'Reduce Phone Usage' Increases in FI and API Condition.

	Dependent Variable: Goal Ranking (Treatment – Baseline)		
	beta	Std. error	p-value
Intercept	-.109	.226	.629
API	-.305	.155	.050
AI	-.127	.156	.413
FI	-.420*	.164	.011
Age	-.001	.004	.821
Gender (Male)	-.130	.111	.241
iPhone	.074	.116	.523
Residual Std. Error	1.051 (df = 363)		
R^2	.027		
Adjusted R^2	.011		
F-statistic	F (6, 363) = 1.671, *p* = .127		

Process measures in the AI and API condition

The following process measure items were asked after the anticipation period 1 (only in the AI and API condition). Responses were given on a 5-point Likert Scale (1 = not at all, 5 = very much).

1. **Practice**: *I wanted to test whether I was capable of reducing my mobile usage before the bonus kicked in.*
2. **Forward-looking habit formation**: *I decided to reduce my mobile usage early, thinking it would make it easier to maintain a lower usage once the bonus kicks in.*
3. **Satiation**: *I wanted to maximize my mobile usage before the bonus kicked in, so I would feel less craving.*
4. **Myopia**: *I can easily lower my mobile usage to achieve the desired reduction in the bonus period – there was no need to prepare beforehand.*

Table W24 presents the regression results predicting period 1 usage, separately, in the AI condition and in the API condition, with the agreement to the process explanations, including controls.

Table W24: Period 1 Usage Is Associated with Agreement to Forward-Looking Habit Formation Explanation in the AI but Not API Condition.

	Dependent Variable: Period 1 Usage (hours)					
	AI Condition			**API Condition**		
	Beta	Std. Error	p-value	Beta	Std. Error	p-value
Intercept	.292	.709	.681	.411	.752	.586
Forward-looking habit formation	-.479**	.168	.006	-.137	.124	.273
Practice	-.164	.167	.330	-.297*	.129	.023
Myopia	-.251*	.125	.048	-.293*	.115	.012
Satiation	.544***	.153	.001	.289*	.121	.019
Age	.015	.009	.129	.013	.010	.196
Gender (Male)	.310	.288	.285	-.022	.276	.938
iPhone	-.475	.308	.126	.135	.283	.633
Residual Std. Error	1.424 (df = 93)			1.363 (df = 98)		
R^2	.276			.164		
Adjusted R^2	.222			.105		
F-statistic	F(7, 93) = 5.087, p < .001			F(7, 98) = 2.753, p = .012		

Device Substitution

In the baseline survey, participants estimated the percentage of their total screen time attributed to different devices (Smartphone, Laptop/PC, TV, Tablet, Other). The same question was repeated in the endline survey, covering the past four weeks. While these results are self-reported, we found no evidence of large-scale substitution. On the contrary, the percentage of screen time attributed to smartphone usage showed a slight increase, $M_{baseline}$ = 51.93, SD = 21.92 vs. $M_{endline}$ = 53.72, SD = 24.27, t(362) = -2.39, p = .017. No significant changes were observed for laptop, TV, or tablet usage, nor did the conditions have a significant impact.

In the endline survey, we further asked participants directly whether, over the past four weeks, they had spent more or less time on devices other than their phone (e.g., laptop, PC, tablet, TV) compared to the four weeks before joining the study. More than half (52%) reported no change in their usage, while 26.52% indicated an increase, and 21.27% reported a decrease. Participants in the treatment conditions were more likely to report spending more time on other devices compared to those in the C condition (C: 13.10%, FI: 30.38%, AI: 32%, API: 29.29%), though these differences were only marginally significant ($\chi^2(6) = 11.05$, $p = .087$). Additionally, we asked participants to estimate how much more or less time they had spent on other devices. On average, participants reported a daily increase of 12.32 minutes, with no significant differences between the conditions.

We also asked participants whether they had spent more or less time on various activities over the past four weeks, including social activities with friends and family, non-screen activities alone, watching TV, and using digital devices other than their phone (1 = a lot less time, 5 = a lot more time). Participants reported spending significantly more time with family and friends ($M = 3.23$, $SD = .76$, $t(361) = 5.87$, $p < .001$) and engaging in non-screen activities ($M = 3.38$, $SD = .85$, $t(361) = 8.50$, $p < .001$, significantly above the scale midpoint 3 = same amount of time). In contrast, time spent watching TV ($M = 2.97$, $SD = .87$, $p = .467$) and on other digital devices ($M = 3.05$, $SD = .89$, $p = .287$) did not significantly differ from the midpoint. We found no significant differences between conditions for time spent with family and friends ($F(3, 358) = 1.68$, $p = .170$) or on other digital devices ($F(3, 358) = .99$, $p = .398$). However, participants in the AI and API conditions reported a significant increase in personal non-screen activities compared to the C condition ($F(3, 358) = 6.33$, $p < .001$). Additionally, participants in the FI condition reported a significant increase in TV watching compared to the C group ($F(3, 358) = 2.90$, $p = .035$).

Overall, these results suggest that while some substitution effects across digital devices may have been occurring, they were not the primary drivers of our findings. Rather, participants appear to have spent more time on alternative offline activities, such as socializing with others or engaging in personal non-screen activities.

Smartphone Addiction Scale

Table W25: Baseline Usage Predicts Higher Perceived Smartphone Addiction at Baseline.

	Dependent Variable: Smartphone Addiction (Baseline)		
	beta	Std. error	p-value
Baseline usage	.036***	.010	.001
Age	-.014***	.002	.000
Male	-.211***	.071	.003
iPhone	-.057	.074	.442
Constant	3.025***	.158	.000
Observations	404		
R2	.119		
Adjusted R2	.11		
Residual Std. Error	.706 (df = 399)		
F Statistic	13.462*** (df = 2, 399) $p < .001$		

Table W26: Period 1 Reduction Predicts Reduction in Perceived Addiction.

	Dependent Variable: Smartphone Addiction (Baseline - Post-treatment)		
	beta	Std. error	p-value
Period 1 usage - baseline	-.044***	.016	.006
Age	-.004	.002	.051
Male	-.073	.053	.172
iPhone	-.086	.056	.125
Constant	.376***	.099	.000
Observations	355		
R2	.041		
Adjusted R2	.030		
Residual Std. Error	.499 (df = 350)		
F Statistic	3.757** (df = 4, 350), $p = .005$		

Table W27: Period 2 Reduction Predicts Reduction in Perceived Addiction.

	Dependent Variable: Smartphone Addiction (Baseline - Post-treatment)		
	beta	Std. error	p-value
Period 2 usage - baseline	-.051***	.013	.000
Age	-.004*	.002	.026
Male	-.083	.053	.118
iPhone	-.084	.055	.118
Constant	.373***	.099	.000
Observations	343		
R2	.064		
Adjusted R2	.053		
Residual Std. Error	.487 (df = 338)		
F Statistic	5.832*** (df = 4, 338), $p < .001$		

Life Satisfaction (Cantril Life Satisfaction Ladder)

Table W28: Period 1 And Period 2 Reduction Does Not Lead to Changes in Life Satisfaction.

	Dependent Variable: Life Satisfaction (Baseline – Post-treatment)					
	Model 1			**Model 2**		
	beta	Std. error	p-value	beta	Std. error	p-value
Period 1 usage - baseline	-.034	.033	.307			
Period 2 usage - baseline				-.003	.211	.269
Age	.004	.004	.341	.002	.004	.569
Male	-.006	.112	.955	.054	.114	.635
iPhone	.168	.116	.148	.099	.118	.402
Constant	-.320	.206	.121	-.234	.211	.269
Observations	336			336		
R2	.011			.003		
Adjusted R2	.000			.000		
Residual Std. Error	1.013 (df = 331)			1.027 (df = 329)		
F Statistic	.963 (df = 4, 331), $p = .428$			.282 (df = 4, 329), $p = .889$		

Web Appendix E: Model Estimation

Model Estimation Procedure

We estimate the habit formation and satiation models using maximum likelihood by assuming normally distributed standard errors ϵ_{ij} and ϵ'_{ij}. Below, we first discuss the estimation of the habit formation model. A similar process is applied for estimating the satiation model. Consider that c_t^{ij} is the period t consumption of a participant i on day j.

$$For\ t = 1,3{:}\ \ c_t^{i,j} = f(\theta_i) + \epsilon_{ij}$$

where f corresponds to the optimal consumption equation in the particular period for the habit formation model. θ_i corresponds to the vector of parameters of the habit formation model (H, γ, α) for a particular participant i and ϵ_{ij} is the independently and normally distributed random variable with mean $E\left(c_t^{ij}\right) = f(\theta_i)$ and $Var\left(c_t^{ij}\right) = \sigma^2$.

The likelihood function given the consumption c_t^{ij} is:

$$L\left(\theta_i \middle| c_{ij}\right) = \Pi_{t=1,3}\Pi_{j=1,\dots,n(t)}\, p\left(c_t^{i,j} \middle| f(\theta_i), \sigma^2\right)$$

$$= \Pi_{t=1,3}\Pi_{j=1,\dots,n(t)}\, \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left(c_t^{i,j} - f(\theta_i)\right)^2/2\sigma^2}$$

where $n(t)$ is the number of observations in period t in the particular RCT. We indicate $N = n(t_1) + n(t_2)$. We take a natural logarithm of the likelihood function to get:

$$\ln\left(L\left(\theta_i \middle| c_{ij}\right)\right) = -\frac{N}{2}\ln(2\pi) - N ln(\sigma) - \frac{\sum_{t=1,3}\sum_{j=1,\dots,n(t)}\left(c_t^{i,j} - f(\theta_i)\right)^2}{2\sigma^2}$$

Finding the first order conditions, the maximum likelihood estimate is given by:

$$\theta_i^{MLE} = argmax_{\theta_i}(\ln\left(L\left(\theta_i \middle| c_t^{ij}\right)\right) = \arg\min_{\theta_i} \sum_{t=1,3}\sum_{\mathrm{j=1,..,n(t)}} \left(c_t^{ij} - \theta_i\right)^2$$

Note that θ^{MLE} is the same as the least square estimate.

To find the maximum likelihood estimate σ^{MLE} of σ, we evaluate $\frac{\partial(\ln(L(\theta_i|c_{ij}))}{\partial\sigma} = 0$ when $\theta = \theta^{MLE}$, and obtain:

$$\sigma^{MLE} = \frac{1}{N}\left(\sum_{t=1,3}\sum_{\mathrm{j=1,..,n(t)}}\left(c_t^{ij} - \theta_i\right)^2\right)$$

Substituting the above into the log-likelihood function, we obtain:

$$\ln\left(L(\theta_i|c_{ij})\right) = -\frac{N}{2}\ln(2\pi + 1) - \frac{N}{2}ln(\sigma) - \frac{\sum_{t=1,3}\sum_{j=1,\dots,n(t)}\left(c_t^{i,j} - f(\theta_i)\right)^2}{N}$$

The AIC is $-2\ln\left(L(\theta_i^{MLE}|c_{ij}^t)\right)$+2(#parameters+1). The lower the AIC, the better the model fits the data.

Individual level estimation

For each participant, we focus on six data points in period 1 and six data points in the post-treatment. We minimize the mean square error by considering the optimal consumption equations in period 1 and the post-treatment. In the optimal consumption equation, we input the observed period 2 consumption for each participant (period 2 is divided into two sub-periods of six days). Note that each participant's period 2 consumption already incorporates the effect of monetary incentives (i.e., if participants found the monetary incentives attractive, then p_t is going to be higher and period 2 consumption is going to be lower). For I, we incorporate the maximum consumption of each participant during the experimental period. The optimization routine (constroptim in R) for each participant is repeated for five different starting values for the parameters to ensure robustness of the fitted values.

Model Estimation Results

Individual level parameter estimates of AI participants in RCT 2

In Table W29, we find that, except for five participants, the habit formation model fits the data better than the satiation model at the individual level.

Table W29: Individual Level Estimates of the Habit Formation and Satiation Model for AI Participants in RCT 2.

	Habit formation model					Satiation model			
#	***H***	***γ***	***α***	**SSE**	**AIC**	***γ′***	***$α_s$***	**SSE**	**AIC**
1	0.14	0.62	0.43	8.94	28.29	0.00	0.27	85.34	53.37
2	0.13	0.03	0.12	11.17	30.96	0.00	0.28	145.23	59.75
3	0.11	0.50	0.46	15.84	35.16	0.00	0.98	83.01	53.04
4	0.13	0.48	0.82	13.96	33.65	0.00	0.98	211.98	64.29
5	0.00	0.00	0.02	24.51	40.40	0.48	1.00	13.22	30.99
6	0.02	1.00	1.00	2.72	14.04	0.00	0.75	8.59	25.81
7	0.14	0.72	0.87	12.06	31.89	0.00	0.96	69.75	50.95
8	0.14	0.93	0.94	6.51	24.48	0.00	0.83	121.24	57.58
9	0.00	1.00	1.00	11.78	31.60	0.00	0.71	27.17	39.64
10	0.40	0.37	0.85	20.23	38.10	0.00	0.97	119.22	57.38
11	0.27	0.45	0.95	26.73	41.44	0.00	0.24	172.83	61.84
12	0.18	0.10	0.05	2.72	14.01	0.00	0.31	32.18	41.66
13	0.00	0.00	0.04	21.58	38.87	0.00	0.73	28.49	40.20
14	0.15	0.00	0.02	18.19	36.82	0.21	1.00	48.23	46.52
16	0.11	1.00	1.00	31.05	43.23	0.00	0.63	74.33	51.71
17	0.05	1.00	1.00	9.65	29.21	0.00	0.95	19.50	35.65
18	0.00	0.00	0.02	11.15	30.94	0.24	1.00	31.73	41.49
19	0.03	1.00	1.00	3.79	17.99	0.00	0.85	104.68	55.82
20	0.43	0.00	0.01	51.63	49.34	0.32	1.00	331.80	69.66
21	0.66	0.11	0.02	22.77	39.51	0.00	0.71	142.05	59.48
22	0.00	0.00	0.01	10.67	30.41	0.11	1.00	10.28	27.97
23	0.08	0.83	0.89	4.68	20.53	0.00	0.26	94.31	54.57
24	0.26	0.84	0.45	13.65	33.37	0.00	0.72	55.54	48.21
25	0.00	0.11	0.06	14.27	33.90	0.22	1.00	138.91	59.21
26	0.00	0.00	0.04	45.29	47.77	0.46	1.00	21.34	36.74
27	0.05	1.00	1.00	6.25	24.00	0.00	0.94	17.12	34.09
28	0.43	0.16	0.06	10.01	29.66	0.00	0.28	13.60	31.33

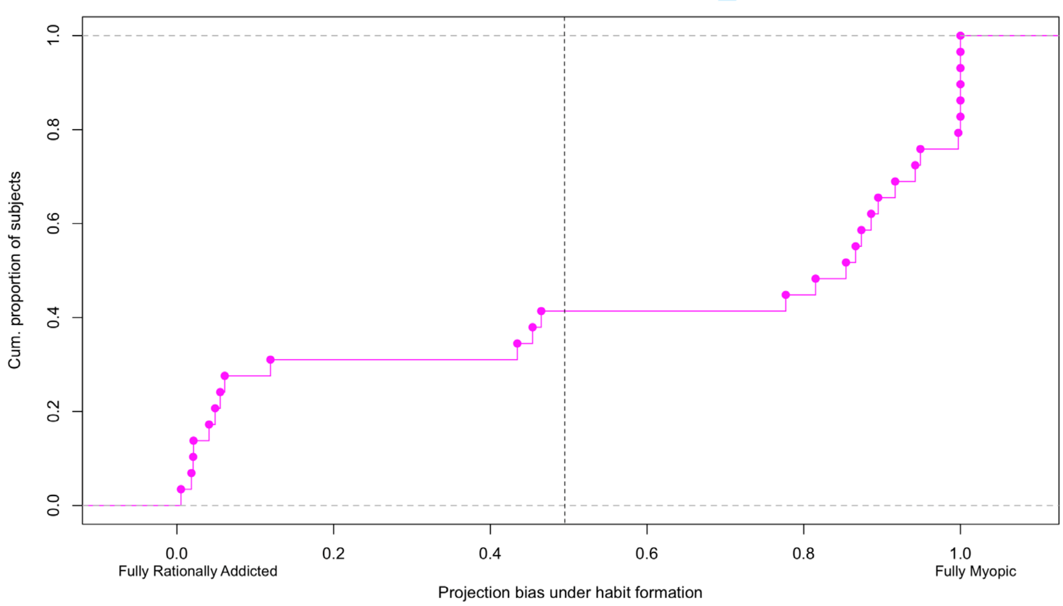
29	0.00	0.05	1.00	5.86	23.22	0.38	1.00	24.99	38.63
30	0.00	0.00	0.09	22.69	39.47	0.52	1.00	9.49	27.01
31	0.08	0.78	0.87	9.53	29.06	0.00	0.30	121.74	57.63
33	0.02	0.38	0.90	3.12	15.66	0.00	0.31	54.44	47.97
34	0.60	0.03	0.80	1.22	4.42	0.00	0.72	1.25	2.64

Notes: Lower AIC between the habit formation and satiation model is indicated in red.

Distribution of projection bias under habit formation (RCT 2)

The habit formation model outperformed the satiation model for 27 out of 32 participants, indicating better fit. For these 27 participants, we plot the distribution of the projection bias parameter (Figure W17). Approximately 37.5% of participants exhibited a projection bias of less than 50%, suggesting that their behavior aligned more with the rational addiction model. On the other hand, 62.5% had a projection bias greater than 50%, indicating behavior more in line with myopic habit formation. Specifically, seven participants displayed fully myopic behavior, while one subject appears to have been fully rationally addicted.

Figure W17: Cumulative Proportion of AI Participants with Different Levels of Projection Bias Under Habit Formation.



Individual level parameter estimates of AI participants in RCT 3

Table W30: Individual Level Estimates of the Habit Formation and Satiation Model for AI Participants in RCT 3.

	Habit formation model					**Satiation model**			
#	$\boldsymbol{H}$	$\boldsymbol{\gamma}$	$\boldsymbol{\alpha}$	**SSE**	**AIC**	$\boldsymbol{\gamma'}$	$\boldsymbol{\alpha_s}$	**SSE**	**AIC**
1	0.22	0.47	0.04	29.03	42.43	0.00	0.83	924.87	81.96
2	0.12	0.80	0.35	7.64	26.41	0.00	0.25	119.25	57.38
4	0.00	0.30	0.62	33.57	44.17	0.54	1.00	46.21	46.01
6	0.00	0.66	0.21	60.30	51.20	0.59	1.00	36.45	43.16
7	0.45	0.00	0.07	44.10	47.44	0.50	1.00	272.85	67.32
10	0.13	0.00	0.61	78.08	54.30	0.57	1.00	147.57	59.94
11	0.17	0.72	0.53	18.65	37.12	0.00	0.18	437.35	72.98
12	0.15	0.76	0.26	36.48	45.17	0.00	0.32	370.26	70.98
13	0.08	0.99	0.45	6.11	23.73	0.00	0.29	126.24	58.07
14	0.05	0.66	0.23	7.54	26.26	0.00	0.99	211.07	64.23
17	0.00	0.70	0.86	546.15	77.64	0.00	0.06	153.62	60.42
18	0.16	0.00	0.59	33.84	44.27	0.41	1.00	61.67	49.47
19	0.18	0.70	0.91	14.32	33.94	0.00	0.18	496.06	74.49
20	0.14	0.69	0.29	17.13	36.10	0.00	0.25	202.20	63.72
21	0.33	0.33	0.79	12.34	32.16	0.00	0.72	340.40	69.97
22	0.30	0.63	0.96	9.52	29.05	0.00	0.83	246.36	66.09
23	0.27	0.36	0.71	31.53	43.42	0.00	0.82	454.17	73.43
24	0.28	0.67	0.99	25.67	40.95	0.00	0.30	268.33	67.11
25	0.00	0.52	0.07	7.26	25.79	0.45	1.00	7.52	24.23
26	0.02	1.00	1.00	30.77	43.13	0.00	0.70	156.49	60.64
27	0.15	0.00	0.03	63.52	51.82	0.57	1.00	118.79	57.34
28	0.10	1.00	0.60	23.44	39.86	0.00	0.38	206.15	63.95
29	0.29	0.57	0.74	62.45	51.62	0.00	0.83	631.01	77.38
30	0.17	0.74	0.44	19.67	37.76	0.00	0.75	431.22	72.81
32	0.00	0.08	0.04	47.29	48.28	0.00	0.24	52.34	47.50
33	0.26	0.00	0.09	110.62	58.48	0.12	1.00	279.20	67.59
34	0.21	0.00	0.17	23.49	39.89	0.23	1.00	147.15	59.91
35	0.30	0.00	0.63	49.60	48.86	0.26	1.00	250.20	66.27
36	0.39	0.48	0.96	15.68	35.03	0.00	0.26	277.11	67.50
37	0.10	0.96	0.76	35.71	44.91	0.00	0.75	417.43	72.42
38	0.02	0.00	0.02	32.86	43.91	0.28	1.00	195.14	63.29

39	0.08	1.00	1.00	53.13	49.68	0.00	0.73	168.81	61.55
40	0.26	0.27	0.73	70.75	53.12	0.00	0.72	376.55	71.18
41	0.17	0.56	0.90	12.72	32.53	0.00	0.25	209.18	64.13
42	0.13	0.93	0.22	98.99	57.15	0.00	0.30	1097.47	84.02
43	0.46	0.33	0.86	46.99	48.21	0.00	0.83	1951.51	90.92
45	0.26	0.00	0.07	172.36	63.80	0.73	1.00	185.23	62.67
46	0.07	0.54	0.01	82.94	55.03	0.00	0.70	211.06	64.23
47	0.33	0.45	0.77	79.80	54.56	0.00	0.27	703.21	78.68
48	0.00	0.10	0.31	88.08	55.75	0.59	1.00	95.19	54.68
50	0.05	0.23	0.99	22.92	39.59	0.00	0.15	166.48	61.39
51	0.04	1.00	1.00	7.60	26.35	0.00	0.95	39.32	44.07
52	0.27	0.01	0.57	36.52	45.18	0.00	0.72	717.97	78.93
53	0.37	0.10	0.05	0.77	-1.09	0.00	0.72	78.87	52.42
54	0.00	0.05	0.01	90.60	56.09	0.51	1.00	44.28	45.49
55	0.20	0.52	0.72	27.93	41.96	0.00	0.25	254.06	66.46
57	0.05	1.00	1.00	10.63	30.37	0.00	0.39	35.78	42.94
58	0.00	0.30	0.61	71.58	53.26	0.38	1.00	152.50	60.33
59	0.00	0.29	0.61	18.74	37.18	0.10	1.00	38.12	43.70
60	0.08	0.00	0.10	2.59	13.45	0.05	1.00	42.19	44.91
61	0.06	0.00	0.01	18.55	37.05	0.61	1.00	53.66	47.80
62	0.21	0.00	0.56	7.39	26.01	0.10	1.00	242.11	65.88
63	0.14	0.87	0.63	38.75	45.89	0.00	0.75	547.63	75.68
64	0.00	0.32	0.61	16.68	35.77	0.26	1.00	50.78	47.14
65	0.04	0.89	0.87	36.52	45.18	0.00	0.73	203.95	63.82
67	0.07	0.98	0.61	30.02	42.83	0.00	0.73	85.19	53.35
68	0.07	0.94	0.38	7.08	25.50	0.00	0.38	204.56	63.86
69	0.09	0.88	0.06	9.49	29.01	0.00	0.75	60.41	49.22
71	0.29	0.00	0.65	35.01	44.67	0.23	1.00	268.13	67.11
72	0.08	0.96	0.45	12.08	31.90	0.00	0.84	159.91	60.90
73	0.13	0.49	0.81	140.52	61.35	0.00	0.18	412.74	72.28
74	0.00	0.65	0.11	35.12	44.71	0.00	0.31	22.14	37.17
75	0.24	0.14	0.65	8.82	28.13	0.00	0.99	116.35	57.09
76	0.16	0.58	0.15	22.15	39.18	0.00	0.27	218.43	64.65
78	0.12	1.00	0.64	9.79	29.39	0.00	0.75	228.25	65.17
79	0.00	0.74	0.20	7.67	26.46	0.00	0.24	16.77	33.84
80	0.19	0.57	0.70	69.97	52.98	0.00	0.18	521.72	75.09
81	0.35	0.00	0.66	40.56	46.44	0.65	1.00	92.43	54.33
82	0.28	0.40	0.81	48.88	48.68	0.00	0.13	609.30	76.96
84	0.13	0.00	0.00	100.25	57.30	0.61	1.00	123.48	57.80

85	0.34	0.07	0.03	49.88	48.92	0.00	0.84	654.67	77.82
86	0.15	0.86	0.62	41.77	46.79	0.00	0.73	452.19	73.38
87	0.02	1.00	1.00	76.62	54.07	0.00	0.26	140.60	59.36
88	0.26	0.41	0.05	21.68	38.93	0.00	0.32	416.41	72.39
89	0.11	0.00	0.00	24.10	40.20	0.05	1.00	90.57	54.08
90	0.12	0.36	0.89	47.04	48.22	0.00	0.85	741.16	79.31
91	0.19	0.15	0.07	8.82	28.13	0.00	0.24	80.83	52.72
92	0.10	0.30	0.10	98.42	57.08	0.00	0.75	227.75	65.15
93	0.19	0.65	0.32	67.07	52.48	0.00	0.30	1264.62	85.72
96	0.34	0.31	0.14	33.99	44.32	0.00	0.72	312.65	68.95
97	0.08	1.00	1.00	102.50	57.57	0.00	0.24	270.70	67.22
98	0.09	0.61	0.95	8.20	27.26	0.00	0.24	24.27	38.28
99	0.27	0.45	0.22	29.55	42.64	0.00	0.27	466.42	73.75
101	0.01	1.00	1.00	19.76	37.81	0.00	0.26	110.99	56.52
102	0.08	1.00	1.00	7.82	26.69	0.00	0.85	33.65	42.20
103	0.34	0.29	0.74	20.83	38.44	0.00	0.16	257.80	66.63
104	0.22	0.00	0.60	152.85	62.36	0.41	1.00	199.10	63.53
105	0.00	0.52	0.07	15.52	34.91	0.18	1.00	23.37	37.83
106	0.00	0.39	0.70	66.18	52.32	0.00	0.70	56.77	48.48
107	0.09	0.11	0.56	92.09	56.28	0.00	1.00	207.11	64.01
108	0.14	0.94	0.58	32.74	43.87	0.00	0.30	304.42	68.63
109	0.20	0.60	0.31	27.56	41.80	0.00	0.26	276.31	67.47
110	0.00	0.24	0.61	30.50	43.02	0.22	1.00	129.96	58.41
111	0.10	0.85	0.29	8.86	28.18	0.00	0.28	414.76	72.34
112	0.21	0.53	0.06	15.67	35.03	0.00	0.37	294.94	68.25
113	0.05	0.96	0.78	31.75	43.50	0.00	0.70	103.52	55.69

Notes: Lower AIC between the habit formation and satiation models is indicated in red.

Distribution of projection bias under habit formation (RCT 3)

The habit formation model outperformed the satiation model for 87 of 96 participants, indicating better fit. For these 87 participants, we plot the distribution of the projection bias parameter (Figure W18). Approximately 36.4% exhibited a projection bias of less than 50%, suggesting that their behavior aligned more with the rational addiction model. On the other hand, 63.6% had a projection bias greater than 50%, indicating that their behavior was more in line with myopic habit formation. Specifically, nine participants displayed fully myopic behavior with respect to habit formation, while one participant appears to have been fully rationally addicted.

Figure W18: Cumulative Proportion of AI Participants with Different Levels of Projection Bias Under Habit Formation in RCT 3.

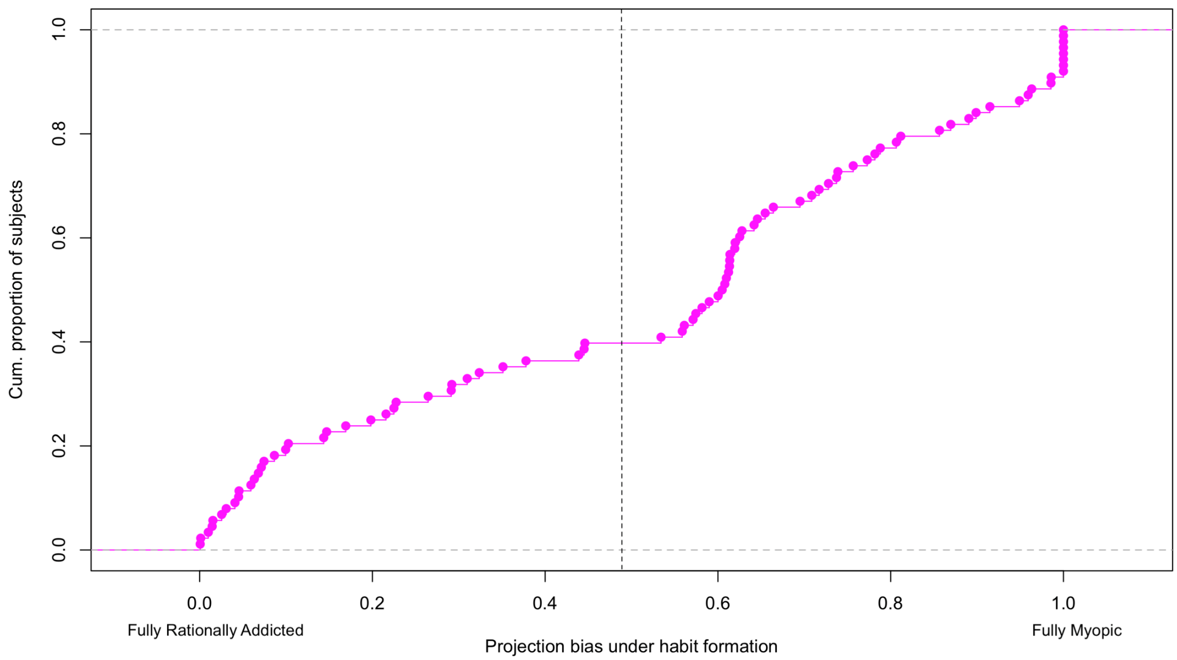


Table W31: Summary of Individual Level Estimates of the Satiation Model.

Satiation Model	**RCT 2**		**RCT 3**	
	Mean (SE)	**Median**	**Mean (SE)**	**Median**
$\boldsymbol{\gamma'}$ (Habituation retention factor)	.08 (.03)	~0	.11 (.02)	~0
$\boldsymbol{\alpha_s}$ (Projection bias)	.70 (.05)	.76	.65 (.028)	.75
SSE (Sum of square errors)	81.53		278.61	
AIC (Akaike information)	48.02		62.103	

Aggregate-level estimation

At the aggregate level, we do not have a good proxy for I. As a result, to estimate the model at the aggregate level, as in Allcott, Gentzkow, and Song (2022), we consider the treatment effect or the change in consumption during period 1, period 2 and post-treatment in the AI condition vis-a-vis C condition. The difference between the optimal consumption in the AI and C condition helps to net out I from the empirical Equations 3 and 4. Indicating optimal consumption in period i in the C condition using c_i and in the AI condition using c_{Ai}, the new empirical equations are as follows:

- Period 1: $c_{1Aij} - c_1 = H(k_{A1i} - k_{C1i}) + (1-\alpha)H\gamma(c_{2Ai} - c_{2i}) + (1-\alpha)\gamma^2 H(c_{3Ai} - c_{3i}) + (1-\alpha)\gamma^3 H(c_{4Ai} - c_{4i}) + \epsilon_{ij}$ (3)
- Post-treatment: $c_{4Aij} - c_1 = H \times (\gamma(c_{3Ai} - c_{3i}) + \gamma^2 c_{2Ai} - c_{2i}) + \gamma^3(c_{1i} - c_{1Ai} + k_{1i} - k_{1Ai}))) + \epsilon_{ij}$

We consider the difference in optimal consumption in the AI and C condition in period 1 and the post-treatment. We consider six days of period 1 and six days of the post-treatment for all 113 AI participants (a total of 1,376 data points). We plug in the difference in period 2 consumption between the AI and C condition to the optimal consumption equations to estimate the parameters. The aggregate level parameters are consistent with the individual level parameters. The projection bias is .47. Based on the AIC criterion, the habit formation model fits the data better than the satiation model.

Table W32: Aggregate Level Estimation of the Habit Formation Model in RCT 3.

Habit formation model	
H (adjacent complementarity)	.61
$\boldsymbol{\gamma}$ (Habituation retention factor)	.64
$\boldsymbol{\alpha}$ (Projection bias)	.57
SSE (Sum of square errors)	5752
AIC (Akaike information crit.)	4659

Table W33: Aggregate Level Estimation of the Satiation Model in RCT 3.

Satiation model	
$\boldsymbol{\gamma'}$ (Satiation retention factor)	~0
$\boldsymbol{\alpha_s}$ (Projection bias)	0.43
SSE (Sum of square errors)	22299.77
AIC (Akaike information)	6495.28

Robustness Checks

For robustness checks, we exclusively focus on the RCT 3 data.

Discount factor $\delta = .9$

We estimate the habit formation model at the aggregate and individual level when the discount factor is .9. The projection bias parameter tends to become smaller. Almost 45% of participants had a projection bias parameter smaller than .5 (or were more rationally addicted).

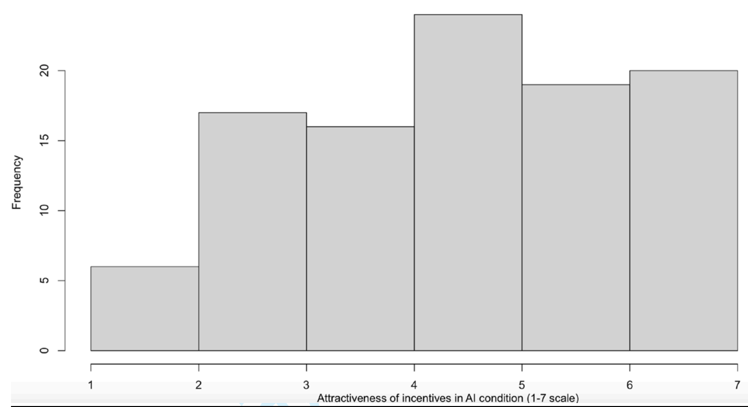
Table W34: Aggregate and Individual Level Estimate of the Habit Formation Model in RCT 3 with δ = .9.

Habit formation model	**RCT 3 with delta = 0.9**		
	Aggregate level	**Individual level**	
		Mean (SE)	Median
H (adjacent complementarity)	.61	.21 (.16)	.17
$\boldsymbol{\gamma}$ (Habituation retention factor)	.65	.42 (.33)	.39
$\boldsymbol{\alpha}$ (Projection bias)	.47	.47 (.35)	.53
SSE (Sum of square errors)	5736.07	45.38	
AIC (Akaike information)	4656.103	33.01	

Focusing only on participants who perceive monetary incentives to be highly attractive

We find that 85 out of 113 participants in the AI condition find the monetary incentive attractive (rating ≥ 4 on a 7-point Likert scale).

Figure W19: Histogram of Attractiveness of Monetary Incentives.



When we focus on the 85 participants with a high attractiveness rating of the monetary incentives, we observe that the parameter estimates are similar to the numbers in Table 7 in the main manuscript. However, the projection bias is slightly higher, although the difference is not statistically significant.

Table W35: Aggregate and Individual Level Estimate of the Habit Formation Model in RCT 3 for those with High Monetary Incentive Attractiveness (Rating ≥ 4 on a 1–7 Scale).

Habit formation model			
	Aggregate level	**Individual level**	
		Mean (SE)	Median
H (adjacent complementarity)	.71	.37 (.02)	.5
$\boldsymbol{\gamma}$ (Habituation retention factor)	.59	.28 (.03)	.17
$\boldsymbol{\alpha}$ (Projection bias)	.67	.67 (.04)	.77
SSE (Sum of square errors)	4040.5	47.5	
AIC (Akaike information)	3437.83	44.35	

Allowing for habit formation and satiation in the same model

We can incorporate both habit formation and satiation in the same model by allowing b_2 in Equation 1 to take values above and below zero. If $\mathrm{b}_2 < 0$, the stock of past consumption reduces marginal utility of current consumption (satiation). If $b_2 > 0$, the stock of past consumption increases marginal utility of current consumption (habit formation). In the empirical Equations 3 - 4, this implies that $H > 0$ is habit formation and $H < 0$ is satiation. In RCT 3, we find that 95 out of 113 participants (84%) have H > 0, exhibiting habit formation. On the other hand, 18 out of 113 participants (16%) have H < 0, therefore exhibiting satiation. The average H > 0 (p <.001), thus habit formation is the dominant phenomenon.

Table W36: Individual-Level Results of Fitting Habit Formation and Satiation in the Same Model (RCT 3)

Model parameters	**Mean (Std. error)**	**Median**
H (adjacent complementarity)	.17 (.020)	.16
$\boldsymbol{\gamma}$ (Habituation retention factor)	.47 (.034)	.50
$\boldsymbol{\alpha}$ (Projection bias)	.61 (.03)	.75
SSE (Sum of square errors)	32.78	
AIC (Akaike information)	38.71	

Fitting the model at the individual level in the API condition

We find that the estimates in the API condition are similar to those in the AI condition. This suggests that participants in the API condition may have followed a strategy to reduce their usage by making adjustments to their usage within period 2.

Table W37: Individual-Level Results of Fitting Habit Formation in the API Condition.

Model parameters	Mean (Std. error)	Median
H (adjacent complementarity)	.21 (.014)	.17
$\boldsymbol{\gamma}$ (Habituation retention factor)	.49 (.03)	.51
$\boldsymbol{\alpha}$ (Projection bias)	.56 (.03)	.64
SSE (Sum of square errors)	39.32	
AIC (Akaike information)	38.71	

References

Allcott, H., Gentzkow, M., & Song, L. (2022). Digital Addiction. *American Economic Review*, *112*(7), 2424-2463.

Baucells, M., & Sarin, R. K. (2007). Satiation in discounted utility. *Operations research*, *55*(1), 170-181.

Baucells, M., & Sarin, R. K. (2010). Predicting utility under satiation and habit formation. *Management Science*, *56*(2), 286-301.

Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. *Journal of political Economy*, *96*(4), 675-700.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives, 19* (4), 25-42.

Gruber, J., & Köszegi, B. (2001). Is addiction "rational"? Theory and evidence. *The Quarterly Journal of Economics*, *116*(4), 1261-1303.

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics, 107* (2), 573-597.

Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *the Quarterly Journal of economics*, *118*(4), 1209-1248.

Ward, A. F., Duke, K., Gneezy, A., & Bos, M. W. (2017). Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research, 2* (2), 140-154.