Why Don't We Sleep Enough?
A Field Experiment Among College Students

Mallory Avery
University of Pittsburgh*
Osea Giuntella
University of Pittsburgh and IZA†
Peiran Jiao
Maastricht University and Nuffield College ‡

May 4, 2020

Abstract

Sleep deprivation is prevalent in modern societies leading to negative health and economic consequences. However, we know little about why people decide to sleep less than the recommended number of hours. This study investigates the mechanisms affecting sleep choice and explores whether commitment devices and monetary incentives can be used to promote healthier sleep habits. To this end, we conducted a field experiment with college students, providing them incentives to sleep and collecting data from wearable activity trackers, surveys, and time-use diaries. Our results are consistent with partially sophisticated time-inconsistent preferences and overconfidence. The subjects in the treatment group responded to the monetary incentives by significantly increasing the likelihood of sleeping between 7 and 9 hours (+19%). We uncover evidence of demand for commitment. Overall, 63% of our subjects were sophisticated enough to take up commitment, and commitment improved sleep for the less overconfident among them. Using time-use diaries, we show that during the intervention, there was a reduction in screen time near bedtime (-48%) among treated subjects. Individuals in the treatment group were less likely to sleep insufficiently than at baseline even after removal of the incentive (-16%), which is consistent with habit formation. Finally, our treatment also had positive (albeit small) effects on health and academic outcomes. Our results have implications for the effectiveness of information interventions on sleep in the field and the possibility of improving sleep by providing proper incentive and commitment devices.

JEL Codes: B49, C93, I1

Keywords: Sleep, Health Behaviors, Sophistication, Present Bias, Habit Formation, Incentives

*University of Pittsburgh. Department of Economics. Email: mla49@pitt.edu.

†University of Pittsburgh. Department of Economics. Email: osea.giuntella@pitt.edu.

‡Maastricht University. Department of Finance. And Nuffield College. Email: p.jiao@maastrichtuniversity.nl. We thank participants to seminars at the University of Pittsburgh, University of Rome 2 Tor Vergata, University of Milano Statale, Maastricht University, the Nuffield Centre for Experimental Social Sciences, the Pittsburgh Experimental Economics Lab, the Circadian Rhythms and Sleep Conference Grands, the Advances in Field Experiments Conference, the Workshop on Risky Health Behaviors, the American-European Health Study Group, the IV Workshop on Behavioral and Experimental Health Economics, the Center for Health Incentives and Behavioral Economics (CHIBE) Roybal Retreat, and the Maastricht Behavioral and Experimental Economics Symposium. We benefited from discussion with Domenico De Palo, David Dickinson, Daniel Hamermesh, Ben Handel, David Huffman, Gautam Rao, Silvia Saccardo, Frank Schilbach, Sally Sadoff, Heather Schofield, Jeffrey Shrader, Justin Sydnor, Severine Toussaert, and Lise Vesterlund.

1 Introduction

Sleep deprivation is an emerging public health challenge. According to the Center for Disease Control and Prevention, more than a third of American adults sleep less than the recommended minimum of seven hours (Liu, 2016). Some scholars consider it the most prevalent risky behavior in modern societies and evidence suggests that in many countries people may be sleeping between one and two hours less than what their ancestors used to sleep a hundred years ago (Roenneberg, 2013). Growing evidence documents the causal effects of sleep deprivation on chronic diseases, health, cognitive skills, decision making, human capital, and productivity (McKenna et al., 2007; Luyster et al., 2012; Jin et al., 2015; Giuntella et al., 2017; Smith, 2016; Hafner et al., 2017; Gibson and Shrader, 2018; Heissel and Norris, 2018; Giuntella and Mazzonna, 2019). Firms, athletes, and military training programs increasingly recognize how sleep deprivation can impair performance.[1]

Despite sleep being increasingly recognized as a fundamental contributor to health and human capital and despite economists' interest in time-use (Becker, 1965; Aguiar and Hurst, 2007; Aguiar et al., 2013; Hamermesh, 2019), sleep behavior has received little attention in the economic literature. Given that we spend approximately a third of our time –one of our scarcest resources– sleeping, and given the substantial economic and health impacts of sleep deprivation, sleep behavior should be an object of natural interest to economists (Mullainathan, 2014). However, most economic models analyzing time allocation regard sleeping as a pre-determined and homogeneous constraint on time allocation. While for some individuals sleep duration and quality are influenced by medical conditions (insomnia, sleep apnea etc.), for most individuals bedtime and sleep duration are a choice. Individuals may optimally allocate less time to sleep and delay their bedtime (or anticipate their wake-up time) to work longer or enjoy more leisure. And indeed, the few pioneering studies analyzing sleep choice have assumed individuals choose hours of sleep optimally (Biddle and Hamermesh, 1990). Yet, according to the Royal Philips global sleep survey, 8 in 10 adults worldwide want to improve their sleep and a poll from YouGov suggests

[1]Recently Aetna, an American managed health care company, introduced incentives to increase workers' sleep (see https://www.cnbc.com/2016/04/05/why-aetnas-ceo-pays-workers-up-to-500-to-sleep.html). Concern has been raised regarding sleep deprivation among NBA players (see https://www.espn.com/nba/story/_/id/27767289/dirty-little-secret-everybody-knows-about). Finally, sleep with physical activity and nutrition is also one of the three pillars of the army performance triad (see https://armymedicine.health.mil/Performance-Triad).

that, while 89% of Americans would like to sleep for 7 hours or more each night, more than 40% report to sleep less than that.[2]

Sleep decisions may be characterized by dynamic inconsistency, as delaying bedtime may have immediate benefits (i.e., the utility from watching a further episode of a TV series, or working an extra hour), but delayed costs (i.e., the lack of energy or alertness following a night of poor sleep). This suggests that there may be scope for incentives and commitment devices to promote optimal behavior (O'Donoghue and Rabin, 2003; O'Donoghue et al., 2006; O'Donoghue and Rabin, 2015). The effectiveness of incentives and commitment devices to promote optimal choices in the presence of self-control problems has been analyzed in the context of other health behaviors such as alcohol consumption, unhealthy eating, and exercising (O'Donoghue and Rabin, 2006; Charness and Gneezy, 2009; Volpp et al., 2009; Just and Price, 2013; Acland and Levy, 2015; Royer et al., 2015). However, sleep is a particularly interesting domain in which to investigate the prevalence and persistence of behavioral biases. Firstly, it is an activity that people engage in every day, and about which they have received repeated feedback throughout their lives. Thus, sleep is a domain wherein demand for commitment might be highly relevant. This is because daily experience and feedback with time-inconsistent behavior may raise some individuals' awareness of their problem (Laibson, 1997; O'Donoghue and Rabin, 1999, 2001), and increase demand for commitment (Rabin et al., 1999; DellaVigna and Malmendier, 2006; Laibson, 2015; Schilbach, 2019). Sophisticated individuals may learn to restrict their future choice set without receiving any compensation (e.g., Ashraf et al., 2006; Dupas and Robinson, 2013; Kaur et al., 2015; Toussaert, 2018) or even at a cost (e.g., Milkman et al., 2013; Casaburi and Macchiavello, 2019). Secondly, sleep is also an interesting domain in which to study biased beliefs. If people are persistently overconfident about their own sleep even in the face of extensive experience and feedback (Huffman et al., 2018), this can help us disentangle different mechanisms of biased belief. For instance, if individuals have biased recall of own sleep and fail to debias in the face of continuous information, this may be consistent with a motivated reasoning perspective (Bénabou and Tirole, 2016) rather than misinformation suggesting the potential of using incentives to mitigate the bias (Zimmermann, 2019).

[2] See `https://www.usa.philips.com/c-e/smartsleep/campaign/world-sleep-day.html` and `https://today.yougov.com/topics/health/articles-reports/2019/03/13/sleep-habits-americans-survey-poll`.

This study investigates sleep choice and the role of commitment devices and monetary incentives to promote healthier sleep habits. We conducted a field experiment among college students and collected data from wearable activity trackers, surveys, and time-use diaries. Eliciting preferences and randomizing incentives to go to bed earlier and sleep longer, we shed light on the role of present bias, biased beliefs, demand for commitment, and habit formation in sleep.

While sleep deprivation is a problem for many age groups, there are several reasons for sleep deprivation and sleep choice being of particular interest for college students. First, time management is a major challenge among college students transitioning from high school and home habits to campus life (Misra and McKean, 2000; Trockel et al., 2000). Second, sleep deprivation among college students is increasingly becoming a reason for concern. According to recent statistics published in a report of the National Institute of Health (Hershner and Chervin, 2014), more than 70% of college students sleep less than eight hours a day, 60% say they are "dragging, tired, or sleepy" at least three days a week, and more than 80% say loss of sleep affects their academic performance. Third, sleep deprivation and poor sleep quality has been associated with various aspects of undergraduate mental health (Milojevich and Lukowski, 2016), including symptoms of psychological distress, anxiety, attention deficit, and depression problems (McEwen, 2006; Kahn-Greene et al., 2007).[3] Fourth, college is also a crucial phase to shape one's lifestyle and habits (Buboltz et al., 2001). Indeed, Giuntella et al. (2019), who investigate the age-sleep profile, document that during college years, sleep duration markedly declines before reaching a minimum in the early forties. Fifth, college students are a group that is physically healthier, with fewer social and familial constraints and with more time flexibility, suggesting that this is an appropriate group for our experimental study of sleep choice. Additionally, understanding the behavioral mechanisms behind sleep choice within this population may help design educational programs and interventions aimed at improving sleep duration and quality, with non-negligible effects on students' mental health and with potential long-lasting effects on both habits and health.

We recruited 319 participants at the University of Oxford (163 subjects) and the University of Pittsburgh (156 subjects). The subjects were given wearable devices (Fitbit) to collect data on

[3]This is of particular concern, given that depression, anxiety, and suicide rates are rising among US college students (Eisenberg et al., 2013; Mortier et al., 2018; Liu et al., 2019). Reetz et al. (2014) report that 95% of college counseling center directors said that the number of students with significant psychological problems is a growing concern in their center or on campus. Anxiety was found to be the top concern among college students (41.6%), followed by depression (36.4%).

their sleep, physical activity, and heart rate for 8 weeks. In the incentive treatments, subjects set bedtime and sleep duration targets for themselves each Monday of the three treatment weeks and were rewarded for each night (Monday through Thursday) that both targets were achieved based on Fitbit data. We elicited subjects' time and risk preferences in the lab, and integrated the data collected from wearable devices with weekly surveys, time-use diaries, and a follow-up survey conducted three months after the end of the experiment to examine how behavioral biases, such as present bias and overconfidence, affect sleep choice.

Our monetary incentives were effective in improving sleep behavior. The participants responded to monetary incentives by sleeping longer. They were 19% more likely to sleep the recommended number of hours (between 7 and 9) and 23% less likely to sleep less than 6 hours. This finding is robust to the inclusion of individual fixed effects, accounting for time-invariant individual heterogeneity. Furthermore, we document a persistent improvement in sleep. Even after the intervention was removed, the subjects were 16% less likely to sleep less than 6 hours. Our intervention also had effects on sleep regularity, reducing sleep, bedtime, and (more weakly) wake-up time variance. Additionally, as sleep deprivation has been linked to detrimental effects on health and human capital, we explore the potential indirect effects of our intervention on health and academic achievement. We find suggestive evidence that our intervention improved heart rate efficiency, physical activity, and self-reported health, although the effects are relatively small. There is also evidence of positive effects on academic achievement.[4]

We uncover evidence that the subjects voluntarily opted for commitment devices in the form of more demanding targets and dominated incentive schemes. Our findings are consistent with partially sophisticated time inconsistency and biased beliefs as key behavioral mechanisms underlying poor sleep choices. In total, 63% of our subjects took up some form of commitment. More present-biased subjects reported less sleep at baseline and were more likely to take up commitment devices (+28%). Among present-biased individuals, commitment devices reduced insufficient sleep by at least 25%. Meanwhile, many subjects were overconfident about their own achievement rates, over-remembered their own bedtime and sleep duration, over-placed

[4] Our evidence on the effects of sleep on health and academic achievement adds to the growing literature analyzing sleep behavior and its effects on human capital and health (Jin et al., 2015; Giuntella et al., 2017; Gibson and Shrader, 2018; Heissel and Norris, 2018; Jagnani, 2018; Giuntella and Mazzonna, 2019). While these studies used quasi-experimental variation, we exploit a unique field experiment setting, which provides us greater control, to identify the relationship between more sleep and better health and academic outcomes.

their own sleep duration and quality among peers, and understated personal risk associated with sleep deprivation relative to the risk they predicted for peers. Overconfident subjects were more likely to be sleep deprived at baseline and selected overly optimistic targets. Present-biased individuals were more likely to achieve their targets if they were less overconfident.

Given that the subjects positively responded to our incentive to sleep, a natural question is how the subjects reallocated their time to achieve their targets. To address this question, we collected time-use diaries before, during, and after the intervention, and examined how individuals in the treatment group allocated their time when receiving incentives to go to bed earlier and sleep longer. We find no evidence of significant changes in time spent on studying, working, personal care activities, exercising, or socializing. The only activity that systematically and significantly declined during the intervention was screen time (i.e., watching TV, videos etc.). Interestingly, we show that among those who complied with the treatment, evening screen time (after 8 pm) declined by 48% during the intervention with respect to baseline, and by about 28% after the incentive was removed. We see these results as particularly noteworthy given the growing evidence that digital temptations and the use of blue light technologies near bedtime severely impair sleep (Nie and Hillygus, 2002; Twenge et al., 2017; Billari et al., 2018). Consistent with the evidence that repetition of behavior, such as following fixed routines, increases habit formation (e.g., Wood and Neal, 2007; Lally et al., 2010), adjusting activities before bedtime may help develop better sleep habits.

We directly relate to recent studies analyzing the effects of wearable technology on sleep and health behavior (Patel et al., 2015; Jakicic et al., 2016). Handel and Kolstad (2017) exploit a large-scale intervention in a firm to randomize subjects into treatments to improve sleep and exercise through planning. They find small effects of accessing planning tools. Our findings suggest in the presence of persistent behavioral biases that the introduction of monetary incentives and commitment devices may be more effective than using planning tools alone. Bessone et al. (2018) randomize incentives to sleep longer to analyze the effects on labor market productivity and health in a developing country, finding little evidence of an impact of sleep on short-run economic outcomes, but significant effects of naps on attention and well-being. While their main goal was to induce exogenous variation in sleep (and naps) to assess its effects on human capital and productivity, our study focuses directly on the mechanisms behind sleep choice. Furthermore, the differences in the contexts, sleep conditions (e.g., quality of mattress, noise), and samples are likely to explain the different results found in the two studies when examining the effects of sleep on health and human capital. Finally, by following individuals for eight weeks and surveying them three months after the end of the experiment, we are the first to examine persistence and habit formation effects in the context of sleep decisions. Our findings regarding habit formation are at least twofold. First, we find that sleep changes are persistent within the duration of our experiment after treatment ended, suggesting that temporary incentives could lead to long-run lifestyle changes in the sleep domain.[5] Second, treated subjects' adjustment of screen time and evening routines could imply longer-term habit formation effects.

Our study also contributes to the literature analyzing demand for commitment and the effectiveness of commitment devices (see, e.g., Bryan et al., 2010; Kremer et al., 2019; Schilbach, 2019, for a review). To date, little evidence exists in the sleep domain regarding the effectiveness of commitment devices in improving sleep habits. The evidence on the effectiveness of commitment devices is mixed (Laibson, 2015, 2018). Some studies support the idea that commitment devices may help sophisticated agents with present bias mitigate their future self-control problems (Ashraf et al., 2006; Kaur et al., 2015; Schilbach, 2019). Others argue that uncertainty could undermine the demand for commitment (Laibson, 2015), and that, unless subjects are sufficiently sophisticated, commitment devices may be welfare reducing (Bai et al., 2017). Furthermore, in a recent work, Carrera et al. (2019b) show that commitment contract take-up may reflect, at least in part, demand effects or "noisy valuation" when there is substantial uncertainty about the desirability of an activity, even if subjects are time consistent. However, the continuous experience and immediate feedback that characterize sleep behavior suggest commitment devices may be more effective in this domain. Our experiment provides a relatively soft commitment device in the form of setting bedtime and sleep duration targets, at the cost of forgone rewards.[6] Additionally, we elicit time preferences in incentivized tasks and then measure sophistication in subjects'

[5] Breig et al. (2018) also consider sleep in a study using wearable devices. However, their main focus is task allocation. In a 2-week experiment, they randomize feedback on subject's time allocation and explore how that affects their time use in the following week. Their findings show the role for over optimism in time allocation decisions. Our focus is instead sleep, and we conduct an eight-week field experiment to analyze the effects of randomized incentives to sleep, their effects on time use, and shed light on the role of demand for commitment, overconfidence, and habit formation in the sleep domain.

[6] Previous evidence also suggests that softer commitments may work better than hard commitments (Dupas and Robinson, 2013).

belief about own future performance. We find evidence not only for sufficiently sophisticated individuals taking up commitment but also for the positive effects of commitment on behavior, especially for those with less biased beliefs. Our evidence is consistent with previous findings of partial sophistication: a substantial fraction of people exhibit demand for commitment and only a share (approximately 50%) are successful in reaching their target.

The rest of the paper is organized as follows. Section 2 describes the experimental procedure and the design of our intervention. The data are presented in Section 3. In Section 4, we discuss the role of present bias, overconfidence, and risk preferences in the sleep domain. In Section 5, we present the results of our randomized experiment, discuss the effectiveness and persistence of incentives to sleep, and their effects on time allocation, academic outcomes, and health. Concluding remarks are provided in Section 6.

2 Experimental Procedure, Design, and Data

2.1 Experimental Procedure

The experiment was conducted at the Centre for Experimental Social Sciences (CESS) in Nuffield College, Oxford, UK and at the Pittsburgh Experimental Economics Laboratory (PEEL) at the University of Pittsburgh.[7] The experimental procedure was approved by the Central University Research Ethics Committee of the University of Oxford, the ethical review committee of CESS, and the University of Pittsburgh Institutional Review Board. All subjects provided informed consent before participating in the experiment. The participants were given sufficient information about the nature and tasks of the experiment. In order to mitigate experimenter demand effect, we did not specify our focus as on sleep but framed it broadly as a study about the use of wearable devices. During the experiment, the participants were free to withdraw at any time without penalty.

The results reported in this paper were derived from five waves of experimental sessions. The first three waves were conducted in Oxford: the first from October to December 2016; the second from April to June 2017; and the third from October to December 2017. These periods correspond to the Michaelmas Terms of the 2016–17 and 2017–18 and the Trinity Term of the

[7] In all our estimates, we include dummies for whether the subject was recruited in Oxford or in Pittsburgh.

2016–17 academic years at the University of Oxford, respectively.[8] The fourth and fifth waves were conducted at the University of Pittsburgh between mid-January and mid-March during the Spring semester of the 2017–18 academic year and between mid-September and mid-November during the Fall semester of 2018–19 academic year, respectively. The experiment was first advertised in the University of Oxford and on the campus of the Oxford Brookes University in the Oxford waves (1–3) and on the University of Pittsburgh campus in the Pittsburgh waves (4–5). Interested participants then signed up on our recruiting website. The participants in all the waves were recruited through the Online Recruiting System for Economic Experiments (Greiner, 2015) at CESS and the SONA online management system at PEEL, respectively.

Each wave was conducted over eight weeks (which in Oxford coincided with the length of the academic term). Recruitment occurred a week before the beginning of the experiment (Week 0). In Week 1, the subjects were invited to the lab for an experimental session and were given a Fitbit Charge HR device. We collected baseline data from Fitbit devices for the first two weeks. Experimental surveys and treatments started on Monday morning of Week 3, and all participants' Fitbit data were monitored until the end of Week 8. On Friday of Week 8, the participants returned the devices and received final payments. A show-up fee of GBP 4 ($\approx$ USD 5.3) in Oxford or USD \$6 in Pittsburgh was given both in the Week 1 lab session and when they returned the Fitbit, regardless of their performance in the experiment. Among subjects who successfully completed all parts of the experiment, a lottery was drawn, and 3% of the subjects each won a reward of GBP 150 ($\approx$ USD 199) in Oxford or USD \$200 in Pittsburgh.

During the lab session in Week 1, subjects were given an oral description of the experiment, including the exclusion criteria, before their consent was sought. This lab session was divided into three parts. The first part was an incentivized elicitation of risk and time preferences using multiple price lists, each comparing two options. The subjects needed to make one choice on each list: at which row they would switch from choosing Option A to choosing Option B. We elicited risk preferences using two price lists, each comparing a fixed lottery with various certainty amounts. We elicited time preferences using four price lists, each comparing different sooner payments with a fixed future payment. We varied both the size and timing (immediately or in 4 weeks) of the sooner payments as well as the gap between the sooner and later payments

[8] All our estimates include controls for the wave of the experiment.

(4 or 8 weeks). Finally, one choice was selected from each preference elicitation to determine payments. The risk preference task was paid at the end of the lab session. All time preference payments were made in the form of a gift card sent to the participants' email address. To jointly elicit time and risk preferences using multiple price lists, we adopt a similar method to those in, for instance, Tanaka et al. (2010), Falk et al. (2016), and the Double Multiple Price List in Andreoni and Sprenger (2012).[9] For details, please see Appendix B.

The second part of the lab session involved several survey items, which elicited details on subjects' demographics, health conditions, cognition, lifestyle, health behaviors, and physical activity. Additionally, a survey measure of domain-specific risk attitudes (Weber et al., 2002) was implemented, which included a health domain. One part of the survey was about sleep. We asked participants about their sleep habits before the experiment and their general knowledge about the negative consequences of bad sleep habits. We then let them read a short paragraph on the medical evidence of the negative consequences of sleep deprivation, which was to remove any misinformation about this. After that we asked them to evaluate the probability of suffering negative health effects due to sleep deprivation for themselves and for others. The questions about self and others were kept distant from each other and were framed in different ways to encourage subjects to think about the questions independently. The survey contained similar sections on other health behaviors. To minimize the experimenter demand effect, in all our surveys we included questions on physical exercise, health and mental health.

In the third part of the lab session, each participant was given a Fitbit Charge HR device, registered for a Fitbit account, which was linked with Fitabase for data collection. The device was then synchronized with the account.[10] They were asked to wear the device as much as possible including during sleep, to charge and to synchronize the device regularly, and to return the device on or soon after the Friday of Week 8.

[9] These are based on risk preference elicitation in Holt and Laury (2002) and time preference elicitation in Harrison et al. (2002). While there are more sophisticated ways to elicit time preferences, we chose a relatively simple and easy task to reduce burden on subjects and shorten the duration of the lab session.

[10] Fitabase is a paid service that collected Fitbit data from our subjects.

2.2 Experimental Design

Subjects were assigned to a control condition or one of the three different treatment conditions. In the control group, participants were asked only to wear the device, allowing their Fitbit data to be recorded, and to respond to surveys during the experimental period. Control group participants received two types of surveys. One was a Weekly Survey, sent on the Monday of each week, which asked subjects about their health, activities and sleep in the previous week. Second, we also surveyed subjects on their time use. On two randomly chosen mornings of each week, subjects were asked to fill in diaries recalling what they did during the previous day in 30-minute intervals. Participants could select the activity for each time slot from a drop down menu of categories (e.g., sleeping, grooming, watching TV, surfing the Internet, playing games, working, studying, preparing meals or snacks, eating or drinking, cleaning, laundry, grocery shopping, attending religious service, hanging out with friends, paying bills, exercising, commuting [bus/train], commuting [walk or bike]). Subjects were permitted to not respond if they felt uncomfortable.

Additionally, subjects in the three treatment conditions also completed sleep incentive surveys in the treatment weeks: as part of the Weekly Survey, treated participants were asked to choose a bedtime target (between 10 pm and 1 am) and a sleep duration target (between 7 and 9 hours) for Monday through Thursday nights of the current week and received incentives for achieving the targets.The sleep duration targets were set between 7 and 9 hours to reflect the recommended number of hours of sleep (see Cappuccio et al. (2010)). A target was met if someone fell asleep earlier than the bedtime target and slept longer than the duration target. One may be concerned that the treatment may lead subjects to sleep more than recommended, given the mixed evidence on the health effects of sleeping longer than 9 hours (Watson et al., 2015). In practice we find the share of subjects sleeping more than 9 hours declined during treatment from 11% to 9.90%. The three treatments varied in the timing of the incentivized weeks and the form of incentives. [11] In Treatment 1 (Incentive-Weekly), the treatment weeks were Weeks 3, 4, and 5. Figure 1 and

[11] Of course a standard concern in field experiments in university settings is that subjects in different treatment arms might communicate with each other, compromising the estimated treatment effect. Subjects in our control group did not know others were paid for meeting sleep targets in our experiment. In our analyses, we controlled for seasonality and university fixed effects. Results were indeed similar in waves with and without control groups. Additionally, the stability of our main results to the inclusion of individual fixed effects is reassuring.

2 illustrate the timeline of our main intervention. We used gain/loss framing: each week, these subjects were told that they would be rewarded GBP 10 (USD 15 in Pittsburgh) for participation in the following week. Rewards and punishments were added to this amount. Each reward was GBP 2.5 (USD 3.75) and each punishment was also GBP 2.5 (USD 3.75), so that the largest gain for achieving targets on all four nights was GBP 20 (USD 30). The subjects would achieve their target by complying with both bedtime and sleep duration targets, measured by Fitbit data, on a given day. A failure was to miss either target on a given day. We also provided feedback on performance in the previous week and asked subjects to predict their own performance for all remaining treated weeks. Formally, we tested the following model

$$Y_{it} = \beta_1 * Treatment_{it} + \beta_2(Post - Treatment_{it}) + X_{it} + \eta_{it} + DOW_{it} + \epsilon_{it} \tag{1}$$

where Y_{it} is a sleep outcome of individual i, at time t; $Treatment_{it}$ is a dichotomous variable equal to 1 during the weeks of treatment if individual i in treatment group; $Post - Treatment_{it}$ is a dichotomous variable equal to 1 during the weeks following the last week of treatment if individual i in treatment group; X_{it} are individual and interview characteristics such as gender, age, season and location; η_{it} are week of the semester (or week by wave) fixed effect; DOW_{it} is a vector of dummies for each day of the week. In our preferred specification, we include individual fixed effects. And we estimate standard errors clustered at the individual level.

In each treated week, we also asked subjects to forecast their performance in the current week and in the remaining treated weeks. Depending on the size of the prediction reward, Treatment 1 (Incentive-Weekly) can be further divided into 2 arms: For approximately 40% of the subjects, only one prediction was finally chosen, and a correct prediction was rewarded GBP 2 (USD 3); for the remaining 60%, one prediction was chosen from each of the 3 treatment weeks, and each correct prediction was rewarded GBP 2 (USD 3).

We then tested two slight modifications to this treatment to see how subjects would respond to changes in frequency and structure of the incentives (Carrera et al., 2019a) (Treatment 2, Incentive-Biweekly); and in the size of the monetary incentive (Treatment 3, Small Incentive-Biweekly). Treatment 2 (Incentive-Biweekly) was the same as 1, except that the sleep incentives were given biweekly, in weeks 3, 5, and 7 (see Figure A.1). In Treatment 3 (Small Incentive-

Biweekly), the incentives were given biweekly as in 2, but we did not use gain/loss framing in the incentives; in other words, there was no initial endowment in each week. Instead, subjects could choose between two contracts. The first one was a reward of GBP 2.5 for each night the target was met, and there was no punishment. Therefore, meeting the target on all nights of a week could lead to a total reward of GBP 10. The alternative contract would not only involve the same reward for meeting the goals but also penalize unmet goals. The punishment for each failed night was GBP 2.[12]

In all treatment groups, rewards and punishments were added to their payments on the day they returned the device. One of the 3 treated weeks was selected for each subject to determine payment for their sleep performance. Table A.1 summarizes our treatments. Table 1 reports the differences between control and treatment groups with respect to baseline characteristics. We report these differences for the whole sample (column 1), the sample used in our main Tables (column 2), and separately for the UK and the US (columns 5 and 6).

Overall, subjects are well-balanced with respect to most variables. There is some evidence that individuals in the treatment group are more likely to be white and less likely to be smokers. In the analysis, we test the sensitivity of the analysis to race, smoking status, and age. In particular the age differences are driven by a couple of outliers. Their exclusion does not affect the results. Furthermore, including individual-fixed effects we rule out most of the concerns regarding selection on time-invariant individual characteristics.

3 Data

We advertised the experiment within the subject pools of CESS and PEEL, and a total of 359 participated in the experiment. Among these 319 of them generated usable data; 40 subjects (11%) either felt uncomfortable wearing a Fitbit or withdrew from the study due to other reasons.We check the sensitivity of our results to the inclusion of individuals who dropped out but generated some usable data. We find no evidence of significant association between compliance with the treatment and the likelihood of dropping out before week 8. Furthermore, withdrawing from the experiment does not correlate significantly with baseline characteristics of the subjects (Table

[12]This treatment was done in Oxford only. Subjects were paid GBP 8 for returning the device in addition to the show-up fee in Week 8. One treated week was randomly chosen and any loss was deducted from this amount.

A.2). We also note that while most subjects synced regularly, sleep data were missing for some of the nights. This could happen if subjects did not wear the Fitbit at night or the device was not charged. However, the median subject had sleep data for 82% of the nights. Excluding observations reporting with 25% or more nights of missing data yields similar results, although if anything the results of our intervention are magnified. Furthermore, the inclusion of individual fixed effects further mitigates the concerns of selection.

Among the 319 remaining participants, 107 were in the control condition, 104 in the weekly Treatment 1 (Incentive-Weekly), 76 in the biweekly Treatment 2, and 32 in the weak incentive Treatment 3. We provide the full questionnaires of the surveys conducted during the experiment in the Online Appendix C.

3.1 Measuring Sleep

Measuring sleep is challenging. Previous studies have shown that self-reported measures of sleep, whether based on time-use diaries or survey questions, are prone to severe measurement errors. Self-reports tend to overestimate sleep duration compared to objective measures (Lauderdale et al., 2008b). Time-use diaries may also be subject to overestimation bias, as often, the activity lexicon associated with sleeping includes transition states (e.g., falling asleep) (Basner et al., 2007). Personal wearable devices (such as Fitbit) have been used to study health behavior (e.g., Handel and Kolstad, 2017). Concerns have also been raised regarding the ability of Fitbit devices to provide an accurate measurement of sleep, although some medical studies (e.g., Lee et al., 2017) find wearable activity trackers that detect heart rate perform fairly well in terms of tracking sleep compared to actigraphy, the more sophisticated method used in medical studies (Beattie et al., 2017).[13] Because the devices measure heart rate, they are able to distinguish between time spent in bed not sleeping, such as watching TV, and time spent sleeping. It is also worth noting that if subjects would take off their devices, data would be missing and subjects would not be able to receive any reward. Thus subjects could not manipulate the sleep data in either of these ways.

While other papers have used Fitbits to measure activity, our is one of the few studies integrat-

[13] Beattie et al. (2017) suggest that Fitbit heart rate-tracking devices accurately track light, REM, and deep sleep stages (see also `http://www.sleepreviewmag.com/2017/06/study-shows-Fitbit-heart-rate-tracking-devices-accurately-track-light-rem-deep-sleep-stages/`.)

ing time-use data from three different sources: wearables, time-use diaries, and surveys. Using these data, we contribute to the methodological discussion on sleep measurement (e.g., Lauderdale et al., 2008a), by comparing information on sleep obtained from three of the main sources used in the literature—wearable devices, time-use diaries, and self-reported sleep in surveys. We identify substantial disparities in sleep measurements obtained using these three methods, which partially reflect the distance between beliefs and actual behavior. We have (1) the bedtime, wake-up time, and sleep duration as collected by the Fitbit devices; (2) self-reported information on sleep habits and quality before and during the experiment collected in surveys; and (3) sleep measured through our time-use diaries. Therefore, we can directly compare these three different measures of sleep. Additionally, Fitbit offers limited but useful information about sleep quality through sleep efficiency–the fraction of time spent asleep while in bed–and the number of sleep episodes per night.

Table A.3 compares the different measures of sleep obtained using Fitbit devices, survey data, and time-use diaries. On average, subjects reported 8.15 hours of sleep in time-use diaries and 7.07 hours of sleep for the previous week in self-reported surveys. Thus, time-use data tend to significantly overestimate time allocated to sleep, while self-reported sleep duration is only a few minutes longer than the average sleep duration measured by Fitbit devices (7.02 hours during the week). Further, according to time-use data, only 7% of the subjects reported sleeping less than 6 hours, while the survey-based measure indicated 10% of the subjects slept less than 6 hours—closer to but still significantly smaller than the 23% recorded by Fitbit devices during the school week. These results were also consistent with the overestimation by the subjects of own sleep duration in the first-day survey. As research on sleep choice, its determinants, and its effects advances, understanding the extent to which each of these methods captures both pure sleep duration and biased beliefs will be crucial in identifying best practices in sleep measurement.

3.2 Descriptive Analysis: Pre-Intervention Data

Table 2 reports summary statistics for subjects at baseline. This information was collected in the lab on the first day of the experiment. Subjects were 59% male, with an average age of 21.54

(min: 18; max: 45; median age: 21).[14] Of our respondents, 58% were White, 22% Asian, 9% Black, and 11% other minorities.

We measured subjects' health, well-being, and sleep behavior before the intervention. Subjects were relatively healthy. Only 11% reported poor health status. The average BMI in the sample was 24 (min: 15.5; max: 47.0), with only 5% obesity rate (BMI>30), and 24% overweight status (BMI>25); 23% of the subjects had ever smoked, but 61% of those subjects quit smoking; 26% reported drinking more than once per week.

However, self-reported mental health problem was a cause for some concern in this group. While 45% of the sample reported feeling depressed rarely or never, 36% reported that they had felt depressed 1–2 days over the last week, 15% reported occasional feelings of depression (3–4 days per week), and 4% reported feelings of depression most of the time (5–7 days per week). Moreover, 6% of the sample reported feeling completely satisfied with their life; 44% considered themselves very satisfied; 42% somewhat satisfied; and 9% not satisfied or not satisfied at all.

According to the survey results of sleep patterns at baseline, subjects sleep an average of 7 hours and 15 minutes (Min: 4; Max: 10) each night during the month before the experiment, with women sleeping 15 minutes longer on average— consistent with what has been found in time-use studies (see Hamermesh (2019)).[15] Most subjects reported an ideal sleep of 8 hours (7.97 on average), and 97% of subjects considered it ideal to sleep more than 7 hours. Yet, 46% reported sleeping less than 7 hours on an average night during term (see Figure A.3). Subjects reported falling asleep during the day on 3.79 days over the last month and a quality of sleep of 6.61 on a 1–10 scale. At baseline, 17.7% (19.3%) of subjects expressed that they were definitely willing to improve their sleep by sleeping longer (going to bed earlier); 43% (41%) stated they were probably willing to; the rest were either unwilling to improve or did not know how to (Table A.4). The median bedtime at baseline is 1:07am, with only 25% of the subjects going to bed before midnight and less than 10% going to bed before 11pm on a typical term night (see Figure 5).

Fitbit data of sleep before the intervention are plotted in Figure 3. Most people on most

[14] One part-time student was aged 45. Excluding this observation from the analysis does not affect the results. Figure A.2 displays the age distribution of the subjects in Oxford and Pittsburgh.

[15] The question asks "During the past month, how many hours of actual sleep did you get at night (average hours for one night)? (This may be shorter than the number of hours you spend in bed.)"

days slept between 6 and 9 hours, with subjects in Pittsburgh (dashed line) sleeping less than those in Oxford (solid line). On an average night, in the first 2 weeks before the experiment, subjects in Pittsburgh slept approximately 6 hours and 45 minutes, while subjects in Oxford slept 7 hours and 20 minutes. Women in our sample slept on average 7 hours, men 6 hours and 50 minutes (difference not statistically significant), but at baseline, women were significantly less likely to report sleeping less than 7 hours (-7% with respect to the mean). The gender difference in sleep duration is consistent with previous studies (Hamermesh, 2019). Figure A.4 documents the cumulative distribution of sleep hours. On an average night of the week, 70% of the time individuals slept less than 8 hours, 47% less than 7 hours, 25% less than 6 hours, and 12% less than 5 hours (see Figure 4). Sleep duration was highly irregular—the standard deviation was 2 hours—varying substantially throughout the week, with subjects sleeping significantly less during the week than on weekends (see Figure A.5). Subjects compensated during the weekend for some of their lost sleep hours during the week, wherein approximately 47% of the subjects slept less than 7 hours in the first 2 weeks, while during the weekend the fraction of individuals sleeping less than 7 hours of sleep declined to 39%.

We also document the association between insufficient sleep and self-reported measures of health and well-being at baseline using self-reported data drawn from the survey conducted on the first day of the experiment. Individuals who report sleeping between 7 and 9 hours were more likely to report good health status (+6% with respect to the mean; p-value$<$0.01); they were also 6 percentage points less likely to be obese (p-value$<$0.05) and overweight (-48% with respect to the mean; p-value$<$0.001); 55 percentage points less likely to report feelings of depression (p-value$<$0.05); and more likely to be satisfied with life (+56% with respect to the mean; p-value$<$0.001) (see Table A.5 for details). Individuals who were identified as more likely to take risks were also more likely to sleep less (see Figure A.6).

4 Incentives, Sleep Behavior, and Habit Formation

4.1 Financial Incentives and Sleep Behavior

Table 3 shows our main regression results. The unit of observation is the subject-night. We first pool Treatments 1 (Incentive-Weekly) and 2 (Incentive-Biweekly) and then document the heterogeneous effects of the treatments later in the text. Relative to control, we find that subjects receiving monetary incentives in Treatments 1 (Incentive-Weekly) and 2 (Incentive-Biweekly) were 19% more likely to sleep the recommended number of hours (between 7 and 9 hours, see Cappuccio et al. (2010)) (column 1). This result holds with the inclusion of individual fixed effects (column 2): accounting for persistent individual heterogeneity, the coefficient reduces by 42%, but still indicates an economically and statistically significant effect of the treatment on the likelihood of sleeping between 7 and 9 hours (+11% with respect to the mean). In columns 3 and 4, we examine the effects of treatment on a metric of insufficient sleep (sleeping less than 6 hours). When receiving the monetary incentive, individuals were 23% less likely to sleep less than 6 hours with respect to the average in the sample (column 3), and this effect holds even with the inclusion of individual fixed effects (column 4). Specifically, during treatment, individuals were 12% less likely to sleep less than 6 hours. The results tend in the same direction when analyzing alternative dichotomic outcomes for sleeping less than 7 or 5 hours (see Table A.6). On average, incentives increased sleep duration by 6–12 minutes. Individuals spent, on average, 10 minutes more in bed, 8 of which were minutes spent asleep. Regarding the nights on which subjects complied with the incentives, individuals in the treatment group slept 22 minutes longer than those in the control group. Results hold to the inclusion when we include week-by-wave fixed effects (see Table A.7).

The results on sleep duration are largely driven by earlier bedtimes. When receiving the monetary incentive, the subjects' bedtimes were moved earlier by approximately 20 minutes, while the average wake-up time did not change significantly (see Table A.8). Figure 5 visualizes the shift in sleep duration (upper figure) and bedtime (lower figure) induced by our intervention.

Restricting the sample to the nights individuals reported at least 4 hours and less than 10 hours of sleep, the results are substantially unchanged and, in fact, more precisely estimated, suggesting the main results are not driven by extreme values (the results are available upon request). These effects survived even after removal of the monetary incentive (See subsection 4.2). Interestingly, we find no evidence that sleeping more on incentivized nights (Monday–Thursday) crowded out sleep on non-incentivized nights during the intervention. In fact, subjects in the treatment group were also more likely to sleep the recommended number of hours during weekends (Table A.9) suggesting short-term habit formation.

The commitment devices and monetary incentives were effective. Subjects met their targets approximately 48% of the time. Overall, female subjects were 8% more likely to meet their targets compared to their male counterparts, as female subjects met their targets on at least 49% of the treatment nights while men met their targets only on 45% of those nights.

Subjects who chose dominated bedtimes ended up with better sleep outcomes. When choosing a dominated bedtime (earlier than 1 am), subjects were 14% more likely to achieve the target than those choosing 1 am as a bedtime target (Table A.10). Similarly, subjects choosing a dominated sleep duration target (longer than 7 hours) were more likely to achieve it. Overall, choosing a more demanding target was associated with higher success rates. Subjects choosing a more demanding bedtime (sleep duration) were 13% (20%) less likely to sleep less than 7 hours and those choosing both a demanding bedtime and a demanding sleep duration target were 26% less likely to sleep less than 7 hours (Table A.11, columns 1–3 and 5–7). Similarly, subjects choosing a dominated contract were less likely to report insufficient sleep during the treatment weeks (columns 4 and 8), although the latter result is not precisely estimated due to the small sample size of Treatment 3 (Small Incentive-Biweekly). It is worth noting that all these estimates restrict the sample to the intervention weeks while including controls for insufficient sleep at baseline, partially mitigating concerns of selection bias.

4.2 Post-Intervention

Our experiment had two post-intervention periods. The first was within the 8 experimental weeks, and thus, we still had data drawn from the wearable devices. The second occurred 3 months after the experimental period ended, and consisted of a follow-up survey to additionally investigate the effect of our treatments after the experiment.

4.2.1 Habit Formation and Sleep with Fitbit Data

We first explore the first part of the post-intervention period. In Table 3, we find evidence that the effects of monetary incentives on sleep persist to some extent in the weeks following the termination of treatment. After removing the monetary incentive, subjects in the treatment group were 9% more likely to sleep between 7 and 9 hours, although these results are not precisely estimated (column 1) and are not robust to the inclusion of individual fixed effects (column 2). Yet, when focusing on the left tail of the sleep distribution, we find significant and sizeable effects when removing the financial incentive (column 3)[16]. While the coefficients are marginally smaller, the effects hold even after including individual fixed effects (column 4). In fact, after the removal of the incentive, the effect was even larger (+16% with respect to the mean). Using the natural logarithm of sleep, we find that even after the removal of incentives, treated subjects' sleep was 2% longer than at baseline. The difference in magnitude between the treatment and post-treatment effect is comparable with recent evidence on habit formation effects when using financial incentives to promote exercising (Carrera et al., 2019a). In sum, while we only observe students for three weeks after the end of treatment, this evidence suggests there may be a habit formation effect. In Section 5.4, we document changes of daily routines, such as screen time, which may contribute to maintain healthier sleep habits beyond the duration of our experiment.

Examining other outcomes drawn from the Fitbit data (Table 4), we find no evidence of significant effects of treatment on the efficiency of sleep (columns 1–2), measured as the ratio between sleep duration and time spent in bed (including time awake). There is some weak evidence that treated subjects were more likely to have an efficient resting heart rate, defined as a resting heart rate in the lowest 25th percentile of those reported in the first 2 weeks of the experiment and before the start of the intervention (columns 3–4). Finally, treated subjects spent less time in sedentary activities (columns 5–6) in their wakeful time during treatment weeks (-9 minutes per day). The magnitude of these effects is relatively small. We find no evidence of any significant effect on the number of steps walked.

4.2.2 Follow-up Survey: Effects on Self-Reported Sleep, Health, and Human Capital

In the second part of the post-intervention period, we sent subjects a follow-up survey that included questions about health, sleep, and academic performance three months after the experiment. The response rate to our follow-up survey was 46%, and thus, the results should be interpreted with some caution. However, there is little evidence of systematic selection when examining the baseline characteristics of those who did not respond to the follow-up survey (see Table A.12). Subjects not responding to this survey tended to be older and were more likely to be

[16] Approximately 25% of subjects reported sleeping less than 6 hours at baseline.

African-American than the respondents. Nonetheless, most characteristics are not significantly different between the two samples.

Subjects receiving any incentive during the experiment had significantly higher sleep quality 3 months after the end of the experiment compared to those not treated (see table A.13)Additionally, those who had a higher achievement rate for their bedtime targets also reported better sleep quality. There is little evidence of changes in self-reported health, but treated subjects were 1.4 percentage points less likely to report very poor health status.

Finally, we investigated the qualitative effects of our intervention on academic performance (see Table 5). Sleep duration and regularity were found to be directly related to grade changes—those who slept longer, slept more regularly between 7 and 9 hours, and were less likely to sleep less than 6 hours experienced larger grade increases than those who did not. Having greater variance in sleep was associated with decreases in self-reported percentile rank. Being a part of any treatment group is associated with a 6.3-point increase in percentile rank with respect to one's own percentile rank before the experiment. Having a higher rate of compliance with the treatment, through meeting the target, was also associated with an increase in the letter grade—those who met the target more than 50% of the time had a 0.162-point greater increase in their letter grades than those who met their target less than 50% of the time.

5 Behavioral Biases and Sleep Choice

5.1 Time Inconsistency and Demand for Commitment

Several aspects of our participants' behavior were consistent with partially sophisticated time inconsistency. We correlated our measures of subjects' time preference with baseline sleep patterns and performance in the experiment.[17] The results are reported in Table 6. Columns 1–3 report estimates based on self-reported sleep in the survey conducted on day 1. Columns 4–6 report estimates based on the first two weeks of data collected from Fitbit devices. While estimates are not precise due to the small sample size, we find that before intervention, present-biased subjects were more likely to be sleep deprived. In particular, at baseline, individuals in the top quartile of our measure of present bias were 10% (11%) more (less) likely to report sleeping less

[17] In Appendix B, we describe in detail how we built our measure of present bias and impatience.

than 7 hours (between 7 and 9 hours, columns 1–3). Fitbit data reveal even larger differences (columns 4–6). Present-biased subjects were 19% more likely to sleep less than 7 hours with respect to the mean and 21% less likely to sleep between 7 and 9 hours, sleeping on average 12 minutes less per night. The relationship between sleep duration and impatience appears to be less clear (see columns 4–6, Panel B).

Our experiment included two features that allowed us to directly observe the demand for commitment consistent with sophisticated hyperbolic discounting models. First, in all intervention groups, we asked subjects to choose bedtime targets between 10 pm and 1 am and sleep duration targets between 7 and 9 hours. An agent with standard preferences would maximize rewards by choosing the least binding targets, namely 1 am and 7 hours. By contrast, choosing more restrictive targets is equivalent to disciplining one's future behavior and can serve as a commitment device. Second, in Treatment 3 (Small Incentive-Biweekly), we asked subjects to choose between a contract that only rewards successes and a contract that not only rewards successes to the same extent but also punishes failures. To maximize monetary payoff, an agent with standard preferences would choose the former, whereas an agent who demands commitment would choose the latter (e.g., Kaur et al., 2015). An agent with naive time-inconsistent preferences may mistakenly predict that her/his future self will achieve all targets and thus be indifferent between having and not having a commitment device, whereas a sophisticated agent may anticipate her/his future time-inconsistent behavior and would actively demand for a commitment device even at some cost. The "cost" of the commitment device in our setting is the forgone reward, or explicit punishment in some cases.

We uncover some interesting evidence of demand for commitment. Despite 1 am being a dominant choice for bedtime target, in 50% of the weeks, subjects in the treatment group chose bedtime targets earlier than 1 am. Moreover, 60% of the subjects chose bedtime targets earlier than 1 am in at least 1 week; 24% chose bedtime targets earlier than 1 am in all 3 treatment weeks (Table A.14, column 1). Similarly, despite 7 hours being a dominant choice for sleep duration target, in approximately 48% of the subject-week observations in the treatment group, bedtime targets longer than 7 hours were chosen. Moreover, 60% of the subjects chose sleep duration targets longer than 7 hours in at least 1 week, and 19% chose sleep duration targets longer than 7 hours in all 3 treatment weeks (Table A.14, column 2). These results are comparable in magnitude to those of Schilbach (2019), who find that one-third to half of study participants chose sobriety incentives over unconditional payments, even when this choice implied a cost in terms of forgone payments. It is worth noting that the chance that menu or demand effect (Carrera et al., 2019b) drove the choice of restrictive targets was small, as the least restrictive targets (1 am and 7 hours) are at opposite extremes on their respective lists. Additionally, the correlation between pre-treatment behaviour and choice of commitment cannot be explained by menu or demand effect. In addition, consistent with demand for commitment, in Treatment 3 (Biweekly-Small), in approximately 10% of the subject-week observations, the contract with punishment was chosen (Table A.14, column 3). A total of 13% of the subjects in this treatment chose a contract with punishment. Subjects choosing a dominated contract were also significantly more likely to choose a dominated bedtime target (+50%).

Present-biased and impatient individuals were more likely than other subjects to take up a commitment device (Table A.15). They were 25% more likely to choose a bedtime target before 1 am (column 1) and 6% more likely to choose a sleep duration target longer than 7 hours (column 2). Overall, present-biased subjects were 22% more likely to commit to at least one dominated target (column 3). Similarly, impatient individuals were more likely than other subjects to choose a bedtime target before 1 am (+26%, column 1 Panel B).

Time-inconsistent subjects may be more likely to choose more demanding sleep targets earlier in the day, when the cost of last night's bad sleep choice is still felt. Yet later in the day, when the desire to watch another episode of a TV series sets in, they may be more likely to choose less restrictive targets. We exploit variation in the time the surveys were answered and targets chosen by the subjects. We deliberately randomized the timing of surveys throughout the experiment, although we could not fully control the timing of the answers. Surveys were sent out at different times throughout the study ranging from 7am to 3pm. The same subject received the survey at a different time each week. While only 35% of the subject chose the least binding bedtime (1 am) when responding to the survey before noon, among those responding after noon, 53% of the subjects chose the least binding target. Among subjects responding after 6 pm, this proportion increases to 64% (see Figure A.7). We also find that, in a continuous measure, people who responded later in the day set later bedtime targets. Using the timing at which the survey was sent out, we show that when the survey was sent out earlier in the day, subjects were 12 percentage points more likely to choose an earlier bedtime target (column 1, Table A.16), a 24% increase with respect to the mean. Including individual fixed effects, the coefficient drops, but it is still large and statistically significant at the 10% level, suggesting a 14% increase in the likelihood of selecting an earlier bedtime target. Instrumenting the response to the survey with the time at which the survey was sent (column 3), we find that individuals who answered the survey before 3 pm were 17 percentage points more likely to select a bedtime target earlier than 1am (p-value=0.016), equivalent to a 34% increase with respect to the mean. Including individual fixed effects, and exploiting within-subject variation in the timing at which the survey was sent, we find that answering the survey before 3pm (median response time) increased by 10.8 percentage points (+21.3%) the likelihood of choosing a bedtime earlier than 1am (p-value=0.057), column 4. Additionally, we also find evidence that the later the average actual bedtime the week before, the earlier the bedtime target set by subjects, suggesting their partial sophistication and willingness to improve (see Table A.17).

While the cost of choosing a more binding target should be lower for those with earlier bedtime at baseline, Figure 6 shows that subjects with later bedtimes at baseline (as measured by Fitbit devices in the first two weeks) selected bedtime targets relatively earlier than their baseline sleep, compared to those with earlier bedtimes, as evidenced by the comparison with the 45 degree line. Additionally, consistent with the hypothesis that sophisticated time-inconsistent preferences may be an important factor behind sleep choice behavior, we found that the behavior of opting for the commitment device was correlated with subjects' predicted bedtime (elicited before the target selection). Subjects who expected to go to bed later set earlier bedtime targets than their predicted bedtime (Figure 7), which could reflect subjects with partial sophistication wanting a commitment device.

5.2 Biased Beliefs

As mentioned earlier, our evidence suggests that biased beliefs may contribute to explaining individual sleep choices. First, the data drawn from the survey conducted on the first day of the experiment reveal that subjects systematically reported longer sleep durations, better sleep quality, and lower risks associated with sleep for themselves than what they considered the average for persons of the same age (see Table 7). The majority of subjects (62%) believed they sleep longer than the average person of the same age. Individuals reported sleeping 20 minutes longer than an average person of their age. Similarly, 58% of the subjects thought that their sleep quality was better than that of the average person of their age, with 25% of the subjects rating their sleep quality 2 points higher than average on a 1–10 scale. These are consistent with self-serving belief and an over-placement of own sleep relative to others. It is also worth noting that most subjects (97%) considered it ideal to sleep more than 7 hours. This suggests misinformation on the importance of sleep may play a limited role in explaining biased beliefs. This biased belief of own sleep reflects that individuals may fail to fully integrate information to update belief in face of repeated feedback in everyday life, consistent with the existence of motivated beliefs.

We additionally assessed the motivated beliefs mechanism by analyzing subjects' self-reported sleep duration when they are provided with information from the Fitbit. If subjects do not update their self-reported sleep duration when the fitbit information is available, this may further rule out alternative explanations, such as misinformation or uncertainty, and suggest motivated beliefs may play an important role. Indeed, when looking at subjects in the control group, we find no evidence of significant differences in self-reported sleep duration before and after being provided with the Fitbit and the Fitbit's information, suggesting that they are not fully incorporating information into their beliefs. If anything subjects in the control group reported longer sleep duration at the end of the experiment than in the survey conducted on the first day in the lab, but the differences were small and non-significant (p-value=0.44). Furthermore, we show that the higher the distance between self-reported sleep and actual sleep as measured by Fitbits, the higher the number of hours subjects would predict to sleep in the next week (see Figure 8). In particular, a one standard deviation increase in the difference between self-reported sleep hours for the previous night and the sleep measured by the Fitbit device, would be associated with a .26 standard deviation increase in the number of hours a subject would predict to sleep on a typical night of the following week.

Comparing Fitbit data and self-reported data on sleep duration, we also find evidence that individuals sleeping less than 7 hours were significantly more likely to overestimate their sleep duration, suggesting that overconfidence may be an important factor behind insufficient sleep. As mentioned above, subjects tended to overestimate the duration of own sleep and, consistent with previous evidence, time-use data were particularly prone to this bias (see Table A.3).

We find also evidence of overconfidence with respect to the perceived risks of sleep deprivation: 66% of the subjects estimated for themselves a lower risk of detrimental consequences of sleep deprivation (loss of alertness, weight gain, insomnia, cold, arterial stiffening). In particular, 82% of them thought that others would have a higher likelihood of losing alertness as a consequence of sleep deprivation, with an average of 30-percentage-point higher risk estimated for other individuals of the same age group. Similarly, approximately 65% assessed a higher likelihood for others of the same age group (than themselves) to gain weight and to have insomnia as a result of sleep deprivation. In contrast, differences in the perceived risk of self and others suffering a cold or arterial disease induced by sleep loss were less pronounced.

Using these data, we built an index of overconfidence along these different dimensions. In practice, we summed the overconfidence measures in a single index and defined as overconfident those individuals in the upper quartile of the index. Splitting individuals in this way, overconfident subjects were less likely to report insufficient sleep at baseline based on self-reported data, but more likely to be sleep deprived based on Fitbit data before treatment (Table 8). In other words, while individuals who were overconfident about sleep reported longer sleep duration at baseline, these subjects were also sleeping significantly less than the rest of the sample based on Fitbit data. We did not find significant differences in their likelihood to take up commitment devices (Table A.15, Panel C). However, on average, overconfident individuals chose sleep duration targets that were 1 hour longer than their sleep at baseline, while the rest of the subjects, on average, selected targets that were 8 minutes longer than their sleep at baseline. The difference between the sleep duration target and the usual sleep was approximately 52 minutes longer for overconfident subjects (p-value=0.001). In other words, while overconfident subjects were equally likely to choose dominated targets, given that their bedtime at baseline was significantly later and their baseline sleep duration was significantly shorter, they took up overly optimistic sleep duration and bedtime targets. Furthermore, as mentioned above, among present-biased individuals, overconfident subjects were less likely to achieve targets, and commitment devices were not effective (possibly even welfare diminishing) for them, consistent with Bai et al. (2017). While the differences are not precisely estimated, we find that overconfident subjects with present bias were 12% less likely to sleep the recommended number of hours (p-value=0.27). Among overconfident subjects in the control group, the difference between self-reported and Fitbit-measured sleep duration remained significant and substantially unchanged throughout the experiment. These results suggest that despite the feedback available through the wearable device, subjects kept overestimating sleep duration.

Participants were also asked to predict the likelihood that they would achieve their chosen target in each of the following treated weeks. Correct predictions were rewarded. Table A.18 shows that individuals tended to over-predict their likelihood of achieving the targets. Predictions do not seem to be improving over time: while subjects were revising their predictions down from week to week, they were also increasingly falling short of their targets as the study proceeded. In the first treated week, 62% of the subjects were too optimistic about the number of nights they could achieve; in the second (third) week of treatment 61% (71%) of the subjects were too optimistic. The decreasing achievement rate may be partially explained by increasing demands on time as the semester proceeds. While students might recognize that this is happening, they may be consistently underestimating how the demands on their time will change. We note that incentivizing predictions may be problematic as partially-sophysticated present-biased participants may reinforce their incentives to meet desired goals (Acland and Levy, 2015). However, we find no evidence of increasing prediction incentive affecting prediction accuracy. Furthermore, we find no significant difference in the achievement rate of subjects with and without incentive to predict.

As mentioned above, choosing a dominated target (or contract) was associated with a higher success rate (see Tables A.10 and A.11). However, we find no evidence that choosing dominated targets improved sleep duration among present-biased individuals who were also classified as overconfident (coef. 0.13, p-value=.62), while the effect is large and significant among the rest of the sample (coef. 0.28, p-value=0.009).

Overall, these results, albeit not all precisely estimated, appear consistent with partially sophisticated time inconsistency.

5.3 Risk Preferences

In Table 6 (Panel C), we explore the correlation between subjects' risk aversion and their average sleep duration as estimated during the first lab session. Risk-averse individuals reported longer sleep duration, were less likely to report less than 7 hours of sleep (column 2), and more likely to sleep between 7 and 9 hours (column 3). Overall, Fitbit data confirm these qualitative associations, although the magnitude of the estimates is somehow smaller.

Risk-averse individuals were also less likely (-23%) to choose a demanding target (earlier than 1 am, column 1 of Table A.15, Panel D) and less likely to choose a sleep duration target longer than 7 hours (-10%, column 2). Interestingly, subjects choosing a dominated contract tended to have low risk aversion. In anything, risk-averse individuals were 8% more likely to meet their target (p-value=0.27).

5.4 Incentives to Sleep and Time-Use Allocation

A natural question is whether and how the allocation of time changed in response to our intervention. Individuals may compensate insufficient sleep at night by napping during the day or by sleeping longer during the weekend. Other studies find significant effects of naps on productivity and well-being (Monk et al., 2001; Bessone et al., 2018). We investigated whether our intervention affected the time allocated to naps. Only 5% of the subjects reported any sleep lasting less than 2 hours between 7 am and 7 pm during weekdays. Although nap duration is negatively correlated with sleep duration at night (-0.13) and individuals sleeping between 7 and 9 hours are significantly less likely to report any naps (-3.89%), we find no evidence that our intervention systematically affected the likelihood of taking a nap and the nap duration (see Table A.19, columns 1–4). Thus, unsurprisingly, the results are substantially unchanged when we include controls for napping behavior (columns 5–6). We also find no evidence of subjects changing their weekend sleep duration during the intervention in response to the longer sleep duration induced by the incentives during the week. In fact, during the three weeks of the intervention, treated subjects were sleeping longer also during weekend (Table A.9). Although the effects are less precisely estimated than when analyzing the treated nights, the point estimates are not statistically different. Overall, these results are consistent with habit formation.

The subjects may also reallocate their time devoted to other activities when receiving incentives to go to bed earlier and to sleep longer. Using time-use diaries, we directly examine the effects of our incentives on individual time allocation. Time-use data are available for approximately 72% of the participants, and thus, results should be interpreted with some caution. The subjects not responding to the time-use surveys were younger, more likely to be Blacks, and were 10% more likely to report less than 7 hours of sleep during a typical night of the term, although this difference is only marginally significant (see Table A.20).

As mentioned above, consistent with previous evidence (Lauderdale et al., 2008a), we find that individuals tend to overestimate sleep when using time-use diaries. Indeed, there is no evidence that the subjects sleep longer during treatment when using time-use data and examining the likelihood of reporting between 7 and 9 hours of sleep (Table 9, columns 1–3). However, individuals do report significantly lower likelihood of sleeping less than 6 hours (-66% with respect to the mean, column 3). When examining other activities, we find no significant evidence that the increase in sleep duration was associated with a change in time spent studying, working, on personal care activities, exercising, relaxing, hanging around with friends or on the Internet, although we may be have not sufficient statistical power to identify some of these effects. Interestingly, the only activity that is systematically and significantly less likely to be reported under the intervention is "watching TV videos" (column 5, panel B). Indeed, for those who complied with the treatment, screen time after 8 pm declined by 13 minutes (see column 1 of Table A.21), equivalent to a 48% reduction with respect to the average screen time (45 minutes) observed in the sample. Among those who achieved the target at least half of the times, the coefficient decreases by 40% after the incentive is removed, but it is still economically and statistically significant. We find similar results when considering the likelihood of spending any amount of time on TV, Internet, or video games. Those achieving the target during the intervention are 12.5 percentage points less likely to report any screen time. This is equivalent to 33% of the sample mean. After the incentive is removed, the subjects who achieved the target on most nights are 8 percentage points less likely to report any screen time (a 20% effect with respect to the mean).

These results are consistent with recent research suggesting that screen time near bedtime is associated with lower sleep duration (Nie and Hillygus, 2002; Twenge et al., 2017; Billari et al., 2018).

5.5 Additional Findings: Sleep Regularity, Structure and Size of the Incentives

This section reports some additional findings regarding the effect of our intervention. Interestingly, the monetary incentives affected the regularity of sleep, bedtime, and waking time, reducing their variance (Table A.22). However, these effects did not persist after the removal of the incentive.

We do not find statistically significant differences when examining the role of the frequency and the structure of the incentives (Table A.23). In fact, the weekly incentive has stronger post-treatment effects, although these differences are not precisely estimated, and thus, should be interpreted with caution.[18]

Finally, we explore the role of incentive size. Using a smaller monetary incentive and eliminating loss framing leads to effects that are smaller and non-precisely estimated. In particular, the effects of weaker incentives on the likelihood of reporting between 7 and 9 hours are about 50% lower and non-significant (Table A.24). Furthermore, while the effect of larger monetary incentives survives the inclusion of individual fixed effects, the point estimate of the weak incentive treatment is close to zero. Unsurprisingly, given the lack of in-treatment effects, we find no evidence of post-treatment effect. However, consistent with what we found earlier, the effects are larger and more precisely estimated when focusing on the likelihood of sleeping less than 6 hours. Pooling all the treatments (1–3) in one, we substantially confirm the main results (see Table A.25) while increasing the precision of the point estimates as the sample size increases.

6 Conclusion

Statistics reveal that many individuals sleep less than the recommended number of hours. There are several factors affecting individuals' sleep choices. Understanding how to improve health habits is crucial in designing policies aimed at promoting healthier behavior. As pointed out by Charness and Gneezy (2009), people tend to underestimate the impact of current actions on future utility and discount the future too much. Our evidence suggests that this tendency also characterizes sleep behavior. The prevalence and persistence of behavioral biases in the

[18] In the biweekly treatment, we regard as post-treatment any week after the first week of treatment during which subjects did not receive a monetary incentive. Using an alternative definition and focusing only on the last week of the experiment (week 8), we find similar results.

sleep domain is particularly interesting given the repeated feedback individuals receive on sleep throughout their lives (Huffman et al., 2018).

We studied sleep choice, and whether commitment devices as well as monetary incentives can improve sleep behavior among students. We find supportive evidence for partially sophisticated time-inconsistent preferences in sleep choice. The subjects in our experiment chose commitment devices even if this meant a lower monetary reward in expectation. Present-biased subjects were more likely to be sleep deprived at baseline, but many of them committed to dominated bedtime or sleep duration targets. Subjects choosing more demanding targets were also more likely to achieve them, with the exception of those who were classified as overconfident. Indeed, many subjects tended to be overconfident in their own sleep duration and quality and were more optimistic about themselves than about others when assessing the risks associated with insufficient sleep. Overconfident individuals were more likely to be sleep deprived at baseline and more likely to select overly optimistic targets, and thus less likely to achieve them. Risk aversion was associated with better sleep and a higher likelihood of achieving target during the intervention.

Our incentives improved sleep behavior and led to some habit formation effects, with subjects in the treatment groups sleeping longer even after the incentives were removed. Furthermore, monetary incentives increased sleep regularity, reducing the variance of bedtime, wake-up time, and sleep duration. Finally, we show that incentives to sleep may also have positive effects on academic outcomes, although these results are at best suggestive and further research is needed to establish this finding. When receiving incentives to sleep longer, individuals significantly reduced screen time (watching TV, YouTube videos, or surfing the Internet), while time spent with friends, working, or studying were not affected. Overall, these results give us a more nuanced understanding of sleep choice. Despite many economic models regarding sleep as an exogenous and homogeneous constraint on time, we provide evidence that behavioral biases play an important role and affect the heterogeneity of choice.

Our findings suggest that time inconsistency and biased beliefs can persist in the face of extensive experience and feedback. Thus, interventions based only on information (i.e. educational programs on sleep hygiene or fatigue management) may not be effective in the presence of these behavioral biases. Self-control problems may lead to procrastination with subjects repeatedly placing higher weight on immediate outcomes, and constantly delaying the start of good sleep habits. Also, people with motivated beliefs may be able to suppress the recall of objective feedback challenging their self-image, so that the simple provision of information may be ineffective in correcting misperceptions. Yet, to the extent subjects become more aware of their time inconsistent preferences due to the repeated feedback, sleep is also a domain where demand for commitment may be relevant and commitment devices effective. We show that appropriate incentives can be used to improve an individual's sleep behavior, while we find no evidence of subjects updating their beliefs with the additional information provided by the wearable devices, suggesting they are not fully incorporating information into their beliefs. Incentives to go to bed earlier and to sleep longer sleep were effective, suggesting that there is a cost to sleep, either in effort or in alternative uses of time, which can be compensated with a monetary payment.

Our findings also suggest that commitment devices and incentive structures may be more effective than planning tools at improving sleep behavior (Handel and Kolstad, 2017), and that temporary interventions, as those adopted by some companies, may have persistent effects, particularly when individuals lack a commitment device in natural settings. Providing incentives and commitment devices may help time inconsistent and overconfident individuals improve their sleep habits. Incentives and commitment devices may promote better sleep behaviors among subjects with self-control problems in the form of a time-inconsistent taste for immediate gratification (O'Donoghue et al., 2006), and with overconfidence as a result of motivated beliefs resilient to repeated information (Bénabou and Tirole, 2016). Incentives may also mitigate the role of motivated reasoning (Zimmermann, 2019). At the same time, our results imply that interventions that help individuals form routines conducive to healthy sleep habits (i.e., reduced screen time) may have longer-lasting effect.

The results on academic achievement, self-reported health, and heart rate efficiency support the growing evidence suggesting that sleep is a fundamental input for human capital and health. Taken together, the evidence on the behavioral factors behind sleep choice and the direct effects of sleep on health and productivity indicates the importance of not treating sleep as a mere time constraint, as well as the need to account for its direct effects on productivity of waking hours. The potential effects of our intervention on post-treatment sleep behavior, health outcomes, and human capital suggest the significance of further research along this line. Future research efforts exploiting larger samples may shed further light on the human capital and health effects of interventions aimed at improving sleep duration and quality. Future studies could also explore the relative effectiveness of non-monetary incentives and alternative commitment devices in nudging individuals into healthier and persistent sleep habits.

References

Acland, Dan, and Matthew R Levy (2015) 'Naiveté, projection bias, and habit formation in gym attendance.' *Management Science* 61(1), 146–160

Aguiar, Mark, and Erik Hurst (2007) 'Measuring trends in leisure: The allocation of time over five decades.' *The Quarterly Journal of Economics* pp. 969–1006

Aguiar, Mark, Erik Hurst, and Loukas Karabarbounis (2013) 'Time use during the great recession.' *The American Economic Review* 103(5), 1664–1696

Andreoni, James, and Charles Sprenger (2012) 'Estimating time preferences from convex budgets.' *American Economic Review* 102(7), 3333–56

Ashraf, Nava, Dean Karlan, and Wesley Yin (2006) 'Tying odysseus to the mast: Evidence from a commitment savings product in the philippines.' *Quarterly Journal of Economics* 121(2), 635–672

Bai, Liang, Benjamin Handel, Edward Miguel, and Gautam Rao (2017) 'Self-control and demand for preventive health: Evidence from hypertension in india.' Technical Report, National Bureau of Economic Research Working Paper No.W23727

Basner, Mathias, Kenneth M Fomberstein, Farid M Razavi, Siobhan Banks, Jeffrey H William, Roger R Rosa, and David F Dinges (2007) 'American time use survey: sleep time and its relationship to waking activities.' *Sleep* 30(9), 1085–1095

Beattie, Z, A Pantelopoulos, A Ghoreyshi, Y Oyang, A Statan, and C Heneghan (2017) 'Estimation of sleep stages using cardiac and accelerometer data from a wrist-worn device.' In 'Sleep,' vol. 40 OXFORD UNIV PRESS INC JOURNALS DEPT, 2001 EVANS RD, CARY, NC 27513 USA pp. A26–A26

Becker, Gary S (1965) 'A theory of the allocation of time.' *The Economic Journal* pp. 493–517

Bénabou, Roland, and Jean Tirole (2016) 'Mindful economics: The production, consumption, and value of beliefs.' *Journal of Economic Perspectives* 30(3), 141–64

Bessone, Pedro, Gautam Rao, Frank Schilbach, Heather Schofield, and Mattie Toma (2018) 'Sleepless in chennai: The economic effects of sleep deprivation among the poor.' Technical Report, Working paper. MIT

Biddle, Jeff E, and Daniel S Hamermesh (1990) 'Sleep and the allocation of time.' *Journal of Political Economy* 98(5, Part 1), 922–943

Billari, Francesco C, Osea Giuntella, and Luca Stella (2018) 'Broadband internet, digital temptations, and sleep.' *Journal of Economic Behavior and Organization*

Breig, Zachary, Matthew Gibson, and Jeffrey Shrader (2018) 'Why do we procrastinate? present bias and optimism.' Technical Report, Mimeo

Bryan, Gharad, Dean Karlan, and Scott Nelson (2010) 'Commitment devices.' *Annual Review of Economics* 2(1), 671–698

Buboltz, Walter C, Franklin Brown, and Barlow Soper (2001) 'Sleep habits and patterns of college students: a preliminary study.' *Journal of American College Health* 50(3), 131–135

Cappuccio, Francesco P, Lanfranco D'Elia, Pasquale Strazzullo, and Michelle A Miller (2010) 'Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies.' *Sleep* 33(5), 585

Carrera, Mariana, Heather Royer, Mark Stehr, and Justin Sydnor (2019a) 'The structure of health incentives: evidence from a field experiment.' *Management Science*

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky (2019b) 'How are preferences for commitment revealed?' Technical Report, National Bureau of Economic Research

Casaburi, Lorenzo, and Rocco Macchiavello (2019) 'Demand and supply of infrequent payments as a commitment device: evidence from kenya.' *American Economic Review* 109(2), 523–55

Charness, Gary, and Uri Gneezy (2009) 'Incentives to exercise.' *Econometrica* 77(3), 909–931

DellaVigna, Stefano, and Ulrike Malmendier (2006) 'Paying not to go to the gym.' *American Economic Review* 96(3), 694–719

Dupas, Pascaline, and Jonathan Robinson (2013) 'Why don't the poor save more? evidence from health savings experiments.' *American Economic Review* 103(4), 1138–71

Eisenberg, Daniel, Justin Hunt, and Nicole Speer (2013) 'Mental health in american colleges and universities: variation across student subgroups and across campuses.' *The Journal of Nervous and Mental Disease* 201(1), 60–67

Falk, Armin, Anke Becker, Thomas J Dohmen, David Huffman, and Uwe Sunde (2016) 'The preference survey module: A validated instrument for measuring risk, time, and social preferences.' *Working Paper*

Gibson, Matthew, and Jeffrey Shrader (2018) 'Time use and labor productivity: The returns to sleep.' *Review of Economics and Statistics* 100(5), 783–798

Giuntella, Osea, and Fabrizio Mazzonna (2019) 'Sunset time and the economic effects of social jetlag. evidence from us time zone borders.' *Journal of Health Economics*

Giuntella, Osea, Sally McManus, Redzo Mujcic, Andrew J Oswald, Nattavudh Powdthavee, and Ahmed Tohamy (2019) 'Why is there so much midlife distress in affluent nations?'

Giuntella, Osea, Wei Han, and Fabrizio Mazzonna (2017) 'Circadian rhythms, sleep, and cognitive skills: Evidence from an unsleeping giant.' *Demography* 54(5), 1715–1742

Greiner, Ben (2015) 'Subject pool recruitment procedures: organizing experiments with orsee.' *Journal of the Economic Science Association* 1(1), 114–125

Hafner, Marco, Martin Stepanek, Jirka Taylor, Wendy M Troxel, and Christian van Stolk (2017) 'Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis.' *Rand Health Quarterly*

Hamermesh, Daniel S (2019) *Spending Time: The Most Valuable Resource* (Oxford University Press)

Handel, Benjamin, and Jonathan Kolstad (2017) 'Wearable technologies and health behaviors: new data and new methods to understand population health.' *American Economic Review: Papers and Proceedings* 107(5), 481–85

Harrison, Glenn W, Morten I Lau, and Melonie B Williams (2002) 'Estimating individual discount rates in denmark: A field experiment.' *American Economic Review* 92(5), 1606–1617

Heissel, Jennifer A, and Samuel Norris (2018) 'Rise and shine the effect of school start times on academic performance from childhood through puberty.' *Journal of Human Resources* 53(4), 957–992

Hershner, Shelley D, and Ronald D Chervin (2014) 'Causes and consequences of sleepiness among college students.' *Nature and Science of Sleep* 6, 73

Holt, Charles A, and Susan K Laury (2002) 'Risk aversion and incentive effects.' *American Economic Review* 92(5), 1644–1655

Huffman, David, Collin Raymond, and Julia Shvets (2018) 'Persistent overconfidence and biased memory: Evidence from managers'

Jagnani, Maulik (2018) 'Poor sleep: Sunset time and human capital production.' *Mimeo*

Jakicic, John M, Kelliann K Davis, Renee J Rogers, Wendy C King, Marsha D Marcus, Diane Helsel, Amy D Rickman, Abdus S Wahed, and Steven H Belle (2016) 'Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the idea randomized clinical trial.' *JAMA* 316(11), 1161–1171

Jin, Lawrence, Nicolas R Ziebarth et al. (2015) 'Sleep and human capital: Evidence from daylight saving time.' Technical Report, HEDG, c/o Department of Economics, University of York

Just, David R, and Joseph Price (2013) 'Using incentives to encourage healthy eating in children.' *Journal of Human Resources* 48(4), 855–872

Kahn-Greene, Ellen T, Desiree B Killgore, Gary H Kamimori, Thomas J Balkin, and William DS Killgore (2007) 'The effects of sleep deprivation on symptoms of psychopathology in healthy adults.' *Sleep Medicine* 8(3), 215–221

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan (2015) 'Self-control at work.' *Journal of Political Economy* 123(6), 1227–1277

Kremer, Michael, Gautam Rao, and Frank Schilbach (2019) 'Behavioral development economics.' *Handbook of Behavioral Economics*

Laibson, David (1997) 'Golden eggs and hyperbolic discounting.' *The Quarterly Journal of Economics* 112(2), 443–478

— (2015) 'Why don't present-biased agents make commitments?' *American Economic Review* 105(5), 267–72

— (2018) 'Private paternalism, the commitment puzzle, and model-free equilibrium.' In 'AEA Papers and Proceedings,' vol. 108 pp. 1–21

Lally, Phillippa, Cornelia HM Van Jaarsveld, Henry WW Potts, and Jane Wardle (2010) 'How are habits formed: Modelling habit formation in the real world.' *European journal of social psychology* 40(6), 998–1009

Lauderdale, Diane S, Kristen L Knutson, Lijing L Yan, Kiang Liu, and Paul J Rathouz (2008a) 'Self-reported and measured sleep duration: how similar are they?' *Epidemiology (Cambridge, Mass.)* 19(6), 838–845

— (2008b) 'Sleep duration: How well do self-reports reflect objective measures? The cardia sleep study.' *Epidemiology (Cambridge, Mass.)* 19(6), 838

Lee, Hyun-Ah, Heon-Jeong Lee, Joung-Ho Moon, Taek Lee, Min-Gwan Kim, Hoh In, Chul-Hyun Cho, and Leen Kim (2017) 'Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults.' *Psychiatry Investigation* 14(2), 179

Liu, Cindy H, Courtney Stevens, Sylvia HM Wong, Miwa Yasui, and Justin A Chen (2019) 'The prevalence and predictors of mental health diagnoses and suicide among us college students: Implications for addressing disparities in service use.' *Depression and anxiety* 36(1), 8–17

Liu, Yong (2016) 'Prevalence of healthy sleep duration among adults—united states, 2014.' *MMWR. Morbidity and mortality weekly report*

Luyster, Faith S, Patrick J Strollo Jr, Phyllis C Zee, James K Walsh et al. (2012) 'Sleep: a health imperative.' *Sleep* 35(6), 727–734

McEwen, Bruce S (2006) 'Sleep deprivation as a neurobiologic and physiologic stressor: allostasis and allostatic load.' *Metabolism* 55, S20–S23

McKenna, Benjamin S, David L Dickinson, Henry J Orff, and Sean PA Drummond (2007) 'The effects of one night of sleep deprivation on known-risk and ambiguous-risk decisions.' *Journal of sleep research* 16(3), 245–252

Milkman, Katherine L, Julia A Minson, and Kevin GM Volpp (2013) 'Holding the hunger games hostage at the gym: An evaluation of temptation bundling.' *Management Science* 60(2), 283–299

Milojevich, Helen M, and Angela F Lukowski (2016) 'Sleep and mental health in undergraduate students with generally healthy sleep habits.' *PloS one* 11(6), e0156372

Misra, Ranjita, and Michelle McKean (2000) 'College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction.' *American journal of Health studies* 16(1), 41

Monk, Timothy H, Daniel J Buysse, Julie Carrier, Bart D Billy, and Lynda R Rose (2001) 'Effects of afternoon "siesta" naps on sleep, alertness, performance, and circadian rhythms in the elderly.' *Sleep* 24(6), 680–687

Mortier, Philippe, Randy P Auerbach, Jordi Alonso, Jason Bantjes, Corina Benjet, Pim Cuijpers, David D Ebert, Jennifer Greif Green, Penelope Hasking, Matthew K Nock et al. (2018) 'Suicidal thoughts and behaviors among first-year college students: Results from the wmh-ics project.' *Journal of the American Academy of Child & Adolescent Psychiatry* 57(4), 263–273

Mullainathan, S (2014) 'Get some sleep, and wake up the gdp the new york times.' *The New York Times*

Nie, Norman H, and D Sunshine Hillygus (2002) 'The impact of internet use on sociability: Time-diary findings.' *It & Society* 1(1), 1–20

O'Donoghue, Ted, and Matthew Rabin (1999) 'Doing it now or later.' *American Economic Review* 89(1), 103–124

_ (2001) 'Choice and procrastination.' *The Quarterly Journal of Economics* 116(1), 121–160

_ (2003) 'Studying optimal paternalism, illustrated by a model of sin taxes.' *American Economic Review* 93(2), 186–191

_ (2006) 'Optimal sin taxes.' *Journal of Public Economics* 90(10-11), 1825–1849

_ (2015) 'Present bias: Lessons learned and to be learned.' *American Economic Review* 105(5), 273–79

O'Donoghue, Ted, Matthew Rabin et al. (2006) 'Incentives and self-control.' *Econometric Society Monographs* 42, 215

Patel, Mitesh S, David A Asch, and Kevin G Volpp (2015) 'Wearable devices as facilitators, not drivers, of health behavior change.' *JAMA* 313(5), 459–460

Rabin, Matthew, Ted O'Donoghue et al. (1999) 'Doing it now or later.' *American Economic Review* 89(1), 103–124

Reetz, David R, Brian Krylowicz, and Brian Mistler (2014) 'The association for university and college counseling center directors annual survey.' *Aurora* 51, 60506

Roenneberg, Till (2013) 'Chronobiology: the human sleep project.' *Nature* 498(7455), 427–428

Royer, Heather, Mark Stehr, and Justin Sydnor (2015) 'Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company.' *American Economic Journal: Applied Economics* 7(3), 51–84

Schilbach, Frank (2019) 'Alcohol and self-control: A field experiment in india.' *American Economic Review* 109(4), 1290–1322

Smith, Austin C (2016) 'Spring forward at your own risk: Daylight saving time and fatal vehicle crashes.' *American Economic Journal: Applied Economics* 8(2), 65–91

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen (2010) 'Risk and time preferences: Linking experimental and household survey data from vietnam.' *American Economic Review* 100(1), 557–71

Toussaert, Séverine (2018) 'Eliciting temptation and self-control through menu choices: A lab experiment.' *Econometrica* 86(3), 859–889

Trockel, Mickey T, Michael D Barnes, and Dennis L Egget (2000) 'Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors.' *Journal of American college health* 49(3), 125–131

Twenge, Jean M, Zlatan Krizan, and Garrett Hisler (2017) 'Decreases in self-reported sleep duration among us adolescents 2009–2015 and association with new media screen time.' *Sleep Medicine* 39, 47–53

Volpp, Kevin G, Andrea B Troxel, Mark V Pauly, Henry A Glick, Andrea Puig, David A Asch, Robert Galvin, Jingsan Zhu, Fei Wan, Jill DeGuzman et al. (2009) 'A randomized, controlled trial of financial incentives for smoking cessation.' *New England Journal of Medicine* 360(7), 699–709

Watson, Nathaniel F, M Safwan Badr, Gregory Belenky, Donald L Bliwise, Orfeu M Buxton, Daniel Buysse, David F Dinges, James Gangwisch, Michael A Grandner et al. (2015) 'Recommended amount of sleep for a healthy adult: a joint consensus statement of the american academy of sleep medicine and sleep research society.' *Sleep* 38(6), 843–844

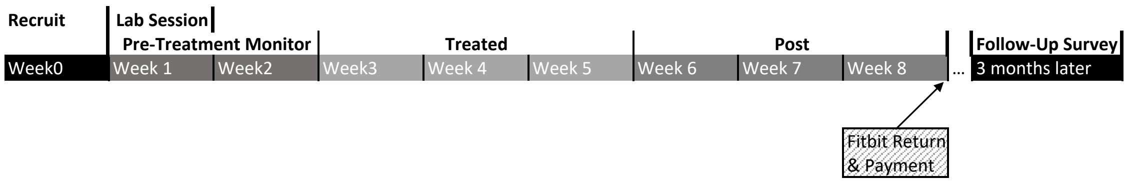
Weber, Elke U, Ann-Renee Blais, and Nancy E Betz (2002) 'A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors.' *Journal of Behavioral Decision Making* 15(4), 263–290

Wood, Wendy, and David T Neal (2007) 'A new look at habits and the habit-goal interface.' *Psychological review* 114(4), 843

Zimmermann, Florian (2019) 'The dynamics of motivated beliefs.' *American Economic Review*

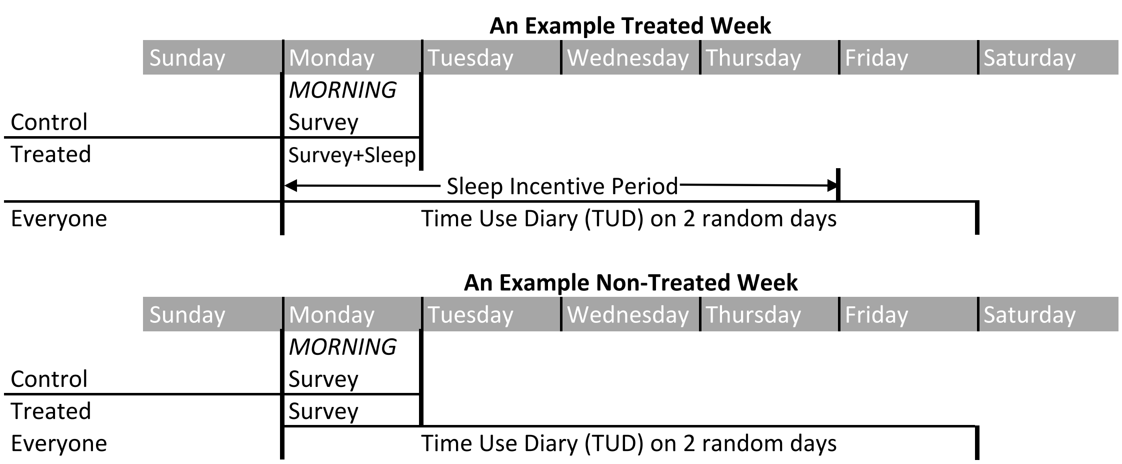
Figures

Figure 1: Design Illustration



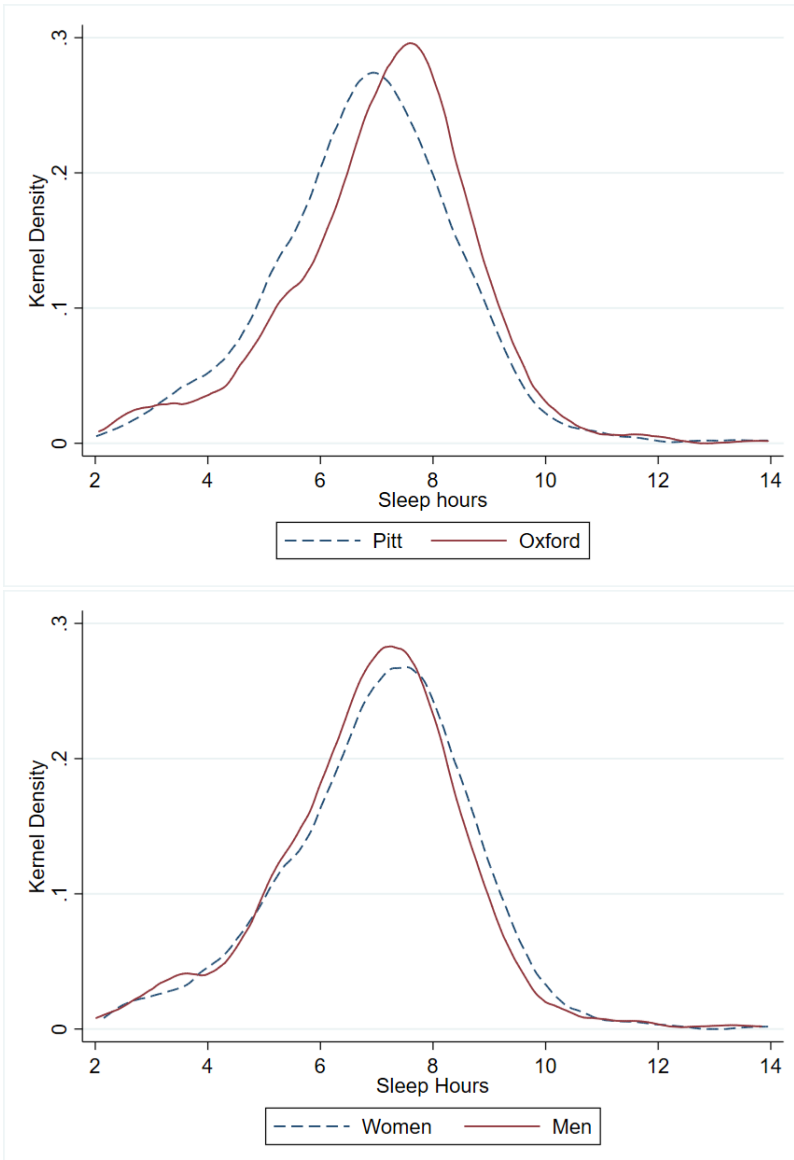
Notes - The above figure describes the timeline of our experiment for individuals in our baseline treatment (Treatment 1).

Figure 2: Treatment Week



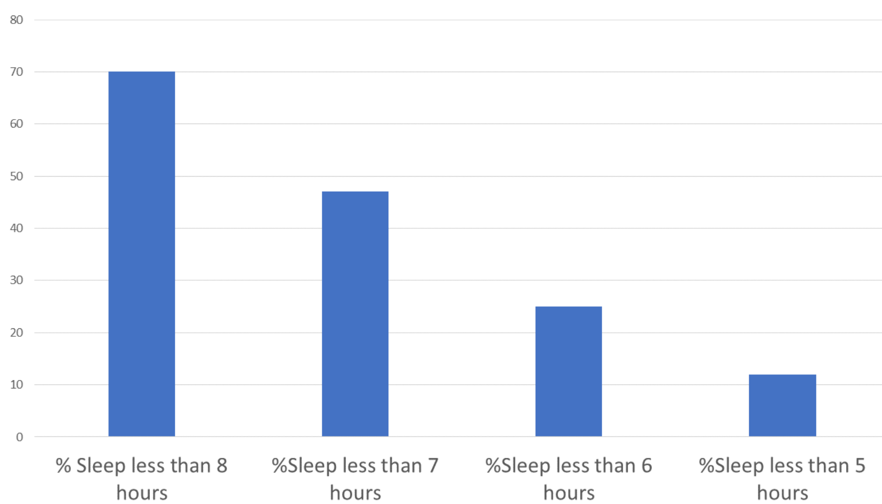
Notes - The above figure describes the typical week timeline during the experiment.

Figure 3: Sleep duration before intervention



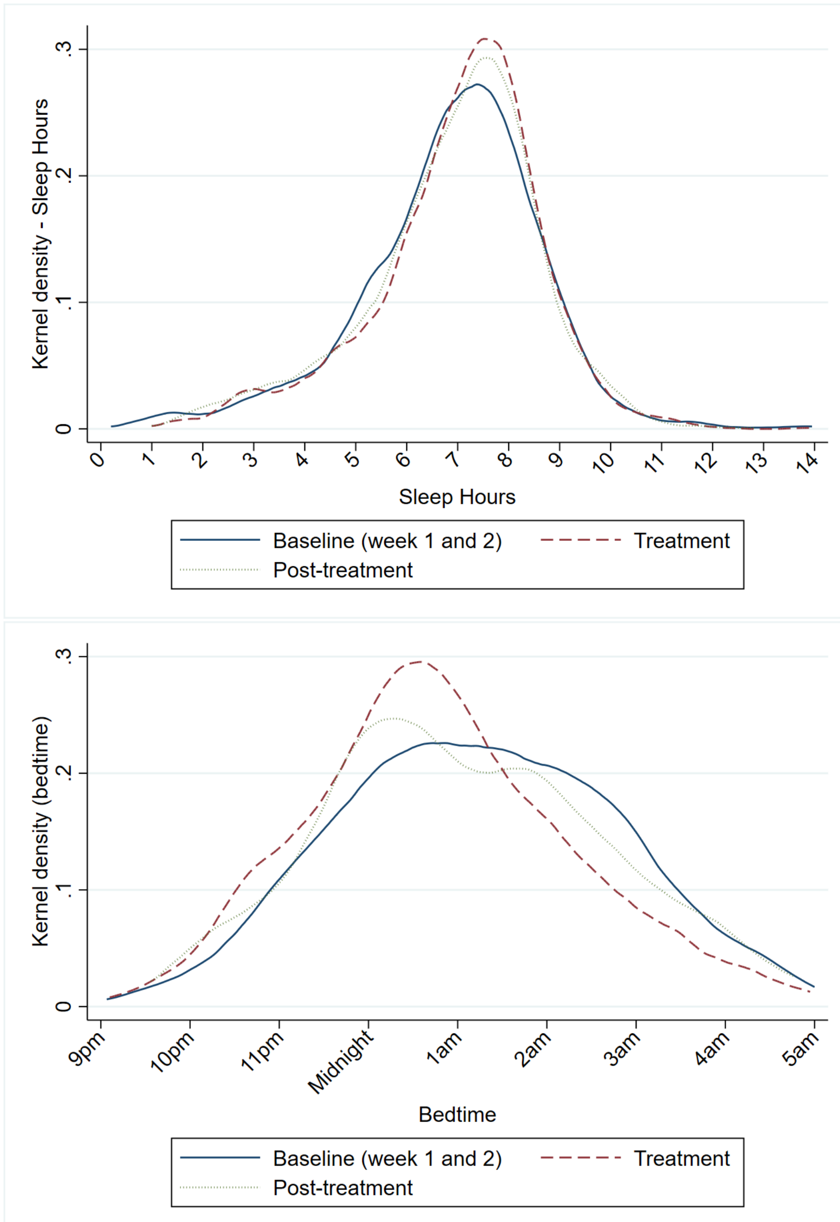
Notes - The top figure plots the distribution of sleep among subjects in Pittsburgh (dashed blue line) and Oxford (solid red line). The bottom figure above plots the distribution of sleep among women (dashed blue line) and men (solid red line). Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.

Figure 4: Insufficient sleep



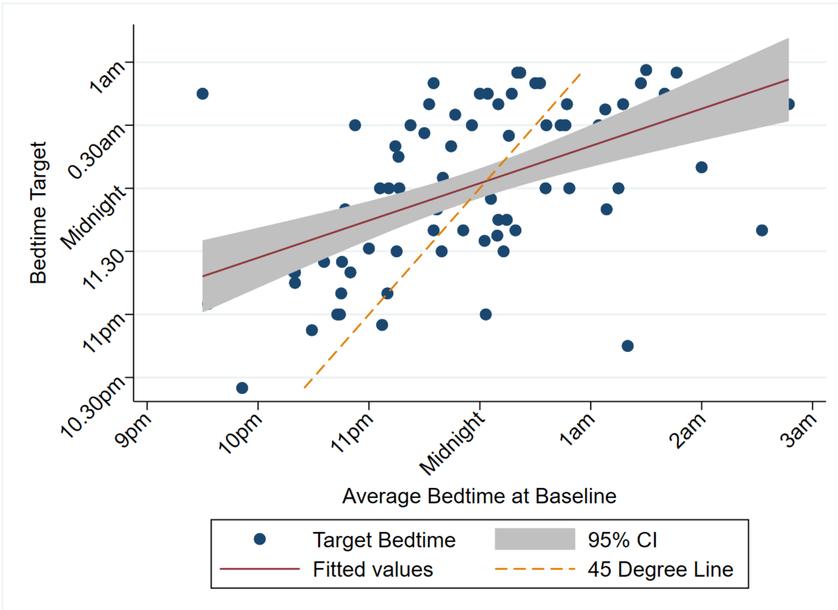
Notes - The figure above documents the share of individuals sleeping less than 8, 7, 6, and 5 hours in our sample at baseline. Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.

Figure 5: Sleep duration and bedtime distribution before and after treatment



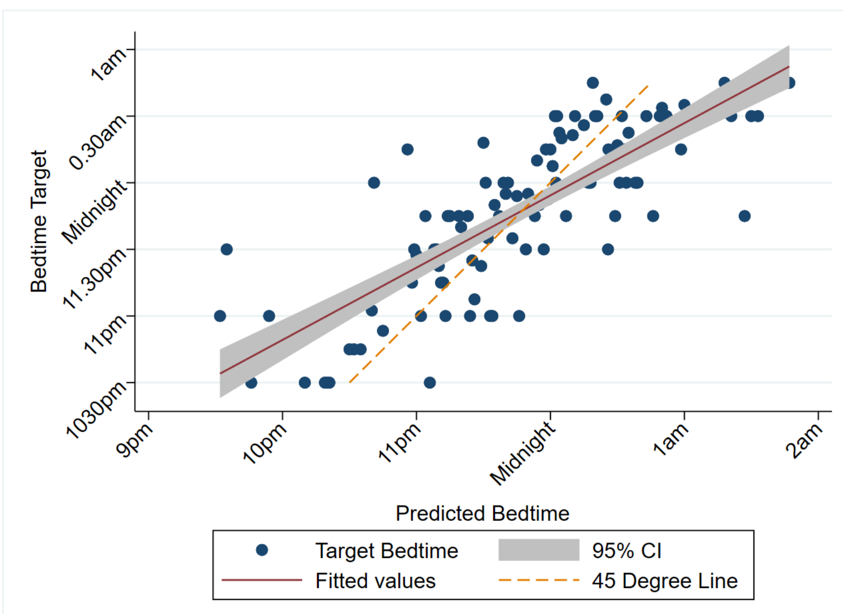
Notes - The top (bottom) figure plots the distribution of sleep duration (bedtime) before, during, and after treatment. In the figure, we pooled all the treatments. Data are drawn from the Fitbit devices.

Figure 6: Bedtime Targets and Pre-treatment Bedtimes



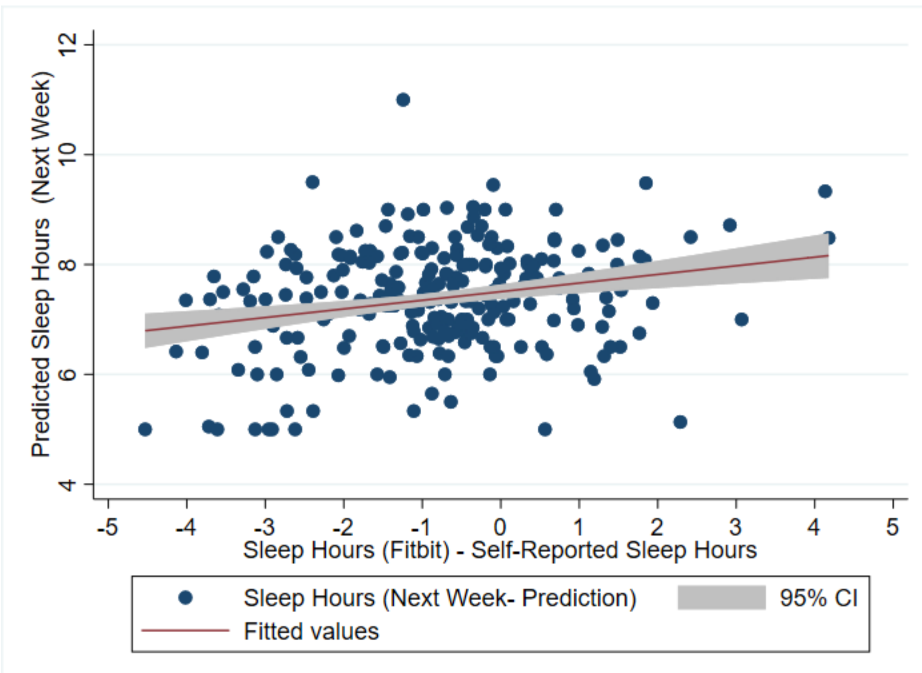
Notes - The figure presents average pre-treatment bedtime and target bedtime in the first treatment week by subjects.

Figure 7: Bedtime Targets and Predicted Bedtime



Notes - The figure presents predicted and target bedtimes by subjects.

Figure 8: Overconfidence and Beliefs



Notes - The figure plots the relationship between how many hours subjects predict to sleep in the following week and the difference between self-reported sleep and sleep as measured by Fitbits the night before the survey.

Tables

Table 1: Differences between Treatment and Control at Baseline

	(1)	(2)	(3)	(4)
	Wave 1-5	Wave 2-5	UK	US
		Treatment-Control		
Dep.Var.:				
Day 1 Survey, Univariate Regressions				
female	0.0454	0.0313	0.1163	-0.0499
	(0.060)	-0.062	-0.086	(0.083)
age	0.0381	-0.1073	-1.1272*	0.7365**
	(0.392)	-0.403	-0.605	(0.372)
white	0.1048*	0.1146*	0.0747	0.1273
	(0.059)	-0.061	-0.085	(0.083)
asian	-0.0190	-0.0181	-0.0142	-0.0260
	(0.053)	-0.054	-0.077	(0.073)
black	-0.0667*	-0.0673*	-0.0151	-0.1002*
	(0.036)	-0.037	-0.033	(0.060)
other	-0.0190	-0.0292	-0.0454	-0.0011
	(0.032)	-0.032	-0.054	(0.037)
smoker	-0.1000*	-0.0936*	-0.097	-0.1054
	(0.052)	-0.053	-0.076	(0.071)
drinker	-0.0000	-0.0068	-0.0949	0.0689
	(0.053)	-0.054	-0.083	(0.065)
poor health	0.0160	0.0158	0.0059	0.0195
	(0.033)	-0.034	-0.054	(0.040)
very satisfied	0.0476	0.0425	0.0081	0.0913
with life	(0.059)	-0.061	-0.086	(0.083)
most depressed	0.0238	0.0336	0.0442**	0.0092
	(0.022)	-0.024	-0.019	(0.039)
ideal sleep (hours)	0.2286	0.2838	-0.1886	0.5768
	(0.283)	-0.29	-0.393	(0.399)
BMI	0.3751	-0.1331	1.2747	-0.5887
	(0.964)	-0.918	-1.684	(0.759)
overweight	-0.0262	-0.0362	-0.0248	-0.0199
	(0.053)	-0.054	-0.074	(0.076)
sleep quality	-0.2048	-0.263	-0.3691	-0.1006
	(0.182)	-0.19	-0.242	(0.267)
Sleep hours	0.0202	0.0107	0.0514	-0.0046
	(0.199)	-0.203	-0.154	(0.344)
Fitbit data				
(first 2 weeks)				
Sleep hours	-0.319	-0.370	-0.671	-0.036
(fitbit measured)	(0.292)	(0.299)	(0.472)	(0.333)
Observations	319	280	163	156

Notes - Data are drawn from the first-day survey and from the first two weeks of fitbit data (baseline period). Each cell reports the coefficient of a univariate regression of covariates on treatment.

Table 2: Summary Statistics, Baseline (Survey-based metrics)

Variable	Mean	Std. Dev.
Demographics		
Female	0.41	0.49
Age	21.54	3.90
White	0.58	0.49
Asian	0.22	0.42
Black	0.09	0.28
Other	0.11	0.31
Health and Behaviors		
Poor health	0.11	0.31
Weight (kg)	69.44	15.39
Height (cm)	171.46	12.14
BMI	23.97	9.72
Obese	0.05	0.22
Overweight	0.24	0.43
Ever smoked	0.23	0.41
Drinks (more than once per week)	0.26	.44
Depression symptoms		
Rarely	0.44	0.49
5-7 days	0.039	0.19
1-2 days	0.36	0.48
3-4 days	0.15	0.35
Life satisfaction		
Completely satisfied	0.06	0.24
Very satisfied	0.44	0.49
Somewhat satisfied	0.41	0.49
Not satisfied (or not at all)	0.09	0.27
Sleep		
Sleep last month	7.17	.97
Sleep during term	6.90	1.32
Less than 7 hrs sleep	0.46	0.50
Ideal sleep	7.97	.78
Sleep quality (1-10)	6.62	1.61
# days falling asleep	3.79	4.94
# days not rested	10.51	6.86

Notes - Summary statistics are drawn from the Day 1 Survey.

Table 3: Incentives and Sleep

VARIABLES	(1) 7<Sleep<9	(2)	(3) Sleep<6 hours	(4)
Treatment	0.0850***	0.0493***	-0.0584***	-0.0316*
	(0.024)	(0.018)	(0.022)	(0.016)
Post-Treatment	0.0418	0.0053	-0.0589*	-0.0422**
	(0.038)	(0.024)	(0.032)	(0.021)
Individual fixed effects		YES		YES
Observations	7,690	7,690	7,690	7,690
Individuals	280	280	280	280
Mean of Dep. Var.	0.453	0.453	0.250	0.250
Std.Dev. of Dep. Var.	0.498	0.498	0.433	0.433

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). The dependent variable in column 1 and 2 is an indicator for sleeping between 7 and 9 hours. The dependent variable in column 3 and 4 is an indicator for sleeping less than 6 hours. Columns 2 and 4 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table 4: Incentives to Sleep and Other Outcomes (Fitbit data)

VARIABLES	(1) Sleep Efficiency	(2)	(3) Low Resting Heart Rate	(4)	(5) Sedentary Minutes	(6)
Treatment	0.1984	-0.2502	0.0392	0.0145	-14.2803**	-9.1444*
	(0.575)	(0.373)	(0.028)	(0.010)	(6.811)	(4.860)
Post-Treatment	0.6906	-0.3316	0.0848*	0.0026	-1.3966	5.2014
	(0.701)	(0.285)	(0.045)	(0.022)	(10.765)	(7.249)
Individual fixed effects		YES		YES		YES
Observations	7,690	7,690	7,690	7,690	7,690	7,690
Mean of Dep. Var.	92.65	92.65	0.203	0.203	720.3	720.3
Std.Dev. of Dep. Var.	8.459	8.459	0.403	0.403	143.7	143.7

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2, 4, and 6 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table 5: Sleep, Incentives to Sleep, and Academic Performance

	(1) Letter Grade Change	(2) Percentile Change
Sleep measures:		
7<Sleep<9	0.710***	8.421*
	(0.183)	(4.636)
Sleep<6	-0.557***	-14.49***
	(0.171)	(4.361)
Sleep Duration	0.0753**	1.654*
	(0.035)	(0.884)
SD of Sleep	-0.0619	-2.681*
	(0.0647)	(1.559)
Incentives:		
Any Treatment	-0.0459	6.304**
	(0.115)	(3.069)
Large Treatment	-0.0626	8.738**
	(0.127)	(3.636)
Compliance:		
Achievement Rate	0.422***	3.386
	(0.151)	(3.023)
High Achiever	0.162*	2.627
	(0.084)	(2.384)

Notes - Each cell reports a separate regression of self-reported letter grade change or percentile change on sleep, treatment, or compliance indicators. The grade and percentile changes are drawn the follow-up survey. All regressions control for the wave of the experiment and the previous semester grade or percentile.

Table 6: Time Preferences and Sleep Duration at Baseline

	(1)	(2)	(3)	(4)	(5)	(6)
	Self-reported (Day 1 Survey)			Actual Sleep (Fitbit)		
	Sleep hours	Sleep<7hrs	7≤Sleep≤9	Sleep hours	Sleep<7hrs	7≤Sleep≤9
Panel A: Present-Bias						
Present-biased	-0.1067	0.0438	-0.0566	-0.2914	0.0929	-0.1034
	(0.142)	(0.069)	(0.069)	(0.284)	(0.069)	(0.068)
Observations	319	319	319	319	319	319
Mean of Dep. Var.	6.845	0.458	0.522	7.078	0.465	0.468
Std.Dev. of Dep. Var.	0.984	0.499	0.500	1.979	0.500	0.500
Panel B: Impatience						
Impatient	0.2374	0.0140	-0.0381	0.1315	-0.0352	0.0099
	(0.269)	(0.066)	(0.067)	(0.243)	(0.066)	(0.066)
Observations	319	319	319	319	319	319
Mean of Dep. Var.	6.895	0.462	0.516	6.895	0.500	0.465
Std.Dev. of Dep. Var.	1.380	0.499	0.501	1.667	0.501	0.500
Panel C: Risk Aversion						
Risk Averse	0.2873**	-0.1507**	0.1747**	0.2043	-0.0878	0.0970
	(0.134)	(0.073)	(0.073)	(0.341)	(0.076)	(0.077)
Observations	319	319	319	319	319	319
Mean of Dep. Var.	6.845	0.458	0.522	7.078	0.465	0.468
Std.Dev. of Dep. Var.	0.984	0.499	0.500	1.979	0.500	0.500

Notes - Data for time and risk preferences are drawn from the first-day survey (columns 1-3). Data on sleep are drawn from fitbit data collected in first two weeks of the experiment and before the intervention (columns 4-6).

Table 7: Perceived own and other's sleep quality and sleep deprivation risks

	Own		Others		Difference	
Variables	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
Sleep quality (1-10)	6.67	1.58	6.08	1.35	0.58	1.73
Sleep duration	6.92	0.91	6.60	0.97	0.31	1.24
Sleep deprivation risks for:						
Mental alertness (1-100)	25.96	12.80	59.73	24.26	-33.93	24.76
Weight gain (1-100)	39.20	24.54	51.17	22.40	-12.00	22.95
Insomnia (1-100)	23.10	17.86	35.32	21.95	-12.72	19.88
Getting a cold (1-100)	37.84	23.50	45.46	25.01	-7.88	20.86
Arterial (1-100)	30.65	21.98	34.51	22.16	-3.34	18.47
Average risk	31.81	13.02	45.40	16.78	-13.72	13.74
Observations	319	319	319	319	319	319

Notes - We report averages and standard deviations obtained from our day 1 of the experiment survey.

Table 8: Overconfidence and Sleep Duration (Self-reported vs Fitbit data)

	Self-reported (Day 1 Survey)			Actual Sleep (Fitbit)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Sleep hours	Sleep<7hrs	$7\leq$Sleep≤ 9	Sleep Hours	Sleep<7hrs	$7\leq$Sleep≤ 9
Overconfident	0.8867***	-0.3441***	0.3695***	-1.1449***	0.2824***	-0.2179***
	(0.109)	(0.059)	(0.059)	(0.259)	(0.066)	(0.066)
Observations	319	319	319	319	319	319
Mean of Dep. Var.	6.845	0.458	0.522	7.078	0.465	0.468
Std.Dev. of Dep. Var.	0.984	0.499	0.500	1.979	0.500	0.500

Notes - Data are drawn from the first-day survey (columns 1-3) and the Fitbit data for the first two weeks of the experiment before intervention (columns 4-6). Our index of overconfidence is defined in Section 5.2.

Table 9: Incentives and Time Allocation

Panel A	(1) Sleep (hours)	(2) Time Use Sleep 7$<$ hours $<$ 9	(3) Time Use Sleep $<$6 hours	(4) Total Study hours	(5) Total Work hours	(6) Total Care hours	(7) Total Exercise hours
Treatment	0.041	-0.004	-0.045**	0.055	-0.106	-0.013	-0.022
	-0.144	(0.044)	(0.022)	(0.206)	(0.153)	(0.087)	(0.053)
Post-Treatment	-0.06	-0.066	-0.029	-0.103	0.107	-0.065	-0.062
	0.155	(0.045)	(0.025)	(0.246)	(0.170)	(0.100)	(0.072)
Observations	1,363	1,363	1,363	1,363	1,363	1,363	1,363
Number of id	215	215	215	215	215	215	215
Mean of Dep. Var.	8.212	0.591	0.0602	4.864	1.982	2.900	0.425
Std.Dev. of Dep. Var.	1.727	0.492	0.238	3.340	2.823	1.329	0.773
Panel B	**Total Relaxing hours**	**Total Other hours**	**Total Social hours**	**Total TV&Internet hours**	**Total TV hours**	**Total Internet hours**	**Total Gaming hours**
Treatment	0.051	-0.008	0.165	-0.114	-0.175*	0.112	-0.052
	(0.171)	(0.094)	(0.130)	(0.140)	(0.105)	(0.092)	(0.066)
Post-Treatment	0.244	-0.188	0.338**	-0.094	-0.328***	0.340***	-0.107**
	(0.223)	(0.119)	(0.167)	(0.166)	(0.117)	(0.126)	(0.050)
Observations	1,363	1,363	1,363	1,363	1,363	1,363	1,363
Number of id	215	215	215	215	215	215	215
Mean of Dep. Var.	4.066	1.550	1.800	2.266	0.858	1.145	0.263
Std.Dev. of Dep. Var.	2.511	1.530	2.065	2.076	1.376	1.569	0.848

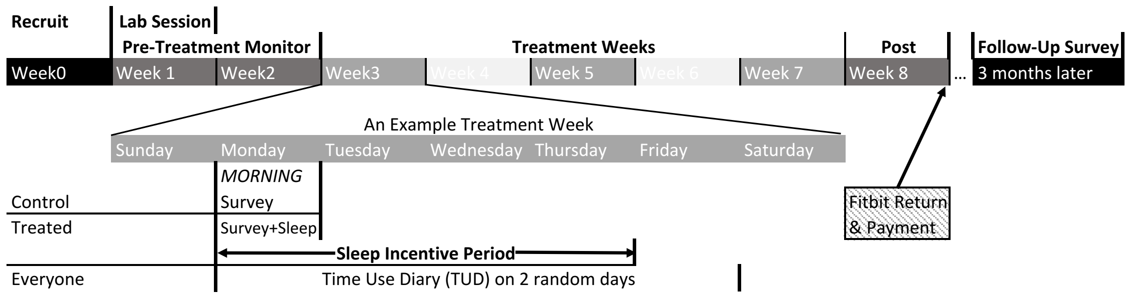
Notes - All the dependent variables are drawn from the time-use surveys collected throughout the experimetn. All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). All estimate include individual fixed effects. Standard errors are clustered at the individual level. Data for Panel A Column 1 comes from the Fitbit Data. The remaining data comes from the Time Use Surveys.

APPENDIX

A Additional Tables and Figures

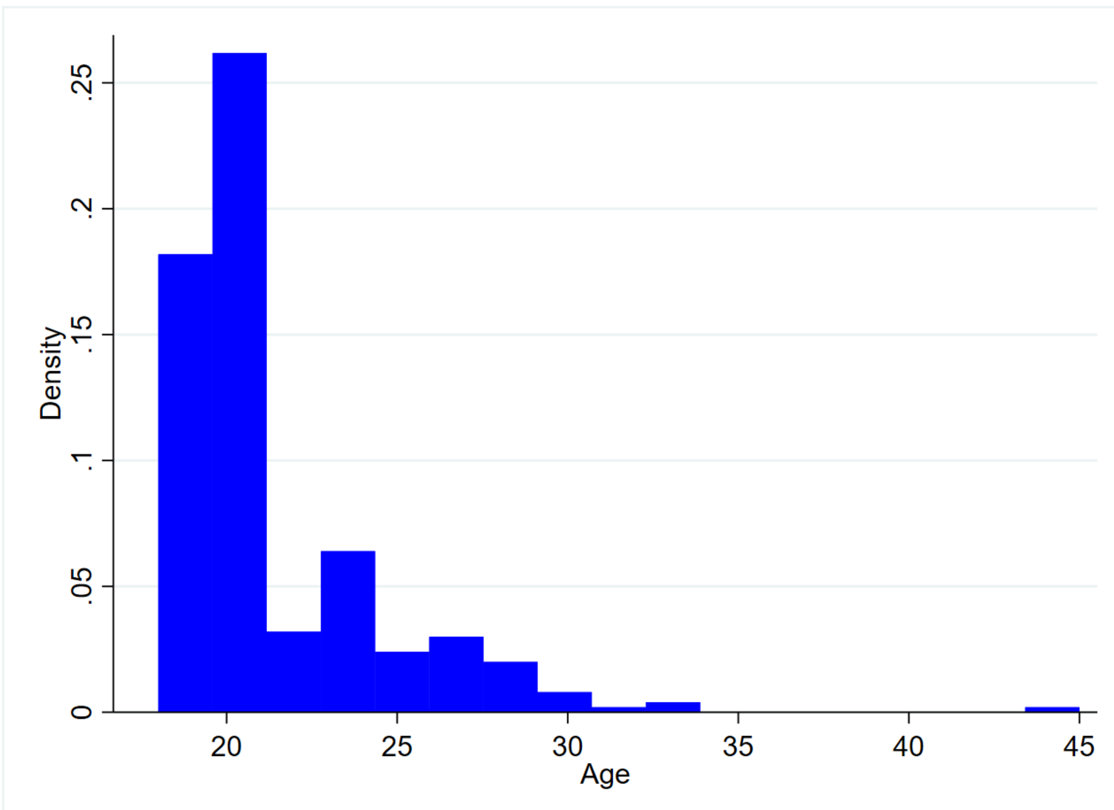
Appendix Figures

Figure A.1: Design Illustration: Biweekly Intervention



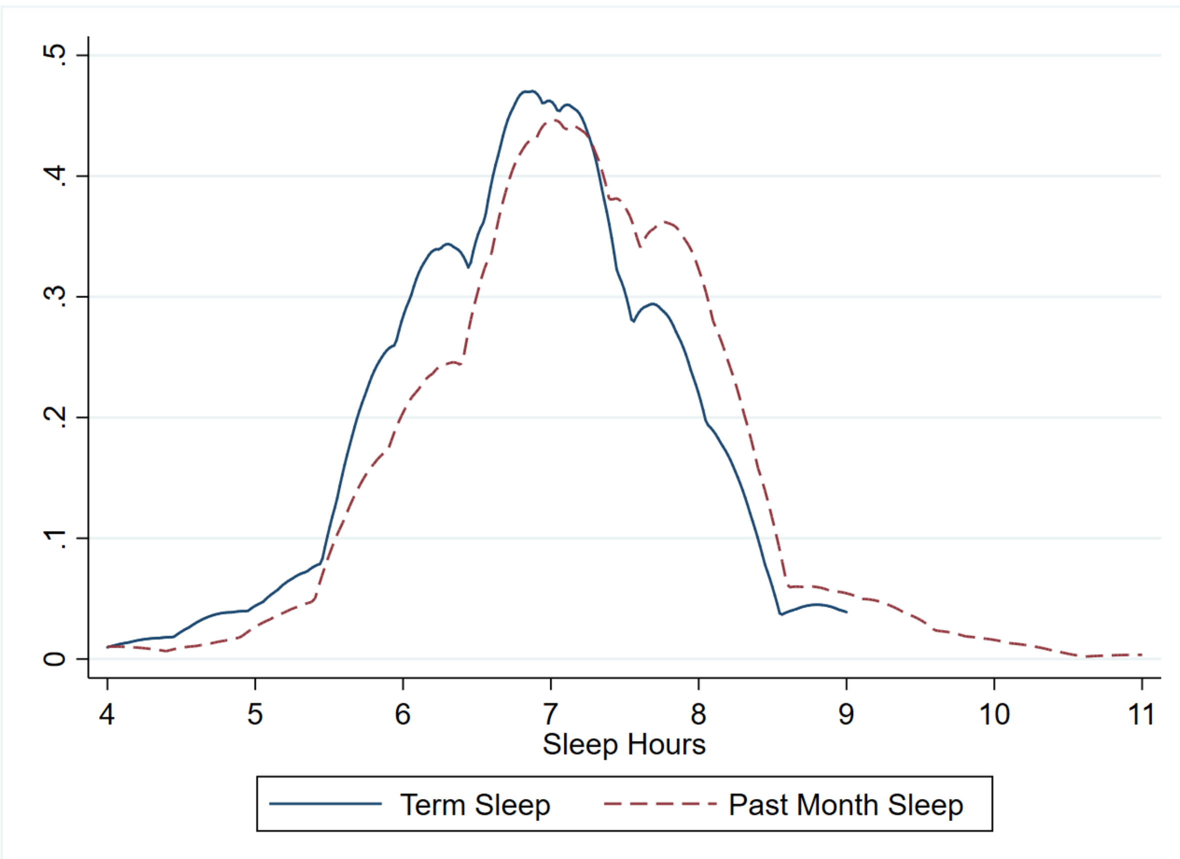
Notes - The above figure describes the timeline of our experiment for individuals in our biweekly treatment.

Figure A.2: Age of Participants



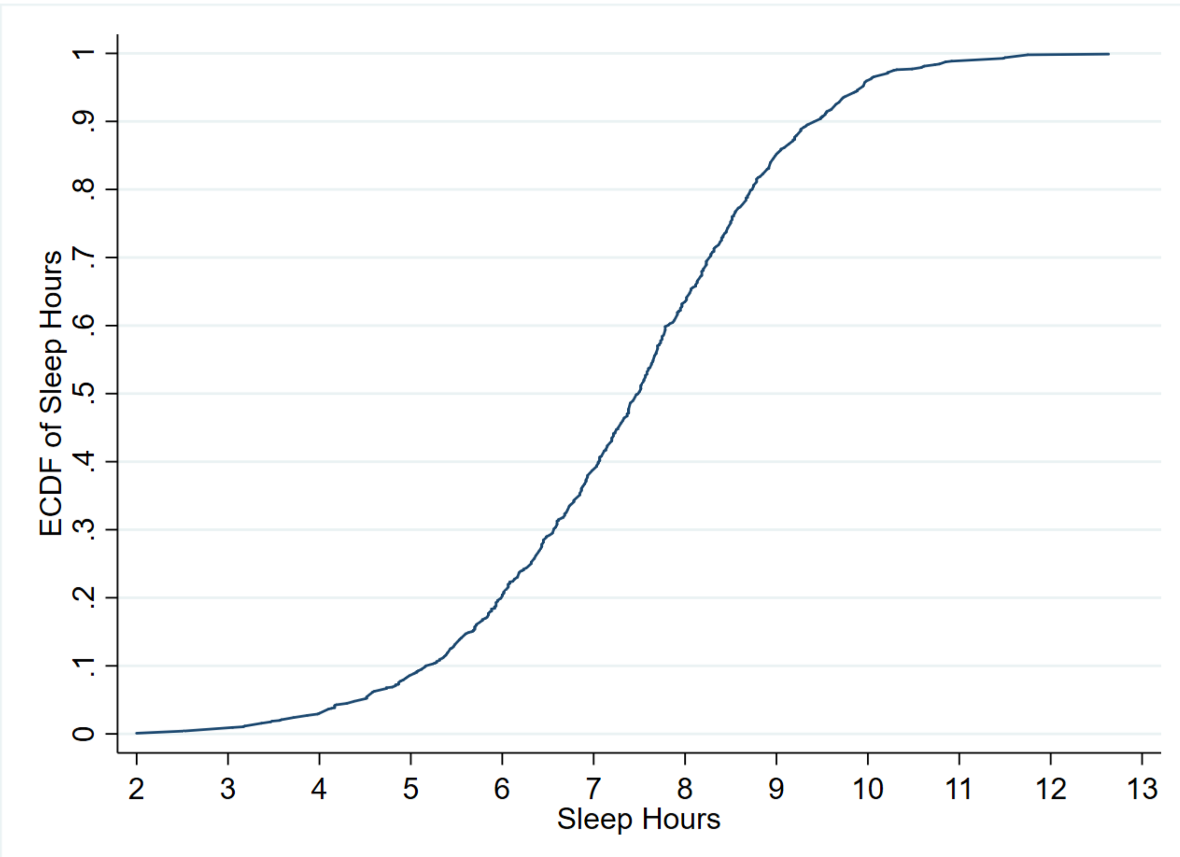
Notes - The above histogram describe the age distribution in our sample.

Figure A.3: Self-reported sleep duration at baseline



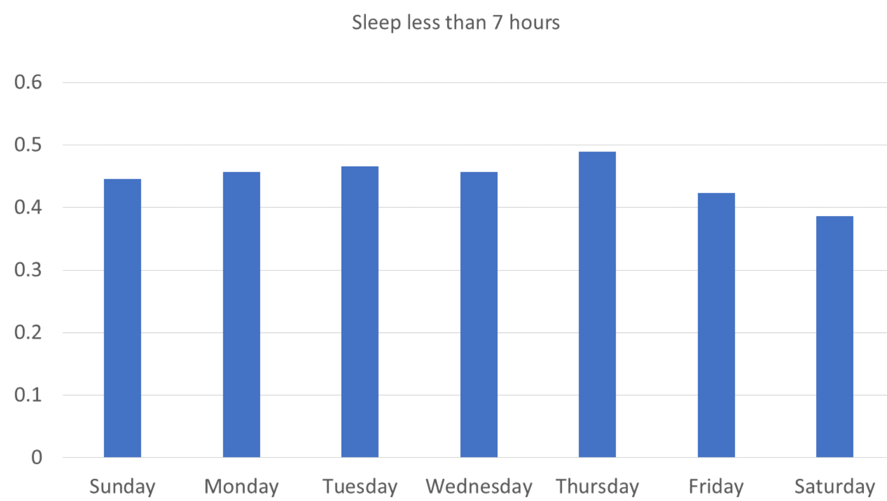
Notes - The figure reports self-reported sleep duration during term (solid) and over the month preceding the survey (dashed) from the Day 1 Survey, which occurred at the beginning of the term.

Figure A.4: Sleep duration before intervention



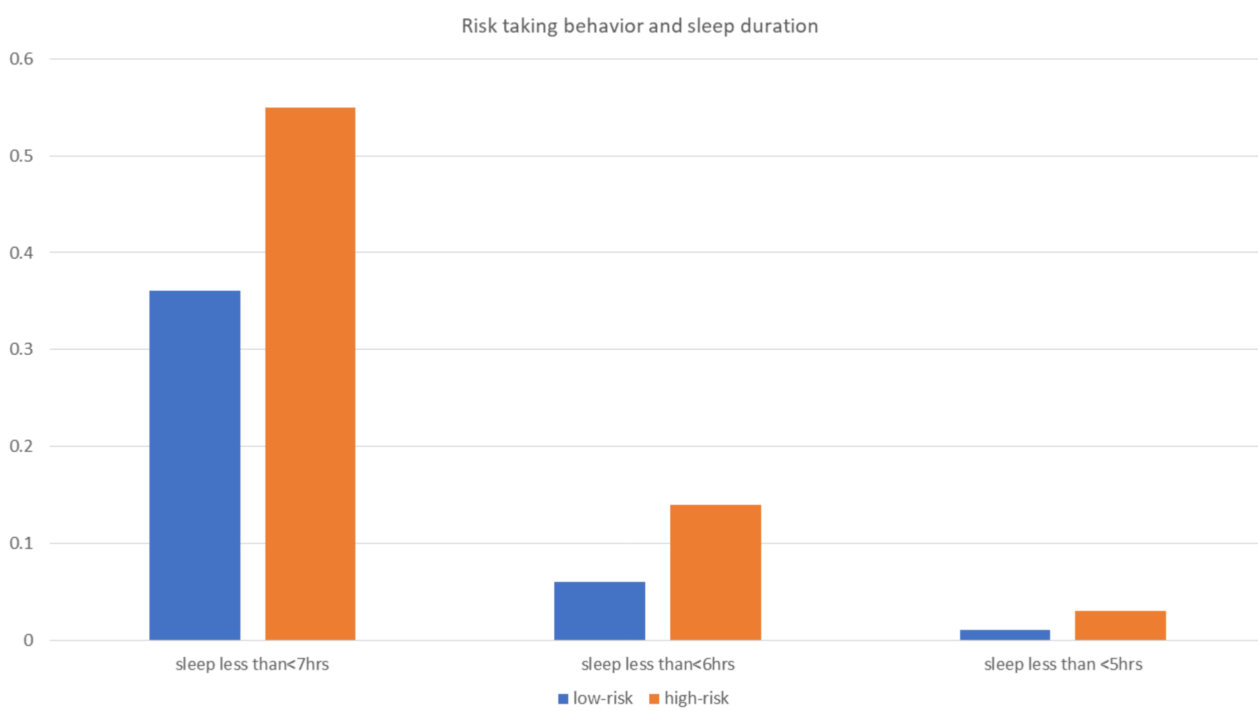
Notes - The figure plots the cumulative distribution function of sleep hours in our sample at baseline. Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.

Figure A.5: Sleep duration over the week



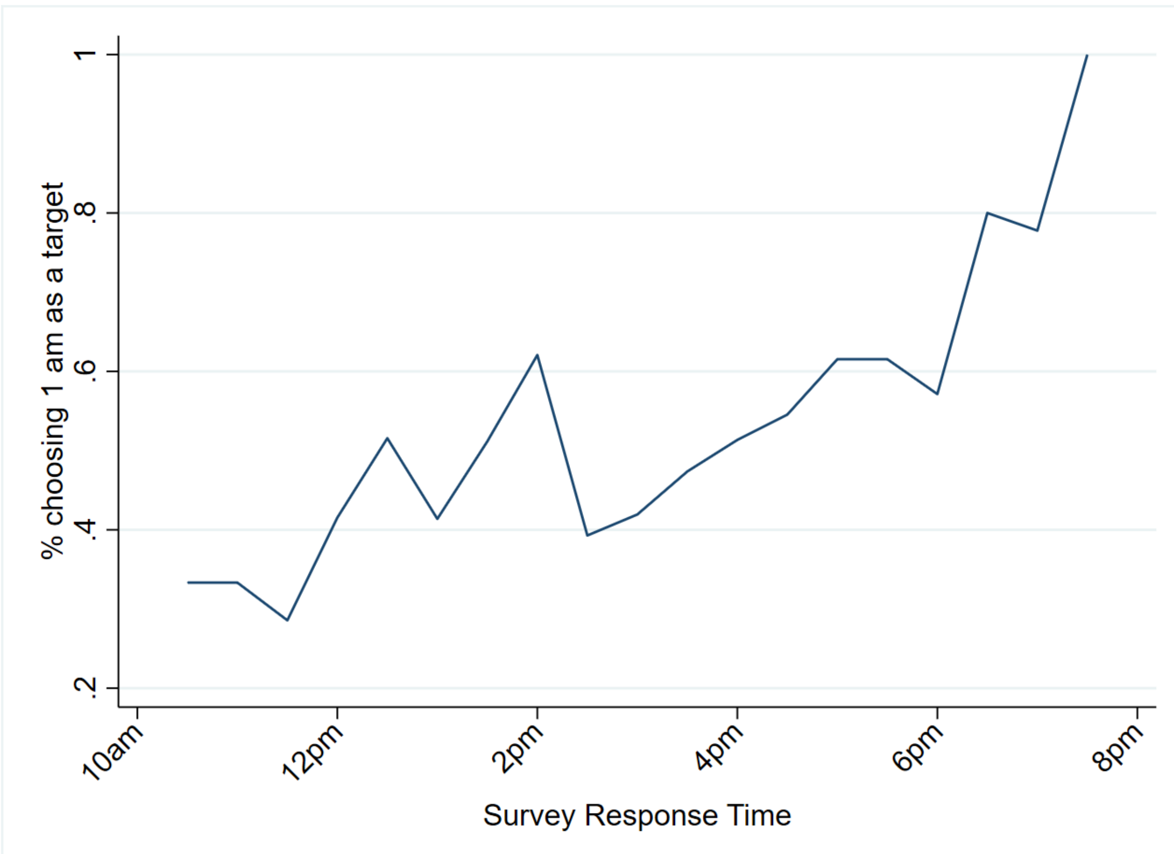
Notes - The figure describes the proportion of subjects sleeping less than 7 hours throughout the week. Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.

Figure A.6: Self-reported sleep duration at baseline and risk taking behavior



Notes - The figure reports the share of individuals self-reporting sleeping less than 7, 6, and 5 hours, respectively, in the Day 1 Survey among low-risk individuals (blue) and high-risk individuals (orange), were low-risk (high-risk) are subjects with a risk aversion below (above) the median.

Figure A.7: Timing of Survey Response and Bedtime Target Choice



Notes - The figure reports the share of individuals choosing the least binding bedtime target (1 am) by the timing of the survey (10 am-8 pm). The data are drawn from the weekly surveys during intervention weeks.

Appendix Tables

Table A.1: Summary of Treatments

Treatment	Wave	Location	Time	Incentive	Prediction Reward
Control	1	Oxford	Oct-Dec 2016	None	No
Treatment	1	Oxford	Oct-Dec 2016	Biweekly, Weak	No
Treatment	2	Oxford	Apr-Jun 2017	Weekly, Strong	Yes, 1
Treatment	3	Oxford	Oct-Dec 2017	Weekly, Strong	Yes, 3
Treatment	4	Pittsburgh	Jan-Mar 2018	Biweekly, Strong	Yes, 1
Control	5	Pittsburgh	Sep-Nov 2018	None	No
Treatment	5	Pittsburgh	Sep-Nov 2018	Weekly, Strong	Yes, 3

Notes - The table above describes the location, timing and incentive structure used in the different waves of the experiment.

Table A.2: Baseline Characteristics and Attrition

Dep. Var.	Female	Age	White	Black	Asian	Other	Last month sleep
Withdrew from the experiment	0.100	0.272	-0.102	0.053	-0.027	0.076	-0.182
	(0.079)	(0.486)	(0.079)	(0.045)	(0.067)	(0.050)	(0.155)
Observations	359	359	359	359	359	359	359
Dep. Var.	**Sleep during term**	**Slee $<$ 7 hrs during term**	**Ever smoked**	**Ideal sleep hours**	**BMI**	**Overweight**	**Obese**
Withdrew from the experiment	0.093	-0.116	0.091	-0.167	-1.041	-0.071	0.012
	(0.212)	(0.079)	(0.067)	(0.127)	(1.564)	(0.070)	(0.039)
Observations	359	359	359	359	359	359	359

Notes - Data are drawn by the Day 1 Survey. Each column reports a univariate regression estimate of the dependent variable (baseline characteristics) on a dummy indicating whether the individual withdrew from the experiment.

Table A.3: Comparisons of Sleep Measurements

	Sleep Duration	$7 \leq \text{Sleep} \leq 9$	$\text{Sleep} < 6$
Fitbit	7.02	0.47	0.23
	(1.76)	(0.50)	(0.42)
Self-Reported	7.07	0.61	0.10
	(1.08)	(0.49)	(0.31)
Time Use	8.15	0.59	0.07
	(1.74)	(0.49)	(0.25)

Notes - This table compares averages of our three different measures of sleep collected before our intervention started. Standard deviations are reported in parentheses. The first row (Fitbit) reports the sleep measures derived from the Fitbit data. The second row (Self-Reported) reports the sleep measures elicited in Day 1 Survey. The third row (Time Use) reports the sleep measures based on the time use surveys.

Table A.4: Intention to improve sleep

	Wants to improve sleep duration	Wants to improve bedtime
	%	%
Definitely yes	17.77	19.34
Probably yes	43.39	41.56
Might or might not	22.73	22.22
Probably not	14.88	15.64
Definitely not	1.24	1.23
Observations	359	359

Notes - Data are drawn from Day 1 Survey.

Table A.5: Correlations between sleep, health, and well-being

VARIABLES	Good Health	Obese	Overweight	Depressed	Satisfied
$7 \leq$ Sleep ≤ 9	0.057*	-0.064**	-0.123***	-0.054**	0.259***
	(0.034)	(0.026)	(0.046)	(0.021)	(0.051)
Observations	359	359	359	359	359
R-squared	0.008	0.018	0.020	0.018	0.067
Mean of Dep. Var.	0.880	0.0616	0.252	0.0410	0.489
Std.Dev. of Dep. Var.	0.326	0.241	0.435	0.199	0.501
Sleep Less than 7hrs	-0.030	0.066**	0.120***	0.057***	-0.246***
	(0.033)	(0.026)	(0.046)	(0.022)	(0.051)
Observations	359	359	359	v	359
Mean of Dep. Var.	0.890	0.0616	0.252	0.0414	0.494
Std.Dev. of Dep. Var.	0.314	0.241	0.435	0.200	0.501

Notes - Data are drawn from Day 1 Survey. For this analysis, we used the self-reported measure of sleep duration obtained in the survey.

Table A.6: Incentives and Sleep, Other Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep.Var.	Sleep Hours		ln(Sleep Hours)		Sleep<7		Sleep<6		Sleep<5	
Treatment	0.2395***	0.1243*	0.0488***	0.0293**	-0.0789***	-0.0413**	-0.0584***	-0.0316*	-0.0329**	-0.0147
	(0.089)	(0.071)	(0.016)	(0.013)	(0.025)	(0.018)	(0.022)	(0.016)	(0.016)	(0.013)
Post-Treatment	0.2060	0.1043	0.0459*	0.0255	-0.0394	-0.0057	-0.0589*	-0.0422**	-0.0305	-0.0096
	(0.127)	(0.082)	(0.023)	(0.016)	(0.038)	(0.023)	(0.032)	(0.021)	(0.023)	(0.015)
Individual fixed effects		YES		YES		YES		YES		YES
Observations	7,690	7,690	7,690	7,690	7,690	7,690	7,690	7,690	7,690	7,690
Mean of Dep. Var.	6.922	6.922	1.891	1.891	0.466	0.466	0.250	0.250	0.131	0.131
Std.Dev. of Dep. Var.	1.849	1.849	0.332	0.332	0.499	0.499	0.433	0.433	0.337	0.337
Number of id		280		280		280		280		280

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2, 4, 6, 8 and 10 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

Table A.7: Incentives and Sleep, Waves 1 and 5 (including week-by-wave fixed effects)

	(1)	(2)	(3)	(4)	(5)
Dep.Var.:	sleep hours	7≤sleep≤9	sleep<7hr	sleep<6hr	sleep<5hrs
Treatment	0.3923*	0.1349**	-0.1230*	-0.1269**	-0.0897**
	(0.233)	(0.065)	(0.065)	(0.062)	(0.045)
Post-Treatment	0.4728*	0.0995	-0.1119	-0.1378*	-0.0851*
	(0.260)	(0.076)	(0.078)	(0.070)	(0.048)
Observations	3,785	3,785	3,785	3,785	3,785
Mean of Dep. Var.	6.903	0.453	0.476	0.25	0.121
Std.Dev. of Dep. Var.	1.799	0.498	0.499	0.433	0.326

Notes - All estimates include controls for gender, a quadratic in age, week-by-wave fixed effects, and a control for the country. Standard errors clutered at the individual level are reported in parentheses.

Table A.8: Incentives, Bedtime, and Wake up time

	(1)	(2)	(3)	(4)
VARIABLES	Bedtime		Wake up Time	
Treatment	-0.3222***	-0.2117***	-0.1111	-0.1023
	(0.105)	(0.058)	(0.127)	(0.092)
Post-Treatment	0.0015	0.0368	0.2669	0.2491
	(0.181)	(0.076)	(0.197)	(0.154)
Individual fixed effects		YES		YES
Observations	7,690	7,690	7,690	7,690
Mean of Dep. Var.	00.94	00.94	8.011	8.011
Std.Dev. of Dep. Var.	1.631	1.631	3.046	3.046
Number of id		273		273

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

Table A.9: Incentives and Sleep, Weekends

VARIABLES	(1) 7<Sleep<9	(2)	(3) Sleep<6 hours	(4)
Treatment	0.0707	0.0698	-0.137**	-0.102
	-0.0682	-0.0692	-0.0676	-0.0669
Individual fixed effects		YES		YES
Observations	3342	3342	3342	3342
Individuals	280	280	280	280
Mean of Dep. Var.	0.453	0.453	0.250	0.250
Std.Dev. of Dep. Var.	0.498	0.498	0.433	0.433

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects.Standard errors clutered at the individual level are reported in parentheses.

Table A.10: Commitment devices and Target Achievement

	Bedtime target Before 1am	Bedtime target 1am	% Difference	p-value
% Target achieved	53%	46%	7%	0.17
Achieved at least once	93%	84%	9%	0.065
	Sleep target >7hrs	Sleep target 7hrs	% Difference	p-value
% Target achieved	51%	48%	3%	0.48
Achieved at least once	92%	85%	6%	0.16
	Bedtime before 1 am & Sleep >7hrs	Bedtime 1am or Sleep=7hrs	% Difference	p-value
% Target achieved	55%	47%	8%	0.08
Achieved at least once	95%	85%	10%	0.02

Notes - Data are drawn from Fitbit data and weekly surveys collected during the weeks of the intervention.

Table A.11: Demand for Commitment and Sleep

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		$7 \leq$ Sleep ≤ 9				Sleep less than 7 hours		
Bedtime <1 am	0.0702**				-0.0591*			
	(0.033)				(0.034)			
Sleep Duration>7hrs		0.1266***				-0.1132***		
		(0.025)				(0.023)		
Bedtime<1 am & Sleep Duration>7hrs			0.1531***				-0.1349***	
			(0.027)				(0.026)	
Dominated Contract (Treatment 3)				0.0496				-0.0551
				(0.080)				(0.077)
Observations	1,420	4,566	4,566	710	1,420	4,566	4,566	710
Mean of Dep. Var.	0.511	0.146	0.102	0.0549	0.511	0.146	0.102	0.0549
Std.Dev. of Dep. Var.	0.500	0.353	0.303	0.228	0.500	0.353	0.303	0.228

Notes - All estimates include a control for insufficient sleep at baseline. The sample is restricted to treated subjects during the treatment weeks. Standard errors are clustered at the individual level and are reported in parentheses.

Table A.12: Baseline Characteristics and Sample Attrition in Follow-Up Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Female	Age	White	Black	Asian	Other	Last month sleep
No follow up	0.044	0.685**	-0.041	0.054*	0.030	-0.043	0.123
	(0.053)	(0.326)	(0.053)	(0.030)	(0.045)	(0.034)	(0.104)
Observations	359	359	359	358	359	359	359
Dep. Var.	Sleep during term	Sleep$\¡$hrs during term	Ever smoked	Ideal sleep hours	BMI	Overweight	Obese
No follow up	0.190	0.008	0.066	-0.034	1.033	-0.032	0.019
	(0.142)	(0.054)	(0.045)	(0.085)	(1.037)	(0.047)	(0.026)
Observations	359	359	359	359	359	359	359

Notes - Data are drawn by the Day 1 Survey. Each column reports a univariate regression estimate of the dependent variable (baseline characteristics) on a dummy indicating whether the individual did not respond to the follow-up survey.

Table A.13: Follow-up Sleep Quality

Variables	Sleep Quality Z-Score	
Treatment	0.459**	
	(0.229)	
Achievement Rate		0.433**
		(0.211)
R-Squared	0.132	0.164
Mean of Dep. Var.	-0.018	0.164
Std.Dev. of Dep. Var.	0.596	0.573

Notes - The table above shows regressions of treatment and bedtime target achievement rate on a Z-score of sleep quality formed from the sleep quality questions asked in the follow up survey. All estimates include controls for wave, gender, and a quadratic in age.

Table A.14: Demand for Commitment

	Bedtime target earlier than 1am	Sleep target longer than 7 hours	Dominated contract
At least 1 week	60%	60%	13%
All weeks	24%	19%	10%

Notes - The sample in columns 1-2 is restricted to subjects receiving monetary incentives in treatment weeks (N=207). The sample in column 3 is restricted to subjects receiving monetary incentives in treatment weeks in treatment 3 (Biweekly-Small) (N=32).

Table A.15: Present-Bias, Impatience, Overconfidence and Commitment

	(1) Before 1 am	(2) More than 7hrs	(3) Either
Panel A: Present Bias			
Present-Biased	0.1400*	0.0353	0.1665***
	(0.078)	(0.082)	(0.061)
Observations	207	207	207
Mean of Dep. Var.	0.595	0.590	0.745
Std.Dev. of Dep. Var.	0.492	0.493	0.437
Panel B: Impatience			
Impatient	0.1498*	0.0416	0.0423
	(0.079)	(0.083)	(0.072)
Observations	207	207	207
R-squared	0.016	0.001	0.002
Mean of Dep. Var.	0.595	0.590	0.745
Std.Dev. of Dep. Var.	0.492	0.493	0.437
Panel C: Overconfidence			
Overconfidence	-0.0470	0.0187	-0.0603
	(0.086)	(0.085)	(0.078)
Observations	207	207	207
Mean of Dep. Var.	0.595	0.590	0.745
Std.Dev. of Dep. Var.	0.492	0.493	0.437
Panel D: Risk Aversion			
Risk-averse	-0.1331	-0.0938	-0.1182
	(0.091)	(0.091)	(0.086)
Observations	207	207	207
Mean of Dep. Var.	0.595	0.590	0.745
Std.Dev. of Dep. Var.	0.492	0.493	0.437

Notes - Data are drawn from Fitbit data, weekly surveys collected during the weeks of the intervention, and the first-day survey. The sample is restricted to subjects in the treatment group.

Table A.16: Survey Time and Bedtime Target

VARIABLES	(1) Reduced-form	(2) Reduced-form	(3) 2SLS	(4) 2SLS
Early email	0.1223**	0.0708*		
(before noon)	(0.051)	(0.038)		
Response before			0.1727**	0.1081*
3pm			(0.071)	(0.057)
Observations	611	611	611	611
Individual FE		YES		YES

Notes - Data are drawn from the weeks of treatment. Standard errors clustered at the individual level are reported in parentheses. Analysis restricted to waves 1-4 due to data limitations.

Table A.17: Determinants of Bedtime Targets

	(1)	(2)
Variables	Bedtime Target	
Survey Time	0.0128*	
	(0.0065)	
Average Bedtime in Previous Week		-0.00875**
		(0.00339)
Observations	584	584
R-Squared	0.047	0.059
Mean of Dep. Var.	24.40	24.40
Std.Dev. of Dep. Var.	0.743	0.764

Notes - The table above shows regressions of survey time and average bedtime the previous week on bedtime target. Both regressions include controls for week and subject fixed effects. Standard errors clustered at the individual level are reported in parentheses. Column 1 includes observations with survey times after the first surveys were sent (6 am) and before midnight. The sample is restricted to treated subjects in the treatment weeks.

Table A.18: Predicting Achievement Rate

Prediction	In	Week 1	Week 2	Week 3
For	Week 1	2.83		
	Week 2	2.81	2.71	
	Week 3	2.91	2.71	2.60
Achievement		Week 1	Week 2	Week3
In	Week 1	1.81		
	Week 2	1.78	1.84	
	Week 3	1.47	1.50	1.46
Difference between Prediction and Achievement	In	Week 1	Week 2	Week 3
For	Week 1	1.02		
	Week 2	1.03	0.87	
	Week 3	1.44	1.21	1.14

Notes - This table provides averages for the number of nights subjects predict they will meet their target and the number of nights subjects actually meet their target.

Table A.19: Naps

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Nap		Nap duration		7<Sleep<9	
Treatment	-0.0122	-0.0110	-1.0161	-0.8449	0.0489***	0.0490***
	(0.008)	(0.009)	(0.706)	(0.746)	(0.017)	(0.017)
After treatment	-0.0088	-0.0049	-0.7387	-0.3508	0.0168	0.0169
	(0.011)	(0.012)	(0.916)	(1.004)	(0.023)	(0.023)
Nap					-0.098	
					0.020	
Nap duration						-0.0011***
						(0.000)
Individual fixed effects		YES		YES	YES	YES
Observations	8,738	8,738	8,738	8,738	8,738	8,738
Mean of Dep. Var.	0.0570	0.0570	4.638	4.638	0.456	0.456
Std.Dev. of Dep. Var.	0.232	0.232	19.73	19.73	0.498	0.498
Number of id		319		319	319	319

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2, 4, and 6 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table A.20: Baseline Characteristics and Sample Attrition in Time-Use Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Female	Age	White	Black	Asian	Other	Last month sleep
No follow up	0.045	-0.887***	-0.165***	0.102***	0.073	-0.010	-0.200*
	(0.053)	(0.326)	(0.053)	(0.030)	(0.045)	(0.034)	(0.105)
Observations	359	359	359	359	359	359	359
R-squared	0.002	0.020	0.027	0.031	0.007	0.000	0.010
Dep. Var.	Sleep during term	Sleep < 7hrs during term	Ever smoked	Ideal sleep hours	BMI	Overweight	Obese
No follow up	-0.062	0.095*	0.072	-0.168**	-0.527	0.017	0.033
	(0.144)	(0.054)	(0.045)	(0.085)	(1.047)	(0.047)	(0.026)
Observations	359	359	359	359	359	359	359
R-squared	0.001	0.009	0.007	0.011	0.001	0.000	0.005
0.000	0.003	0.001	0.002				

Notes - Data are drawn from the Day 1 Survey. Each column reports a univariate regression estimate of the dependent variable (baseline characteristics) on a dummy indicating whether the individual did not respond to the time-use survey.

Table A.21: Incentives to Sleep and Screen Time Near Bedtime

	(1) Screen time (hours) after 8 pm	(2) Any screen time after 8 pm
Night on which treatment was achieved	-0.224***	-0.125***
	(0.050)	(0.040)
Post-treatment (achieved 50% of nights)	-0.132*	-0.081*
	(0.077)	(0.048)
Observations	1,106	1,106
R-squared	0.039	0.030
Mean of Dep. Var.	0.452	0.372
Std.Dev. of Dep. Var.	0.793	0.483

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Standard errors are clustered at the individual level.

Table A.22: Incentives and Sleep Regularity

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Std.Dev. Sleep Hours		Std.Dev. Bedtime		Std.Dev. Wake up time	
Treatment	-0.1025*	-0.0625	-0.1353***	-0.0881**	-0.0895	-0.0473
	(0.053)	(0.040)	(0.050)	(0.045)	(0.069)	(0.059)
Post-treatment	-0.0254	0.0084	0.0237	0.0606	-0.1718*	-0.1575*
	(0.089)	(0.072)	(0.072)	(0.068)	(0.097)	(0.083)
Individual fixed effects		YES		YES		YES
Observations	7,690	7,690	7,690	7,690		7,690
R-squared	0.048	0.049	0.018	0.017	0.023	0.031
Mean of Dep. Var.	1.289	1.289	1.085	1.085	1.155	1.155
Std.Dev. of Dep. Var.	0.854	0.854	0.762	0.762	1.040	1.040

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2, 4, and 6 include individual fixed effects.Standard errors clutered at the individual level are reported in parentheses.

Table A.23: Incentives and Sleep: Timing of the Incentives

	(1)	(2)	(3)	(4)
	Sleep7≤Sleep≤9		Sleep<6 hours	
Weekly incentive	0.0932***	0.0374	-0.0560**	-0.0143
	(0.031)	(0.024)	(0.028)	(0.021)
Post-weekly incentive	0.0560	0.0085	-0.0718*	-0.0458*
	(0.043)	(0.028)	(0.037)	(0.024)
Bi-weekly incentive	0.0675**	0.0537*	-0.0884***	-0.0711***
	(0.034)	(0.031)	(0.028)	(0.025)
Post-biweekly incentive	0.0116	-0.0073	-0.0449	-0.0283
	(0.037)	(0.032)	(0.028)	(0.023)
Individual fixed effects		YES		YES
Observations	8,738	8,738	8,738	8,738
Mean of Dep. Var.	0.456	0.456	0.245	0.245
Std.Dev. of Dep. Var.	0.498	0.498	0.430	0.430

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table A.24: Incentives and Sleep: the Role of the Size of the Financial Incentive

	(1)	(2)	(3)	(4)
	Sleep7≤Sleep≤9		Sleep<6 hours	
Strong Treatmet	0.0845***	0.0495***	-0.0607***	-0.0351**
	(0.024)	(0.018)	(0.021)	(0.016)
Post Strong Treatment	0.0507	0.0175	-0.0609*	-0.0477**
	(0.036)	(0.024)	(0.031)	(0.021)
Weak Treatment	0.0481	0.0214	-0.0473	-0.0173
	(0.052)	(0.040)	(0.039)	(0.032)
Post Weak Treatment	-0.0066	0.0009	-0.0402	-0.0492
	(0.078)	(0.063)	(0.057)	(0.045)
Individual fixed effects		YES		YES
Observations	8,738	8,738	8,738	8,738
R-squared	0.014	0.005	0.015	0.007
Mean of Dep. Var.	0.456	0.456	0.245	0.245
Std.Dev. of Dep. Var.	0.498	0.498	0.430	0.430
Number of id		319		319

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table A.25: Incentives and Sleep

VARIABLES	(1) Sleep7<Sleep<9	(2)	(3) Sleep<6 hours	(4)
Any Incentive	0.0792***	0.0451***	-0.0588***	-0.0325**
	(0.022)	(0.017)	(0.020)	(0.015)
Post-Treatment (any incentive)	0.0483	0.0147	-0.0600**	-0.0469**
	(0.035)	(0.023)	(0.030)	(0.020)
Individual fixed effects		YES		YES
Observations	8,738	8,738	8,738	8,738
Mean of Dep. Var.	0.456	0.456	0.245	0.245
Std.Dev. of Dep. Var.	0.498	0.498	0.430	0.430
Number of id		319		319

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

B Elicitation of Risk and Time Preferences

We used choice lists to elicit participants' risk and time preferences. The subjects could choose from two columns, representing Option A and Option B. On each list, one of the two options was fixed, and the other option changed from one row to the next. In each row, subjects had to indicate their preferred option: Option A or Option B. To avoid multiple switching points on a single list, the subjects only had to choose in which row they wanted to switch from choosing Option A to choosing Option B. The subjects were given examples and the opportunity to practice before making decisions that counted for payment. When payments involved a future date, the subjects would receive the corresponding amount via email in the form of a gift card.

To elicit the risk preference parameter, we used two lists. On each list, Option A was a fixed lottery: a 50% chance of getting GBP 6 and a 50% chance of getting GBP 0. Option B was always a sure amount. The lists we used are illustrated in Figures B.1 and B.2.

To elicit the time preference parameters, we used four lists. On each list, Option A was associated with a monetary payment at a sooner time and Option B implied some monetary payment at a later time. The amount to be gained at the later time is fixed at GBP 6, and the amount to be gained at the sooner time varied on each list. Among the lists, the sooner time is either today or in 4 weeks, and the delay between the later and the sooner time is either 4 weeks or 8 weeks. The lists we used are illustrated in Figures B.3, B.4, B.5 and B.6.

Figure B.1: Choice List for Risk Preference 1

Option A	Option B
50% Chance of £6 and 50% Chance of £0	£0.00
50% Chance of £6 and 50% Chance of £0	£0.30
50% Chance of £6 and 50% Chance of £0	£0.60
50% Chance of £6 and 50% Chance of £0	£0.90
50% Chance of £6 and 50% Chance of £0	£1.20
50% Chance of £6 and 50% Chance of £0	£1.50
50% Chance of £6 and 50% Chance of £0	£1.80
50% Chance of £6 and 50% Chance of £0	£2.10
50% Chance of £6 and 50% Chance of £0	£2.40
50% Chance of £6 and 50% Chance of £0	£2.70
50% Chance of £6 and 50% Chance of £0	£3.00
50% Chance of £6 and 50% Chance of £0	£3.30
50% Chance of £6 and 50% Chance of £0	£3.60
50% Chance of £6 and 50% Chance of £0	£3.90
50% Chance of £6 and 50% Chance of £0	£4.20
50% Chance of £6 and 50% Chance of £0	£4.50
50% Chance of £6 and 50% Chance of £0	£4.80
50% Chance of £6 and 50% Chance of £0	£5.10
50% Chance of £6 and 50% Chance of £0	£5.40
50% Chance of £6 and 50% Chance of £0	£5.70
50% Chance of £6 and 50% Chance of £0	£6.00

Figure B.2: Choice List for Risk Preference 2

Option A	Option B
50% Chance of £6 and 50% Chance of £0	£0.00
50% Chance of £6 and 50% Chance of £0	£0.30
50% Chance of £6 and 50% Chance of £0	£0.60
50% Chance of £6 and 50% Chance of £0	£0.90
50% Chance of £6 and 50% Chance of £0	£1.20
50% Chance of £6 and 50% Chance of £0	£1.50
50% Chance of £6 and 50% Chance of £0	£1.80
50% Chance of £6 and 50% Chance of £0	£2.10
50% Chance of £6 and 50% Chance of £0	£2.40
50% Chance of £6 and 50% Chance of £0	£2.70
50% Chance of £6 and 50% Chance of £0	£3.00
50% Chance of £6 and 50% Chance of £0	£3.30
50% Chance of £6 and 50% Chance of £0	£3.60
50% Chance of £6 and 50% Chance of £0	£3.90
50% Chance of £6 and 50% Chance of £0	£4.20
50% Chance of £6 and 50% Chance of £0	£4.50
50% Chance of £6 and 50% Chance of £0	£4.80
50% Chance of £6 and 50% Chance of £0	£5.10
50% Chance of £6 and 50% Chance of £0	£5.40
50% Chance of £6 and 50% Chance of £0	£5.70
50% Chance of £6 and 50% Chance of £0	£6.00

Figure B.3: Choice List for Time Preference 1

Option A	Option B
Receive £5.80 today	Receive £6 in 4 weeks
Receive £5.60 today	Receive £6 in 4 weeks
Receive £5.40 today	Receive £6 in 4 weeks
Receive £5.20 today	Receive £6 in 4 weeks
Receive £5.00 today	Receive £6 in 4 weeks
Receive £4.80 today	Receive £6 in 4 weeks
Receive £4.60 today	Receive £6 in 4 weeks
Receive £4.40 today	Receive £6 in 4 weeks
Receive £4.20 today	Receive £6 in 4 weeks
Receive £4.00 today	Receive £6 in 4 weeks
Receive £3.80 today	Receive £6 in 4 weeks
Receive £3.60 today	Receive £6 in 4 weeks
Receive £3.40 today	Receive £6 in 4 weeks
Receive £3.20 today	Receive £6 in 4 weeks
Receive £3.00 today	Receive £6 in 4 weeks
Receive £2.80 today	Receive £6 in 4 weeks
Receive £2.60 today	Receive £6 in 4 weeks
Receive £2.40 today	Receive £6 in 4 weeks
Receive £2.20 today	Receive £6 in 4 weeks
Receive £2.00 today	Receive £6 in 4 weeks
Receive £1.80 today	Receive £6 in 4 weeks
Receive £1.60 today	Receive £6 in 4 weeks
Receive £1.40 today	Receive £6 in 4 weeks
Receive £1.20 today	Receive £6 in 4 weeks
Receive £1.00 today	Receive £6 in 4 weeks
Receive £0.80 today	Receive £6 in 4 weeks
Receive £0.60 today	Receive £6 in 4 weeks
Receive £0.40 today	Receive £6 in 4 weeks
Receive £0.20 today	Receive £6 in 4 weeks

Figure B.4: Choice List for Time Preference 2

Option A	Option B
Receive £5.80 today	Receive £6 in 8 weeks
Receive £5.60 today	Receive £6 in 8 weeks
Receive £5.40 today	Receive £6 in 8 weeks
Receive £5.20 today	Receive £6 in 8 weeks
Receive £5.00 today	Receive £6 in 8 weeks
Receive £4.80 today	Receive £6 in 8 weeks
Receive £4.60 today	Receive £6 in 8 weeks
Receive £4.40 today	Receive £6 in 8 weeks
Receive £4.20 today	Receive £6 in 8 weeks
Receive £4.00 today	Receive £6 in 8 weeks
Receive £3.80 today	Receive £6 in 8 weeks
Receive £3.60 today	Receive £6 in 8 weeks
Receive £3.40 today	Receive £6 in 8 weeks
Receive £3.20 today	Receive £6 in 8 weeks
Receive £3.00 today	Receive £6 in 8 weeks
Receive £2.80 today	Receive £6 in 8 weeks
Receive £2.60 today	Receive £6 in 8 weeks
Receive £2.40 today	Receive £6 in 8 weeks
Receive £2.20 today	Receive £6 in 8 weeks
Receive £2.00 today	Receive £6 in 8 weeks
Receive £1.80 today	Receive £6 in 8 weeks
Receive £1.60 today	Receive £6 in 8 weeks
Receive £1.40 today	Receive £6 in 8 weeks
Receive £1.20 today	Receive £6 in 8 weeks
Receive £1.00 today	Receive £6 in 8 weeks
Receive £0.80 today	Receive £6 in 8 weeks
Receive £0.60 today	Receive £6 in 8 weeks
Receive £0.40 today	Receive £6 in 8 weeks
Receive £0.20 today	Receive £6 in 8 weeks

Figure B.5: Choice List for Time Preference 3

Option A	Option B
Receive £5.80 in 4 weeks	Receive £6 in 8 weeks
Receive £5.60 in 4 weeks	Receive £6 in 8 weeks
Receive £5.40 in 4 weeks	Receive £6 in 8 weeks
Receive £5.20 in 4 weeks	Receive £6 in 8 weeks
Receive £5.00 in 4 weeks	Receive £6 in 8 weeks
Receive £4.80 in 4 weeks	Receive £6 in 8 weeks
Receive £4.60 in 4 weeks	Receive £6 in 8 weeks
Receive £4.40 in 4 weeks	Receive £6 in 8 weeks
Receive £4.20 in 4 weeks	Receive £6 in 8 weeks
Receive £4.00 in 4 weeks	Receive £6 in 8 weeks
Receive £3.80 in 4 weeks	Receive £6 in 8 weeks
Receive £3.60 in 4 weeks	Receive £6 in 8 weeks
Receive £3.40 in 4 weeks	Receive £6 in 8 weeks
Receive £3.20 in 4 weeks	Receive £6 in 8 weeks
Receive £3.00 in 4 weeks	Receive £6 in 8 weeks
Receive £2.80 in 4 weeks	Receive £6 in 8 weeks
Receive £2.60 in 4 weeks	Receive £6 in 8 weeks
Receive £2.40 in 4 weeks	Receive £6 in 8 weeks
Receive £2.20 in 4 weeks	Receive £6 in 8 weeks
Receive £2.00 in 4 weeks	Receive £6 in 8 weeks
Receive £1.80 in 4 weeks	Receive £6 in 8 weeks
Receive £1.60 in 4 weeks	Receive £6 in 8 weeks
Receive £1.40 in 4 weeks	Receive £6 in 8 weeks
Receive £1.20 in 4 weeks	Receive £6 in 8 weeks
Receive £1.00 in 4 weeks	Receive £6 in 8 weeks
Receive £0.80 in 4 weeks	Receive £6 in 8 weeks
Receive £0.60 in 4 weeks	Receive £6 in 8 weeks
Receive £0.40 in 4 weeks	Receive £6 in 8 weeks
Receive £0.20 in 4 weeks	Receive £6 in 8 weeks

Figure B.6: Choice List for Time Preference 4

Option A	Option B
Receive £5.80 in 4 weeks	Receive £6 in 12 weeks
Receive £5.60 in 4 weeks	Receive £6 in 12 weeks
Receive £5.40 in 4 weeks	Receive £6 in 12 weeks
Receive £5.20 in 4 weeks	Receive £6 in 12 weeks
Receive £5.00 in 4 weeks	Receive £6 in 12 weeks
Receive £4.80 in 4 weeks	Receive £6 in 12 weeks
Receive £4.60 in 4 weeks	Receive £6 in 12 weeks
Receive £4.40 in 4 weeks	Receive £6 in 12 weeks
Receive £4.20 in 4 weeks	Receive £6 in 12 weeks
Receive £4.00 in 4 weeks	Receive £6 in 12 weeks
Receive £3.80 in 4 weeks	Receive £6 in 12 weeks
Receive £3.60 in 4 weeks	Receive £6 in 12 weeks
Receive £3.40 in 4 weeks	Receive £6 in 12 weeks
Receive £3.20 in 4 weeks	Receive £6 in 12 weeks
Receive £3.00 in 4 weeks	Receive £6 in 12 weeks
Receive £2.80 in 4 weeks	Receive £6 in 12 weeks
Receive £2.60 in 4 weeks	Receive £6 in 12 weeks
Receive £2.40 in 4 weeks	Receive £6 in 12 weeks
Receive £2.20 in 4 weeks	Receive £6 in 12 weeks
Receive £2.00 in 4 weeks	Receive £6 in 12 weeks
Receive £1.80 in 4 weeks	Receive £6 in 12 weeks
Receive £1.60 in 4 weeks	Receive £6 in 12 weeks
Receive £1.40 in 4 weeks	Receive £6 in 12 weeks
Receive £1.20 in 4 weeks	Receive £6 in 12 weeks
Receive £1.00 in 4 weeks	Receive £6 in 12 weeks
Receive £0.80 in 4 weeks	Receive £6 in 12 weeks
Receive £0.60 in 4 weeks	Receive £6 in 12 weeks
Receive £0.40 in 4 weeks	Receive £6 in 12 weeks
Receive £0.20 in 4 weeks	Receive £6 in 12 weeks