Why Do We Procrastinate? Present Bias and Optimism

Zachary Breig, Matthew Gibson, and Jeffrey G. Shrader*

December 13, 2019

Abstract

Research has shown that procrastination has significant adverse effects on individuals, including lower savings and poorer health. Procrastination is typically modeled as resulting from present bias. In this paper we study an alternative: excessively optimistic beliefs about future demands on an individual's time. The models can be distinguished by how individuals respond to information on their past choices. Experimental results refute the hypothesis that present bias is the sole source of dynamic inconsistency, but they are consistent with optimism. The findings offer an explanation for low takeup of commitment and suggest that personalized information on past choices can mitigate procrastination. (JEL: D90,D84,D15,J22)

*Affiliations. Breig: School of Economics, The University of Queensland, Level 6 Colin Clark Building (39), St Lucia, Brisbane, Australia, 4072 (e-mail: z.breig@uq.edu.au); Gibson: Department of Economics, Williams College, 24 Hopkins Hall Drive, Williamstown MA 01267 (e-mail: mg17@williams.edu); Shrader: School of International and Public Affairs, Columbia University, 420 West 118th Street, New York NY 10027 (e-mail: jgs2103@columbia.edu). We are grateful for helpful comments from James Andreoni, Gordon Dahl, Marco Faravelli, Osea Giuntella, Jeff Kline, Bentley MacLeod, Frank Schilbach, Charles Sprenger, and Jack Willis as well as seminar participants at ASHEcon, the North American ESA Meeting, UCSD, University of Melbourne, and University of Queensland. George Kassabian and Anmolbir Mann provided excellent research assistance. This study was approved by the IRB at the University of California, San Diego.

1 Introduction

Procrastination is an important feature of everyday life. It is a common topic of conversation at work and at home, and economists have documented it in consequential settings including retirement saving, exercise, and education (Thaler and Benartzi, 2004, DellaVigna and Malmendier, 2006, Ariely and Wertenbroch, 2002). Such procrastination is commonly modeled as originating from present bias: discounting that favors the present at the expense of the future (Strotz, 1955, Laibson, 1997, O'Donoghue and Rabin, 1999, Barro, 1999, Ashraf et al., 2006, Augenblick et al., 2015).[1] We study an alternative model in which dynamic inconsistency arises from excessive optimism about future demands on an individual's time. While both models predict dynamically inconsistent choices, they predict different responses to information about past procrastination. We test these predictions experimentally and reject the hypothesis that present bias is the sole source of dynamic inconsistency. Instead we find evidence that both discount rates and beliefs matter. Our results suggest that the typical policy prescription—offering people the chance to tie themselves to the mast, committing to decisions in advance—is incomplete, and that personalized historical information is an important additional tool for people making decisions over time.

Biased beliefs about future time shocks can cause choices made ahead of time to differ from choices made in the moment. Consider an agent who does not accurately anticipate the arrival of a time-consuming task. Colloquially we say that such an agent is optimistic about her time shocks. Once the task arrives, the agent will need to defer planned time use to accommodate the unanticipated shock. If the agent has systematically biased beliefs over future time shocks, then such procrastination can occur even with neoclassical discounting. We refer to this as *belief-based dynamic inconsistency*.[2]

Discounting-based dynamic inconsistency, in contrast, models dynamically inconsistent choices as originating from a utility function that places lower weight on the more distant future relative to the present or the immediate future. This leads

[1] In the quasi-hyperbolic model of Laibson (1997), the agent discounts at rate δ between future periods, but between the current period and the next period at rate $\beta\delta$ with $\beta < 1$. This heavier discounting leads to "present biased" allocative choices.

[2] In contrast to Halevy (2008) and Andreoni and Sprenger (2012b), this inconsistency is a result of the decision maker having incorrect beliefs rather than a utility function which does not take the expected utility form.

the agent to exhibit present-biased dynamic inconsistency because choices made far enough in advance will be governed by geometric discounting, while choices made about the immediate future will not. If an agent is naïve about her own present-biased discounting, she believes that she will behave more consistently than she actually does.[3]

Because these two models lead to similar dynamically inconsistent choices, a research design seeking to distinguish discounting-based from belief-based dynamic inconsistency cannot rely solely on revealed procrastination.[4] Providing agents with information on past time-inconsistent decisions resolves this difficulty. Discounting- and belief-based models make different predictions about how agents will respond to information about their own past procrastination.

First, the two models give different predictions for how effort allocation will change in response to information. Discounting-based dynamically inconsistent agents have a clear idea of the time shocks that they face, but have trouble committing to time use choices. Such agents will not change effort allocation decisions in response to information. In contrast, belief-based dynamically inconsistent agents have erroneous expectations about time shocks. Correcting these beliefs will cause them to change their effort allocation decisions to better conform to the true state of the world.

Second, information can cause naïve present-biased agents to learn about their own present bias. For instance, an agent might learn that her discounting is more present-biased than she previously thought.[5] This will increase commitment demand for time-use choices made far in advance. If agents have biased beliefs over time shocks, however, this prediction need not hold. Information on past dynamically inconsistent decisions should help optimistic agents bring their beliefs in line with the true state, but this does not necessarily lead them to demand costly commitment (Laibson, 2015).

We tested these predictions in an experiment over two weeks. The first week provided a baseline measure of dynamically inconsistent behavior for each subject. On the first morning of the experiment, subjects divided their required tasks between

---

[3] Again, in the quasi-hyperbolic model of Laibson (1997), naïve agents have true discounting parameters β and δ but believe their present-bias parameter is $\hat{\beta}$ where $\beta < \hat{\beta} \leq 1$.

[4] A recent working paper by Browning and Tobacman (2015) makes a similar theoretical argument. Gabaix and Laibson (2017) show that a similar identification problem can occur due to imperfect (but unbiased) forecasting of the future.

[5] More formally, she has a lower β than she previously thought.

a period later that day and a period two days in the future. Subjects were able to pay a price in terms of additional tasks to commit to this morning choice.[6] On the evening of day 1, subjects could revise their task allocation, conditional on the morning commitment decision. Procrastination was indicated by the subject moving tasks to the later date. In addition, we gathered information on routine time use and subjects' predicted and actual bedtimes so we could compare task allocation behavior to real-world decisions. At the beginning of week 2, treated subjects were presented with information on their own task procrastination. They were also given information on how well they were able to forecast their own bedtimes, a decision where subjects also routinely exhibit procrastination. To reduce the probability of experimenter demand effects, this information was cast as a neutral reporting of past behavior. All subjects then engaged in the same task decisions as in week 1.

Our experimental results indicate that both beliefs and discounting are important determinants of time inconsistency. Evidence on beliefs comes from testing the effect of treatment on task allocation. Among subjects who deferred work in week 1, the treatment caused a practically large and statistically significant reduction in deferral of tasks in week 2. This is consistent with belief-based dynamic inconsistency and inconsistent with discounting-based dynamic inconsistency. Evidence on present-biased utility comes from testing the effect of treatment on commitment demand in week 2 for individuals who were procrastinators in week 1. For these individuals, treatment increased week 2 commitment demand, consistent with discounting-based dynamic inconsistency. Heterogeneity analysis using baseline measures of naïvete about time preferences and time shocks shows that these results are stronger for naifs.

Next we assess the prevalence of belief- and discounting-based dynamic inconsistency. Each treatment-group subject who reallocated tasks in week 1 is matched with a control-group peer. If such a subject reduced the number of tasks deferred in week 2 more than her control group peer, we classify her behavior as belief-based. If such a subject increased commitment demand more than her control group peer, we classify her behavior as discounting-based. Under this taxonomy, 25% of subjects exhibited discounting-based inconsistency, 39% exhibited belief-based inconsistency, and 21% exhibited both. The remaining 15% exhibited behavior inconsistent with either model.

[6] This in-kind price could take on both positive and negative values.

We then show that subjects' dynamic inconsistency extended to consequential, real-world time use. At baseline, subjects systematically mis-predicted their own bedtimes, going to bed later than planned on average. Treated subjects reduced their forecast error in week two. Consistent with the task-based results, this reduction was larger for those with larger week-one forecast errors. This is evidence that the treatment affected subjects' decision problem in the time domain. To investigate further, we collected a panel of time use data from subjects over the course of the experiment. Panel data on time use is rare (Frazis and Stewart, 2012), and it allows us to examine how our treatment affected behavior outside the experiment. We show that when subjects were randomly induced to spend more time on our experiment, on average they spent less time studying and working, but more time watching television.

This study provides evidence on the sources of dynamically inconsistent behavior and the real-world consequences of such behavior. We observe procrastination in a controlled setting and tie that behavior to consequential decisions like the timing of sleep (Gibson and Shrader, 2018). Firms may be exploiting time inconsistency in sleep decisions and elsewhere. For instance, Netflix CEO Reed Hastings has argued that tempting services like streaming video may affect sleep decisions, saying in an earnings call that "We're competing with sleep, on the margin." If individuals want to avoid such lures, the appropriate action depends on the source of their time inconsistency. The policy prescription from the time inconsistency literature has primarily been to encourage commitment by sophisticated present-biased agents. Workers are often urged to contribute to retirement plans with early withdrawal penalties or commit to smoking cessation through a website like stickK.com. Our results suggest that this policy prescription is incomplete. If some procrastination instead stems from overestimation of future earnings or underestimation of how difficult it will be to quit smoking, then organizations and individuals seeking to correct dynamic inconsistency should provide personalized information as well. This hypothesis is consistent with the widespread sale of goods—like fitness trackers and planners—that help consumers reflect on execution of their own plans.[7] Models of long-run forecasting errors due to bounded rationality (Gabaix, 2014) similarly suggest that targeted information may correct consistent planning errors.

In addition, our study makes two contributions related to the demand for costly

[7]Paul Krugman has made this point when reflecting on his own fitness tracker use, writing that "what fitness devices do, at least for me, is make it harder to lie to myself" (Krugman, 2015).

commitment. First, our findings help explain the widely observed, low take-up of such commitment. Subjects whose dynamic inconsistency originates solely from optimism will not demand costly commitment. Schilbach (2019) observes that in the majority of past experiments, subjects were either unwilling to pay for commitment or were willing to pay only very small amounts.[8] In significant exceptions, Schilbach does find high demand for commitment in the domain of alcohol consumption, and Casaburi and Macchiavello (2019) find demand for costly commitment in the Kenyan dairy sector. Second, our experimental design makes a methodological contribution in its elicitation of commitment demand. Our design begins from the convex time budget techniques of Andreoni and Sprenger (2012a).[9] Specifically, it uses real-effort tasks similar to those employed by Augenblick et al. (2015) and hews closely to the overall experimental design of that paper to clarify the importance of belief-based dynamic inconsistency. In contrast to Augenblick et al. (2015), our commitment price is denominated in tasks rather than money.[10] By keeping all choices in the task domain, we reduce the tendency of commitment demand to spike sharply at a zero price. We find that 20% of subjects were willing to commit to their time use choices at positive task-denominated prices.

Finally, our results contribute to a growing body of research demonstrating the importance of a decision maker's beliefs in how they make choices involving time. DellaVigna and Malmendier (2006) and Acland and Levy (2015) both study gym membership and attendance, showing that consumers systematically overestimate how often they will go to the gym in the future even when this choice entails monetary costs. Avery et al. (2019) show that students routinely sleep less than the medically recommended amount due to both impatience and over-confidence. Börsch-Supan et al. (2018) demonstrate that a much larger portion of regret about not having saved more earlier in life is explained by positive and negative financial shocks than present bias. Consistent with the common lack of commitment demand in experimental

[8]While commitment demand is typically low, Carrera et al. (2019) show that misperception of contracts can lead an agent to demand too much commitment. In their context, the authors find that information on own and peers' past choices reduces commitment demand.

[9]The approach here differs from previous convex time budget experiments in that it does not vary the rate at which subjects trade off between present and future consumption. This simplifies the experiment and its instructions but does so at the cost of not being able to estimate discounting parameters. For an overview of designs used to estimate time preferences, see Frederick et al. (2002).

[10]To the best of our knowledge Toussaert (2018) is the only other experiment that elicits commitment demand with prices denominated in tasks.

subjects, Augenblick and Rabin (2018) find that individuals' *predictions* about the choices they will make in the future suggest that they do not understand their own present bias. Furthermore, subjects who make choices for the future immediately after completing tasks volunteer for less work in the future than those asked just before completing tasks. While the authors interpret this as evidence of projection bias, it is also consistent with decision makers who are optimistic about their desire to complete future tasks, but who update after getting information.

The paper proceeds as follows. Section 2 presents the models of time inconsistency due to discounting and beliefs, and it lays out the model predictions. Section 3 gives the experimental design. Section 4 describes the data. Section 5 presents tests of our theoretical hypotheses. Section 6 links our experimental and theoretical results to real-world behavior. Section 7 concludes.

2 Theory and Hypotheses

This section provides a theoretical model that yields the hypotheses we test in the experiment. We begin by discussing more formally the discounting- and belief-based sources of dynamic inconsistency that motivate our study. We then create a general framework that embeds the possibility of dynamic consistency, belief-based dynamic inconsistency, and discounting-based dynamic inconsistency. Both sources of dynamic inconsistency lead to observed procrastination. We then show that decision makers with these sources of dynamic inconsistency may react differently to information about their previous decisions.

2.1 Sources of Dynamic Inconsistency

A popular way to model present bias is through the use of β-δ preferences, in which δ captures the standard "exponential" part of discounting, while β is the "present bias" parameter, which places a lower weight on all future sources of utility (Laibson, 1997). The β-δ model is used in part because it generates the dynamic inconsistency that is often seen in choice data. In a two-period model, the inconsistency arises from the difference in how the decision maker trades off utility coming from periods 1 and 2 when the decision is made at or before period 1. In the former case, the rate of discount between the two periods is $\beta\delta$, while in the latter it is δ.

In models where decision makers are aware of their present bias, they have an incentive to seek out commitment. While previous work has found evidence of present bias, evidence for commitment demand has been more elusive. One explanation is that individuals are unaware of their present bias, or are "naïve" in the sense that they believe that their β parameter is closer to 1 than it actually is.

While choice revisions can arise from present biased discounting, they can also arise from biased beliefs over time shocks. Consider a decision maker solving an effort allocation problem over periods 1 and 2. She may do so under the belief that the distribution of period-one shocks, $F(\theta)$, is more favorable than it truly is. In this case, the time shocks in period 1 will tend to be surprisingly high, and the decision maker will want to complete fewer tasks than originally planned.

We do not model the source of incorrect beliefs, instead taking them as given and studying their implications. However, a number of existing models could lead to these optimistic beliefs. Kahneman and Tversky (1982) coined the term "planning fallacy," and provided an intuitive model in which decision makers neglect distributional information, leading to optimistic beliefs about outcomes like task duration or earnings. Beliefs and updating rules have also been modeled as a choice variable from the point of view of the decision maker (Bénabou and Tirole, 2002, Brunnermeier and Parker, 2005, Brunnermeier et al., 2016). Agents in these models trade off between the distortions caused by incorrect beliefs and their benefits, such as improved self-esteem or higher motivation.[11] Although these different models are important for the welfare implications of an intervention to reduce bias in beliefs, they are not important for the purpose of this study—to assess the roles of beliefs and discounting in dynamically inconsistent behavior.

There is a great deal of empirical evidence suggesting that decision makers exhibit optimistic beliefs about the future. Roy et al. (2005) survey the literature on the planning fallacy, in which individuals underestimate the length of time it will take to complete tasks. This optimism is found in situations as diverse as predicting the amount of time it takes to fill out tax forms and predicting the amount of time one will have to wait in line for gas. Similarly, people have a tendency to think they are more likely than their peers to experience positive events and less likely to experience negative events (Taylor and Brown, 1988). Overconfidence is an especially consequential form of optimism. Research has shown that individuals are system-

[11]In accord with these theoretical results, there is evidence that subjects do not update their beliefs according to Bayes' rule (Falk et al., 2006, Eil and Rao, 2011).

atically overconfident, with the vast majority thinking they are smarter (Larwood and Whittaker, 1977) or better drivers (Svenson, 1981) than their average peer. This overconfidence continues to be experimentally observable even when it is costly to the decision makers (Camerer and Lovallo, 1999, Niederle and Vesterlund, 2007). It also manifests in decisions outside of the lab, such as changing patterns in corporate investment and mergers (Malmendier and Tate, 2005, 2008) and job search behavior (Mueller et al., 2018). These sorts of excess optimism and overconfidence might lead decision makers to start a task later than they expected, or fail to save sufficiently, in a way that makes them appear present biased.

2.2 Model

We study a decision maker whom we observe over two weeks. Each week, the decision maker is given tasks and decides how to allocate them between two predetermined days. When the tasks are to be completed, a time shock is realized (imagine that this is a problem set to finish or unexpected car trouble). The decision maker suffers costs that are quadratic in the sum of the time shock and the number of tasks to be completed.

In week t, the decision maker has w tasks to to complete, which she is allowed to split between two days as w_t and $w - w_t$. The time shocks that the decision maker faces on the first and second days are denoted $\theta_{t,1}$ and $\theta_{t,2}$ respectively, both of which are weakly positive. For $i \in \{1, 2\}$, $\theta_{t,i}$ has a distribution $F(\theta_{t,i}; \alpha_i)$, where α_i is an index on the distribution such that if $\alpha > \alpha'$, then $F(\theta_{t,i}; \alpha)$ first order stochastically dominates $F(\theta_{t,i}; \alpha')$. These shocks are independent with the same mean and have distributions that are symmetric.

The decision maker may be present biased with $\beta \leq 1$. For simplicity, she has a long-term discount rate of $\delta = 1$. When the decision maker is offered the chance to split up the tasks well in advance of the tasks being completed, we refer to this choice as being "committed" (denoted by subscript C). If instead the choice is being made just before the tasks are completed, we refer to the choice as being "not committed" (denoted by subscript NC).

When deciding how to split up these tasks before the tasks must be completed, the decision maker solves

$$\min_{w_{t,C}} \; \beta \mathbb{E}\left[(w_{t,C} + \theta_{t,1})^2 + \delta(w - w_{t,C} + \theta_{t,2})^2\right], \tag{1}$$

which given our assumptions on shocks and $\delta = 1$ has solution

$$w^*_{t,C} = \frac{1}{2}w \tag{2}$$

When given the same decision after the first time shock is observed and when the work has to be done, the decision maker instead chooses a workload to solve

$$\min_{w_{t,NC}} \; (w_{t,NC} + \theta_{t,1})^2 + \beta\delta \mathbb{E}\left[(w - w_{t,NC} + \theta_2)^2\right], \tag{3}$$

which has the solution

$$w^*_{t,NC}(\theta_{t,1}) = \min\left\{\max\left\{0, \frac{\beta}{1+\beta}w + \frac{\beta}{1+\beta}\mathbb{E}[\theta_{t,2}] - \frac{1}{1+\beta}\theta_{t,1}\right\}, w\right\} \tag{4}$$

Previous research has demonstrated that individuals often make choices that appear present biased: they allocate more work to the earlier date when the decision is made for the future compared to when it is made in the present. We refer to the number of tasks that are deferred in period t as

$$D_t(\theta_{t,1}) = w_{t,C} - w_{t,NC}(\theta_{t,1})$$

Both belief-based and discounting-based dynamic inconsistency are consistent with behavior that looks like procrastination: planning to do work, then putting it off to a later date when given the chance. With belief-based dynamic inconsistency, this is a result of the average value of $\theta_{t,1}$ being higher than the decision maker's perception of the expected value of $\theta_{t,1}$. With discounting-based dynamic inconsistency, this is a result of β being less than one. An agent exhibiting either or both forms of dynamic inconsistency will procrastinate. Formally, for such an agent $\mathbb{E}[D_t] > 0$. An agent with $\beta = 1$ and unbiased beliefs over time shocks will not procrastinate ($\mathbb{E}[D_t] = 0$).

In what follows, we discuss both belief-based and discounting-based dynamically inconsistent decision makers. When we refer to a belief-based dynamically inconsistent decision maker, specifically we mean a decision maker in the above model who has $\beta = 1$ but an overly optimistic belief distribution ($\hat{\alpha} < \alpha$). Alternatively, when we refer to a discounting-based dynamically inconsistent agent, specifically we mean a decision maker who has $\beta < 1$ and is naïve about this present bias, but has correct beliefs about time shocks ($\hat{\alpha} = \alpha$).

2.3 Information Provision and Testable Hypotheses

Models of biased beliefs and present bias generate the same predictions for effort allocation behavior. But because both models rely on beliefs (either over α or β) being incorrect to generate patterns seen in the data, one way to differentiate between the two models is to observe the effects of information provision on subsequent choice. Suppose the decision maker is given the information that when making past choices, the amount of work she agreed to complete earlier in the day was higher than the amount of work she chose when the tasks actually had to be completed. This could cause the decision maker to update to a higher belief about α or a lower belief about β. The predicted responses of work allocations and commitment to these two types of updating differ.

In the experimental design that follows (see Section 3), we treat some subjects with information about their week one choices before they make choices in week two. We use a tilde to refer to choices made by a decision maker who received this information treatment. Thus, $\tilde{D}_2(\theta_{2,1})$ refers to the number of tasks deferred by a decision maker in the second week after she has received a reminder of her previous choices, while $D_2(\theta_{2,1})$ is the same task deferral, but by someone who did not receive the information treatment.

In the naïve β-δ model, the decision maker has incorrect beliefs about the present bias parameter that will govern her decisions in the future. Because she cares about what her future self will choose, a change in these beliefs might lead to changes in her willingness to commit to her present actions (O'Donoghue and Rabin, 1999). In prior work, researchers have typically interpreted a lack of commitment demand by individuals as evidence that $\hat{\beta}$, the individual's belief about her own present bias, is closer to one than to β, the "true" present bias.

In the experiment, we observe the information subjects are given and the choices they make. However, we do not observe beliefs or how they are updated. Therefore we must impose assumptions about how decision makers of each type update their beliefs when provided with information.

Assumption DB. For discounting-based dynamically inconsistent decision makers, high observations of D_1 (more tasks deferred) cause the decision maker to believe she is more present biased ($\hat{\beta}$ farther from 1) than low observations.

Assumption BB. For belief-based dynamically inconsistent decision makers, high observations of D_1 (more tasks deferred) cause the decision maker to believe shocks on the first day are worse ($\hat{\alpha}$ is higher) than low observations.[12]

When applied to the decision problem described above, the $\hat{\beta}$ of a discounting-based dynamically inconsistent decision maker only affects her willingness to commit—not the allocative choices that she makes when either committed or uncommitted. Thus, the information treatment should not have any effect on procrastination.

Hypothesis DB1. Information provision will have no effect on work allocations for discounting-based dynamically inconsistent agents. Formally, $\tilde{D}_2(\theta_{2,1}) - D_2(\theta_{2,1}) \perp D_1(\theta_{1,1})$.

While β affects the solution to the problem in equation (3), beliefs about β do not. Intuitively, changing beliefs about β affects what the decision-maker believes she will do in the future. However, these beliefs do not affect the decision maker's incentives in the present.

While changing beliefs about one's present bias will not affect work allocations, changing beliefs about the distribution of time shocks will.

Hypothesis BB1. Information provision will decrease procrastination for belief-based dynamically inconsistent agents. Formally, $\mathbb{E}[\tilde{D}_2(\theta_{2,1}) - D_2(\theta_{2,1})]$ is decreasing in $D_1(\theta_{1,1})$.

A decision maker whose beliefs become more pessimistic modifies her choices so that her earlier decisions are more consistent with the decisions she makes later.

Our model of discounting-based dynamic inconsistency also has implications for how commitment demand changes after receiving information. Suppose this decision maker is choosing between the payoffs from equations (3) and (1). To match the

[12] Here we assume that information causes the decision maker to update about *only* the distribution of the first state. This is a reasonable simplification, because the information that the decision maker receives is generated by choices that were made before the value of the second state was observed. However, similar results hold if the decision maker updates beliefs about the distributions of both states, but shifts beliefs about the first state more.

experimental design given in Section 3, we model commitment demand as the number of extra tasks a subject is willing to complete in order to have her committed decision implemented. Mathematically, this is the k such that

$$\begin{aligned}&\mathbb{E}\left[(w_{t,C}+k(\hat{\beta})+\theta_1)^2+\delta(w-w_{t,C}+k(\hat{\beta})+\theta_2)^2\right]\\&\quad=\mathbb{E}\left[(w_{t,NC}(\theta_1;\hat{\beta})+\theta_1)^2+\delta(w-w_{t,NC}(\theta_1;\hat{\beta})+\theta_2)^2\right].\end{aligned}$$

Since the decision maker's beliefs $\hat{\beta}$ about her present bias affect her beliefs about what she will choose just before the tasks need to be completed, they affect demand for commitment.

Hypothesis DB2. Information provision will increase commitment demand for discounting-based dynamically inconsistent individuals. Formally, $\tilde{k}-k$ is increasing in $D_1(\theta_{1,1})$.

The reasoning behind this result is standard. As $\hat{\beta}$ falls, the payoffs that the decision maker expects to receive from the non-committed choice also fall. Since $\hat{\beta}$ only affects committed payoffs through the demand for commitment, $k(\hat{\beta})$ must increase proportionately to the signal the decision maker receives about her own level of procrastination.

The effect of information provision on the commitment demand of an individual with biased beliefs is more ambiguous; in general a clear prediction cannot be made without stronger assumptions either on the structure of decision makers' cost functions, or how beliefs are updated. Even a first order stochastic dominant shift in the distribution can increase or decrease demand.

3 Experimental Design

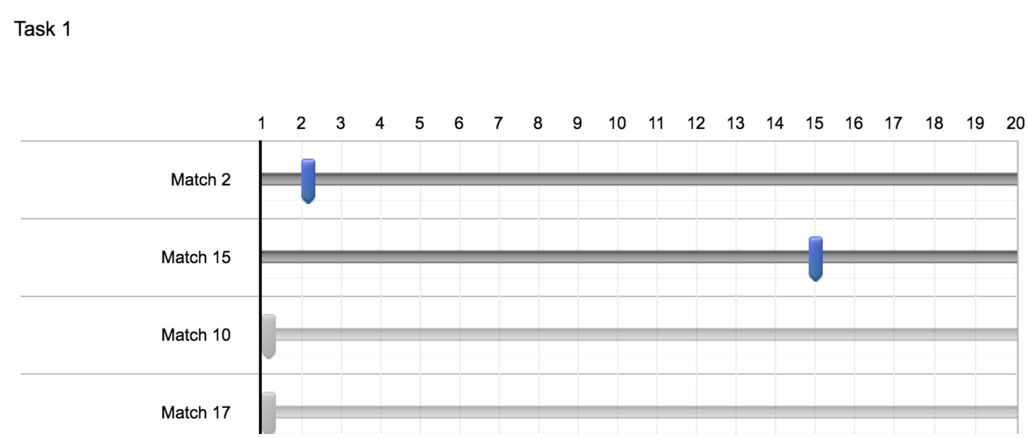
We implemented a longitudinal experiment to test these hypotheses. Undergraduate subjects were recruited through an online system to four different sessions throughout the semester and participated in the experiment for two weeks. To complete the study, subjects were required to complete eight surveys on the mornings of Monday through Thursday of each week and four sets of tasks in the evenings of Monday and Wednesday of each week. All surveys and tasks were distributed through Qualtrics.

To begin, subjects completed an introductory session in a lab. Subjects were first read an overview of the timeline and requirements of the study. They then logged on to the computer to complete a survey that included basic demographic information as well as a measure of present bias. The survey then presented five sample tasks for subjects to complete, and explained how the allocation and commitment decisions would be made. Finally, subjects were required to complete a comprehension quiz before advancing.

After the introductory session, all surveys and tasks were completed outside the lab on subjects' own devices.[13] Subjects were required to complete surveys and tasks at particular times. A link to each survey was sent out at 6 a.m. and subjects were instructed to complete the survey before noon that day. At noon, subjects who had not completed the task were sent a reminder and had two hours to complete the survey. If they did not complete the survey by 2 p.m., they were dropped from the study. A link to the tasks was sent out at 9 p.m. and tasks had to be completed before 4 a.m. the next morning.

3.1 Tasks

Figure 1: The first four sliders of a task



Notes: The figure shows the beginning of a task, showing the first four required sliders. Two of the sliders have been aligned and two remain to be aligned.

The tasks that subjects were required to complete consisted of moving sliders to

[13] While it was possible to complete the surveys on a smartphone, the task interface was easier to use on a computer.

match particular, predetermined levels. Slider tasks have proven useful in experimental settings as tasks that require real effort and focus from subjects (Gill and Prowse, 2012). In other experimental work, subjects have been required to set each slider to its midpoint, with the sliders offset to make the task more difficult. The software we employed did not allow sliders to be offset, so the required level of each slider was varied to increase difficulty. The order of the sliders was randomized for the same reason.

A single task consisted of moving nineteen sliders.[14] Each page included no more than 10 tasks, and subjects were unable to proceed to the next page if the current page was incomplete or if there were any errors. If subjects tried to proceed in these cases, they were informed that the task had a problem, but were not told which slider was incorrect. Figure 1 presents the first four sliders of an example task. The tasks were designed so that each would take about one minute to complete. The actual median time spent per task by subjects was 1 minute and 20 seconds.

3.2 Allocation Decisions and Commitment

Subjects made two allocation decisions each week. Each allocation decision consisted of dividing 10 tasks between Monday and Wednesday evenings. The first allocation was made when completing a survey on Monday morning, imposing at least a seven hour delay between when the allocative decision was made and when the tasks were actually carried out. The second allocation was made immediately before completing the tasks on Monday evening.

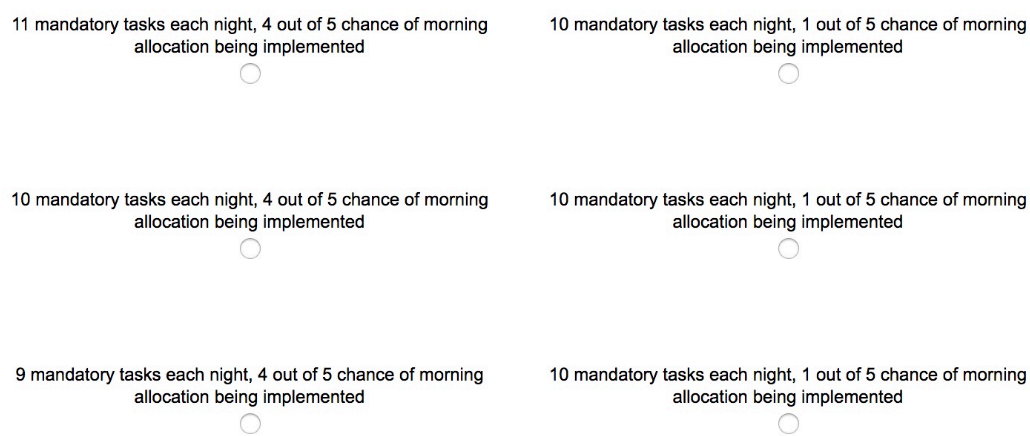
In addition to allocating tasks across evenings, subjects were also offered the chance to "commit," increasing the probability that the morning allocation would be the one implemented. If the subjects did not commit they had a one-in-five chance of the morning allocation being implemented. If the subjects did commit this probability rose to four out of five. The commitment was probabilistic rather than deterministic to preserve the incentive compatibility of the evening choices.

To elicit subjects' demand for commitment, they were given the choice of whether or not to commit at a variety of prices, both positive and negative. Due to previous work, including Augenblick et al. (2015), suggesting that many subjects' money-

[14]Each slider was initialized at the number one, but had to be clicked before it became active. To avoid subjects becoming confused by their tasks not being accepted due to an inactive slider, the number one was omitted from the potential target levels.

denominated willingness to pay for commitment is near 0, the prices were denominated in terms of mandatory tasks that would have to be done each night in addition to the tasks that were allocated to that night. Mandatory tasks could potentially vary between 4 and 16, depending on a subject's choices and which choice was implemented. A portion of the price list subjects faced can be seen in Figure 2.

Figure 2: Commitment price list



11 mandatory tasks each night, 4 out of 5 chance of morning allocation being implemented	10 mandatory tasks each night, 1 out of 5 chance of morning allocation being implemented
10 mandatory tasks each night, 4 out of 5 chance of morning allocation being implemented	10 mandatory tasks each night, 1 out of 5 chance of morning allocation being implemented
9 mandatory tasks each night, 4 out of 5 chance of morning allocation being implemented	10 mandatory tasks each night, 1 out of 5 chance of morning allocation being implemented

Notes: Subjects were asked to choose between pairs of task allocations. The choices in the left column also committed the subjects (with a higher probability) to carrying out their morning allocations.

3.3 Bedtime and Time Use Measurements

Both expected and actual bedtimes were elicited from subjects. In each morning survey, subjects were asked when they went to sleep the night before. Additionally, in both the morning and evening surveys subjects were asked at what time they expected to go to sleep that night. These predictions were deliberately not incentivized. An incentivized prediction could have functioned as a commitment device, and we wanted to make observable reductions in dynamic inconsistency by uncommitted agents. Intuitively, we wanted to see whether subjects would take the treatment to heart and apply its lessons outside the task domain.

Subjects also filled out diaries each morning describing their time use each hour for the previous day. The diaries allowed subjects to choose up to five activities for each hour from a menu.[15] In the data analysis, we allocate time uniformly over activities within the hour, yielding 12-minute resolution. This method of eliciting diaries balances precision against the limits of subjects' recall and the burden of completing the diaries. It is similar to the American Time Use Survey (ATUS) in that all subjects were asked for a sequential list of activities performed during the diary period, with responses constrained to total 24 hours. This method has been shown to yield high-quality estimates of time use (Hamermesh et al., 2005).

3.4 Information Treatment

Within each study wave, N subjects were randomly sorted and the first $N/2$ subjects were assigned to treatment.[16] In the second week of the study, treated subjects were given information about their own past choices. The treatment, a real example of which can be seen in Figure 3, consisted of three main parts. The first described the allocation choices that the individual made the week before, emphasizing whether or not any tasks were reallocated on Monday evening. The second part reported the subject's average actual and predicted bedtimes and gave the difference between them in minutes. Finally, treated subjects were asked why someone's choices and predictions might change throughout the day. Subjects were given a blank space in which they had to type something to proceed.

The treatment information was intentionally neutral to avoid experimenter demand effects. In particular, we did not use judgmental language when describing the change in task allocation. We provided subjects with information that they could have recorded for themselves had they chosen to do so. Finally, we did not mention commitment.

This information was given to treated subjects (and only treated subjects) on Monday morning of the second week. They were shown the information after they reported their bedtime for the previous night and made a prediction for Monday night but before they made the commitment and allocation decisions.

[15] The activity menu included the following: class, exercising, other, sleeping, socializing, studying, TV, and working.

[16] Random sorting was based on a single draw for each subject from a uniform distribution on $[0, 1]$. No re-randomization was performed, nor was any blocking or stratification employed within study wave. For N odd, the first $(N/2) + .5$ subjects were assigned to treatment.

Figure 3: Treatment

Choosing the Implemented Allocation

Last week, on Monday morning you said you'd do 15 tasks on Monday evening and 7 tasks on Wednesday. When you were asked in the evening, you decided to do 16 on Monday, and 6 on Wednesday. **Thus, you moved 1 task from Wednesday to Monday.**

Also, on average you predicted that your bedtime would be 12:30 AM, and your actual average bedtime was 1:42 AM, **so you missed your predicted bedtime by about 72 minutes.**

Why might someone's choices and predictions change throughout the day?

There may be unforeseen things that pop up throughout the day that keep them busier than they thought or they miscalculate how long something will take

Notes: An example of an actual message that one of the treated subjects received at the beginning of week 2 of the experiment, along with the response they entered. The information was provided to subjects just before they made commitment and allocation decisions. The text given in the box is an example of a response that a subject gave to the open-ended question about why someone's choices and predictions might change. The box was empty when subjects were presented with the message.

3.5 Payments

Subjects received $40 total for completing the full study. An initial payment of $10 was made to all subjects on Thursday or Friday of the first week. The second payment of $30 was made to the subjects on Thursday or Friday of the second week, conditional on all portions of the experiment being completed on time.

4 Data

A total of 274 undergraduate subjects were recruited through an online system and completed the introductory session. Twenty-six of these subjects did not complete some surveys and left the experiment having received only the initial payment of $10. The vast majority of those who dropped out of the experiment did so in the first week of their participation. Another 39 subjects missed the completion deadlines for at least one survey, though they eventually did answer all surveys. These subjects are excluded from the primary sample, leaving a final baseline sample of 209 subjects.

Table 1 tests baseline covariate balance. Tests of differences between the two groups are corrected for multiple hypothesis testing using the procedure from List et al. (2019). We find no statistically significant differences across the two groups. The difference in gender, however, could be practically important. We control for

Table 1: Treatment-control balance

	Control Mean/(SD)	Treatment Mean/(SD)	Diff./[p-value]
Commitment demand week 1	-0.92	-0.13	-0.79
	(3.04)	(3.27)	[0.30]
Tasks deferred week 1	0.030	0.0092	0.021
	(2.84)	(1.87)	[0.95]
Bedtime difference from plan (minutes)	36.1	39.8	-3.68
	(68.7)	(53.9)	[0.89]
GPA	3.22	3.31	-0.085
	(0.49)	(0.47)	[0.59]
Female (indicator)	0.54	0.70	-0.16
	(0.50)	(0.46)	[0.12]
Study wave	2.47	2.56	-0.090
	(1.13)	(1.09)	[0.91]
Observations	100	109	

Notes: The significance of the differences is assessed using the procedure from List et al. (2019). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

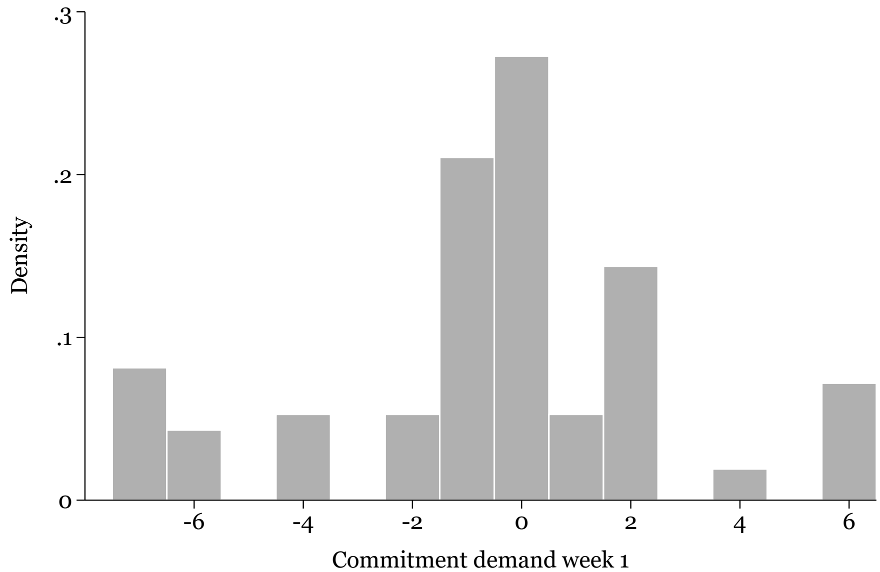
gender in the regression analyses below and test for treatment heterogeneity by gender in Section 5.1. Some 15% of subjects deferred work in the first week, consistent with discounting-based inconsistency ($\beta < 1$), belief-based inconsistency ($\hat{\alpha} < \alpha$), or both. Another 15% of subjects pulled work forward.

Figure 4 shows a histogram of subjects' commitment demand in the first week.[17] As noted in Section 3, subjects were able to choose between: (a) doing 10 tasks and having a low chance of being committed or (b) doing a different number of tasks and having a high chance of being committed. A commitment demand of one indicates that the subject was willing to do one extra task to be committed, but was unwilling to do two. Since this elicitation method is necessarily bounded, having a commitment demand of six indicates that a subject is willing to do *at least* six extra tasks to be committed, while a commitment demand of negative seven indicates that the subject is unwilling to commit even if it lowers the number of tasks she must do by six.

Denominating the commitment price in tasks generated a larger spread in commitment demand than has been seen in similar experiments. Over 28% of subjects were willing to do at least one extra task in order to be committed in the first week,

[17]The commitment demand in week 2 is shown in Figure 6.

Figure 4: Commitment demand in week 1



Notes: The figure shows the distribution of commitment demand for subjects in week 1, before treatment. The x-axis shows the price paid for commitment in terms of extra tasks.

while over 22% were willing to do one extra task to be flexible. Of the remaining subjects whose value of commitment was near zero, just over half chose to commit at a price of zero. This level of commitment demand is higher than in previous experiments that relied on monetary payments. To see this, one can roughly calculate an equivalent cash price as follows. The travel cost literature has estimated that people value time at roughly 72 (Lam and Small, 2001) to 93 percent (Small et al., 2005) of the wage rate. The median wage among subjects was $14, so the implied range of values was $10 to $13. As median task completion time was 1 minute 20 seconds, the cash value of one task was approximately 22 to 29 cents. In contrast, Augenblick et al. (2015) find only 9% of subjects were willing to pay $0.25 to be committed (the lowest price offered), and 10% of subjects were willing to pay $0.25 for flexibility. We find that more than twice as many subjects were willing to be committed when paying in terms of tasks.

Table 2 summarizes attrition in the experiment. Column 1 presents means for the 64 subjects (23%) who did not complete the experiment and Column 2 presents means for those who did, pooling treatment and control. There are no statistically

Table 2: Observables by attrition status

	Did not finish Mean/(SD)	Finished study Mean/(SD)	Diff./[p-value]
Treat	0.44	0.52	-0.084
	(0.50)	(0.50)	[0.67]
Age	20.2	20.5	-0.23
	(1.55)	(1.83)	[0.68]
GPA	3.19	3.27	-0.074
	(0.49)	(0.48)	[0.74]
Female (indicator)	0.56	0.62	-0.060
	(0.50)	(0.49)	[0.64]
Study wave	2.38	2.52	-0.14
	(1.15)	(1.11)	[0.41]
Observations	64	209	

Notes: The significance of the differences is assessed using the procedures in List et al. (2019). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

significant differences across the two groups. Appendix Table 8 reports results from a regression of a study completion dummy on observable characteristics from our baseline survey. Point estimates are uniformly small, and in a joint test of the null hypothesis that all coefficients are zero, we fail to reject.

5 Primary Results: Testing Model Hypotheses

In this section we test the hypotheses from the theory presented in Section 2. We first examine Hypothesis DB1 and Hypothesis BB1 by estimating variants of the following equation.

$$\text{Deferred}_{i2} = \gamma_0 + \gamma_1 \text{Treat}_{i2} + \gamma_2 \text{Deferred}_{i1} + \gamma_3 \text{Treat}_{i2} \text{Deferred}_{i1} + \mathbf{x}_i' \gamma_4 + \varepsilon_{i2} \quad (5)$$

where $Deferred_{it}$ is the number of tasks deferred by subject i in week t, $Treat_{i2}$ is an indicator for being in the treatment group in week 2, $\mathbf{x}_i$ is a vector of control variables, and ε_{i2} is the stochastic error term associated with this regression. We test robustness to alternative control sets across different specifications. First, we report results with indicators for study wave, as randomization into treatment was conditional on wave. Second, we add gender, age, and age squared. Finally, we add controls for grade point average (GPA) prior to the experiment, self-reported busyness during week 1, and employment. Here and in subsequent regressions, we report heteroskedasticity-robust standard errors (White, 1980). We account for multiple testing over our primary hypotheses using the Romano-Wolf stepwise procedure (Romano and Wolf, 2005), which controls the familywise error rate. Specifically, we adjust for multiple testing of estimated coefficients on our regressor of interest, $Treat_{i2} Deferred_{i1}$, using 10,000 bootstrap replications.

Table 3: Effect of treatment on work deferred

	(1) Work deferred week 2	(2) Work deferred week 2	(3) Work deferred week 2
Treat	0.31	0.37	0.29
	(0.38)	(0.37)	(0.39)
Work deferred week 1	0.33**	0.33**	0.34**
	(0.14)	(0.14)	(0.14)
Treat × Work deferred week 1	-0.41**	-0.43**	-0.42**
	(0.20)	(0.21)	(0.21)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Romano-Wolf p-value on interaction	0.044	0.039	0.045
Adj. R^2	0.03	0.02	0.05
Observations	209	209	209

Notes: The table shows results from estimating equation 5. Each column shows the results of a separate regression. All regressions include study wave indicators. Demographic controls are gender, age, and age squared. Additional controls are GPA, GPA squared, self-reported busyness during week 1, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). The Romano and Wolf (2005) p-value is a multiple hypothesis-corrected test of the null hypotheses of 0 for the coefficients on $Treat_{i2} Deferred_{i1}$ here and in Table 4. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 3 presents results based on estimating equation (5), in which the dependent variable is work deferred in week 2. Estimates are similar across the specifications. The estimated treatment effect on subjects who did not defer work in week 1 is positive, but we fail to reject a zero null hypothesis at conventional significance levels.

The second coefficient in each column shows that deferring work in week 1 is positively and significantly associated with deferring work in week 2. Finally, the third coefficient shows that, for subjects who deferred work in week 1, the treatment caused a statistically significant and practically substantial reduction in deferral of tasks in week 2. The magnitudes of the coefficients on work deferred in week 1 and the interaction term are similar but the signs are opposite, so the marginal effect of deferring a task in week 1 on treated subjects is approximately zero. Subjects who are reminded of their past procrastination procrastinate less in week 2 than their week 1 behavior would otherwise predict.

This effect provides evidence against Hypothesis DB1, that information provision will have no effect on work allocation. Because a naïve β-δ agent's allocation depends on present bias (and not beliefs about present bias), Hypothesis DB1 implies a zero coefficient on the interaction of treatment and work deferred in week one. The agent may update her belief $\hat{\beta}$, but this affects commitment demand, not task allocation. The estimates are consistent with Hypothesis BB1 that information provision will decrease the difference between present and future allocations. This is our first empirical evidence that biased beliefs (optimism) influence time-inconsistent behaviors.

We report heterogeneity analysis for this effect in Appendix Table 9. Each morning, subjects provided self-reported measures of how busy they expected to be during the day. For individuals who turned out to be procrastinators but who reported, on the first day of the experiment, that they would be less busy than the median respondent, the coefficient of interest was 55 percent larger than the average effect reported in Table 3. Individuals who were naïve about their own time shocks were more strongly affected by the treatment, consistent with the theory of belief-based dynamic inconsistency.

To understand whether the effects in week 2 operated by changing subjects' morning or evening task allocations, Appendix Table 10 presents estimates similar to Table 3, but the dependent variable is tasks chosen in the morning of day 1 (Monday) or tasks actually performed in the evening of day 1. Estimated effects on morning choices are negative, and 70 percent as large as effects on tasks deferred in Table 3, though less precise. Subjects who were reminded that they deferred work in the prior week planned less work for the first night. This behavior is also inconsistent with Hypothesis DB1 but consistent with Hypothesis BB1.

So far, we have found evidence that subjects make dynamically inconsistent choices due to belief bias. We next test Hypothesis DB2, that treated procrastinators in week 1 will increase their commitment demand in week 2. If subjects do behave consistent with DB2, that provides evidence for discounting-based dynamic inconsistency. The estimating equation models the difference in commitment demand across weeks as a function of treatment interacted with procrastination.

$$\Delta\text{Commitment}_i = \theta_0 + \theta_1\text{Treat}_{i2} + \theta_2\text{Treat}_{i2}\text{Deferred}_{i1} + \mathbf{x}_i'\theta_3 + \nu_i \tag{6}$$

In the equation above, $\Delta Commitment_i$ is the change in commitment for subject i (week 2 minus week 1), ν_i is the stochastic error term for this regression, and other variables are defined as in equation (5). Again we adjust for multiple hypothesis testing of coefficients on $Treat_{i2}Deferred_{i1}$ using the Romano-Wolf stepwise procedure (Romano and Wolf, 2005).

Table 4: Effect of treatment on commitment demand

	(1) Change in commitment demand	(2) Change in commitment demand	(3) Change in commitment demand
Treat	-0.11	-0.11	-0.22
	(0.42)	(0.43)	(0.42)
Treat × Work deferred week 1	0.44**	0.44**	0.45**
	(0.16)	(0.16)	(0.16)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Romano-Wolf p-value on interaction	0.020	0.019	0.017
Adj. R^2	0.02	0.01	0.00
Observations	209	209	209

Notes: The table shows results from estimating equation 6. Each column is a separate regression. Demographic controls are an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. Additional controls are GPA, GPA squared, and self-reported busyness in week 1. In parentheses are heteroskedasticity-robust standard errors (White, 1980). The Romano and Wolf (2005) p-value is a multiple hypothesis-corrected test of the null hypotheses of 0 for the coefficients on $Treat_{i2}Deferred_{i1}$ here and in Table 3. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

The corresponding estimates appear in Table 4. Consistent with theory, the first row shows that the effect of treatment on subjects who defer zero tasks in week

1 is small. The second coefficient shows the interaction of treatment with tasks deferred in week one. The estimated coefficient on the interaction term is positive and statistically significant at the five percent level. This is consistent with Hypothesis DB2. Our model predicts that when a present-biased individual receives information, she will update her belief over her present bias and increase her commitment demand in response. The addition of demographic and busyness controls does not change the estimated coefficient.

Heterogeneity analysis, reported in Table 9, again points to these results being stronger for naifs. We interact treatment and week 1 procrastination with an indicator variable equal to 1 if, at baseline, the subject responded "yes" to a question asking whether they tend to procrastinate. The effect of treatment is roughly twice as large for individuals who stated that they were not procrastinators but who did end up procrastinating on our experimental tasks.

The results presented above refute the hypothesis that discounting alone drives dynamic inconsistency. Both beliefs and discounting are important for explaining agent behavior in this setting. Using commitment demand and task allocation behavior, we can classify individual subjects as time consistent, discounting-based time inconsistent, belief-based time inconsistent, or both discounting- and belief-based time inconsistent. The procedure works as follows. First, we match each treatment group subject to a single, nearest neighbor control group subject using propensity score. We use week 1 commitment demand and tasks deferred as well as the baseline control variables for GPA, gender, age, and experiment wave to perform the matching.

Next, we estimate individual-level versions of the regressions above.[18] If a time inconsistent treatment group member increased her commitment demand more than her control group peer, we classify the subject's behavior as discounting-based. If a time inconsistent treatment group member reduced the amount of work she deferred in week 2 more than her control group peer, then we classify the subject's behavior as belief-based.

Using this classification, we find that conditional on changing their work allocation in week 1, 85% of subjects exhibited behavior consistent with one of these two models. 25% of the subjects behaved in a manner consistent with discounting-based dynamic inconsistency. 39% of subjects behaved in a manner consistent with belief-based

[18] In addition to the individual-level regressions, we also estimate a finite mixture model with four latent classes. These results are reported in Table 11 of Appendix A.

dynamic inconsistency, and 21% of subjects exhibited behavior consistent with both mechanisms.

5.1 Robustness Checks

The initial robustness checks for the headline results appear in Tables 3 and 4. As noted in Section 4, our sample displays a subjectively large treatment-control difference on one potentially important demographic characteristic—gender. The results in Tables 3 and 4, however, show that this imbalance did not have a substantial effect on the estimates. Appendix Tables 12 and 13 interact treatment variables with gender. Marginal effects of treatment on female procrastinators are similar to those for males, and we fail to reject a null of zero difference.

To evaluate whether our results arise from outliers, we Winsorize the upper and lower 2.5% of observations on tasks deferred and commitment demand in both weeks, then re-estimate our primary specifications. Appendix Tables 14 and 15 show that under this procedure point estimates are strongly similar and they remain statistically significant at conventional levels.

Appendix Section B reports the main results, excluding subjects who switched multiple times when choosing from the commitment price list. (In the results reported in the body of the paper, we include all subjects.) If at any point the subject violated the law of demand, we classify the subject as a multiple switcher. Appendix Section B shows that whether or not these subjects are excluded, the results remain substantively unchanged.

Finally, we empirically assess experimenter demand effects by leveraging the two pieces of information given to treatment group members. Recall that treatment subjects were told about any changes in their task allocation and about deviations from their bedtime plan. Assuming that experimenter demand effects would have been stronger when both of these reported figures were consistent—for instance, when the treatment message indicated that the subject both deferred tasks and delayed bedtime—and larger, then we can assess experimenter demand by interacting bedtime prediction error with the other variables in equations (5) and (6). We report these interaction models in Table 16. In neither case does the interaction between the bedtime error message and tasks deferred in week 1 substantially change the conclusions from the baseline analysis. We take this as evidence against strong experimenter demand effects.

6 Secondary Results: Effects on Real-World Behavior

To this point we have focused on testing theoretical predictions about task choices and commitment. These analyses allow for a close connection between theory and data, but our experimental design also allows us to study procrastination behavior in another domain—real-world time use. Bedtime has important effects on sleep duration and well-being (Gibson and Shrader, 2018), but people frequently revise their initial plans and go to bed later (Kroese et al., 2014). In addition, our panel time-use data allow evaluation of how changes in commitment price alter the entire time allocation.

6.1 Effect of Treatment on Bedtime Forecast Error

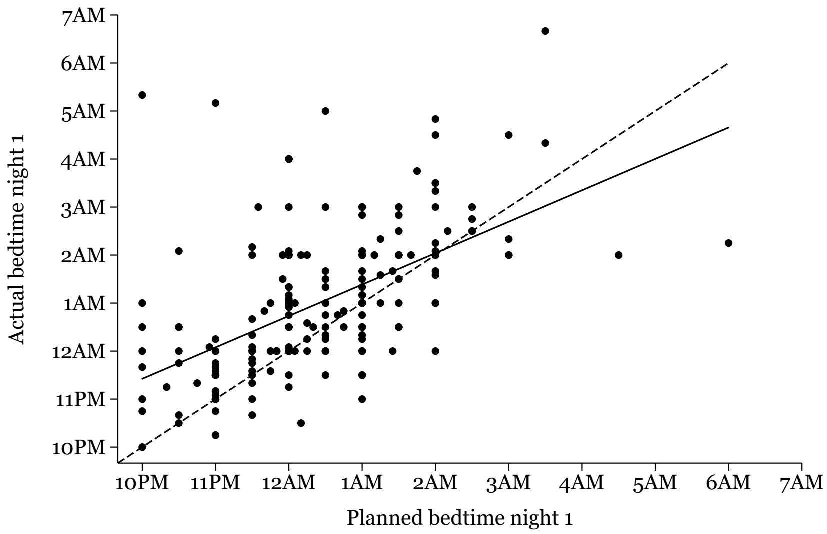
Table 1 provides initial evidence that subjects do indeed procrastinate around bedtime in our setting. The average values for "Bedtime difference from plan" in Table 1 indicate that agents miss their planned bedtimes by 36 to 40 minutes. In a regression of this variable on a constant, the pooled mean of 38 minutes is significantly different from zero (the robust standard error is 4.24). Subjects were not incentivized to meet their bedtime plans, so this behavior plausibly reflects subjects' decision making in a non-experimental setting. It is possible in principle that asking for bedtime predictions generated anchoring effects (Tversky and Kahneman, 1992), but such effects would have reduced forecast errors for both treatment and control subjects.

Delving deeper, Figure 5 compares individual planned bedtimes (horizontal axis) to self-reported actual bedtimes (vertical axis) for the first night of the study.[19] The majority of subjects miss their planned bedtime and appear as points above the 45-degree line, corroborating the earlier finding that subjects are generally optimistic about their bedtime plans. A linear fit (solid, black line) shows that, on average, subjects underestimate their bedtime earlier in the evening and tend to overestimate it later in the evening, though the overestimation is supported by relatively few observations.

Figure 5 displays a single night's planned and actual bedtimes in order to highlight three types of noise in the data. First, there is considerable round-number heaping in

[19] Our choice of the first night is arbitrary; analogous figures for other nights look strongly similar. Comparison of bedtime plans, self-reports and readings from sleep monitors suggests subjects frequently failed to correctly enter a.m. or p.m. For both planned and self-reported bedtimes, we assume reports in the range from 10 a.m. to 3 p.m. reflect this type of error.

Figure 5: Planned and actual bedtimes, first night



Notes: The figure shows planned bedtime (x-axis) versus actual bedtime (y-axis) for all observations on the first night of the study. Points above the 45 degree line (dashed), indicate that subjects went to bed later than their stated plan. A linear fit (solid, black line) shows that, on average, subjects underestimated their planned bedtime when going to bed earlier in the evening and overestimated later in the evening.

both planned and actual bedtimes. Planned bedtimes in particular are likely to fall on the hour or half hour. Second, 35 subjects report actual bedtime exactly equal to planned bedtime. Although these subjects might have gone to bed around the time that they planned, an exact match between plan and realization could reflect misreporting. Finally, some subjects report extreme bedtime plans and realizations. On this night, for instance, 9 subjects reported going to bed later than 4 a.m. Although extreme bedtime values are not necessarily in error, erroneous extreme bedtimes could exert disproportionate influence on regression analysis.

As previously described in Section 3.4, treated subjects were given information about their own time use decisions at the beginning of week 2 of the experiment. To test whether this reduced procrastination, we re-estimate equation (5) using forecast error as the outcome. To reduce the influence of noise in the bedtime data highlighted above, we employ a trichotomized measure of forecast error. We classify bedtime forecast error into one group if the subject underestimated her bedtime, a second group if her prediction exactly matched her actual bedtime, and a third group if she overestimated her bedtime. Table 5 shows results from estimating the effect of treatment on this trichotomized bedtime error variable, using an ordered logit model.[20] On average, treated subjects moved to lower categories, representing bedtime equal to or earlier than forecast. There are multiple potential mechanisms for this change, including commitment devices outside the experiment and changes in salience. Movement across categories was larger for those with larger week 1 forecast errors.

Table 5: Effect of treatment on bedtime forecast error

	(1) Discrete forecast error week 2	(2) Discrete forecast error week 2	(3) Discrete forecast error week 2
Treat	-0.59**	-0.60**	-0.68**
	(0.30)	(0.30)	(0.31)
Forecast error week 1	0.70**	0.70**	0.75***
	(0.29)	(0.30)	(0.29)
Treat × Forecast error week 1	-0.85**	-0.84**	-0.91**
	(0.36)	(0.37)	(0.36)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Observations	202	202	202

Notes: Subjects who do not go to sleep at all are excluded from the sample. Each column of the table shows results from a separate regression. All regressions contain the following controls: an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Marginal effects of treatment, reported in Table 6, allow us to assess the magnitudes of these responses and see whether subjects moved into the group with bedtime equal to forecast, or the group with bedtime before forecast. Column 1 gives marginal effects of treatment with week 1 error at its mean, which is equal to zero because this variable is standardized. Treatment reduced the probability of going to bed later than

[20]Subjects who do not go to sleep at all are excluded from the sample, as their forecast errors are not well defined. Most of these subjects report studying through the night. Assigning these subjects to the first group does not meaningfully change the estimates. Appendix Table 17 reports results without the adjustment of planned bedtimes between 10 a.m. and 3 p.m.. Estimates are strongly similar in both magnitude and statistical significance.

forecast by 13 percentage points, a large change compared to the predicted probability of 62 percent with all variables at their means. The offsetting increases were split roughly equally over the other two categories. Column 2 gives marginal effects of treatment with week 1 error equal to 1 (1 standard deviation above the mean). Comparing to column 1, two differences are apparent. The reduction in the probability of going to bed late is greater, and the increase in the probability of going to bed on time is greater than the increase in the probability of going to bed early (though the latter two are not statistically distinguishable). That is, week 1 procrastinators reduced week 2 procrastination more than did other subjects, and the principal mechanism was going to bed on time. Finally column 3 gives marginal effects of treatment with week 1 error equal to 2 (2 standard deviations above the mean). Here the pattern of column 2 becomes still more pronounced. The probability of procrastination falls by 45 percentage points, and the offsetting increase in the probability of going to bed on time is 10 percentage points greater than the increase in the probability of going to bed early. Again this suggests procrastinators responded more strongly to treatment, and that they did so by going to bed on time.

Table 6: Marginal effect of treatment, varying week 1 forecast error

	(1) Week 1 forecast error at mean	(2) Week 1 forecast error = 1	(3) Week 1 forecast error = 2
Pr(Bedtime < forecast)	0.063*	0.13***	0.17**
	(0.036)	(0.049)	(0.076)
Pr(Bedtime = forecast)	0.067*	0.18***	0.27***
	(0.037)	(0.068)	(0.10)
Pr(Bedtime > forecast)	-0.13*	-0.31***	-0.44***
	(0.071)	(0.11)	(0.17)
Observations	202	202	202

Notes: Subjects who do not go to sleep at all are excluded from the sample. Each column of the table shows the marginal effects of treatment, estimated based on column 2 of Table 5, on the probabilities of being in each of the three error categories: bedtime prior to forecast, bedtime equal to forecast, and bedtime after forecast. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Taken together, our bedtime results indicate that the type of dynamic inconsistency we study matters for consequential real-world decisions. They corroborate our task-based results showing that information provision can alter procrastination behavior, suggesting a purely discounting-based model of dynamic inconsistency is incomplete.

6.2 Effect of Treatment on Other Activities

Bedtime is not the only element of an agent's time allocation problem that may be influenced by time shocks and beliefs over those shocks. Indeed time shocks may influence an agent's entire time allocation, and optimizing responses to such shocks are substantially understudied. Shifts in within-day time use can affect health and productivity (Bessone et al., 2019). Using time budget recall data, we are able to investigate the effect of an experimentally administered time shock on other time use choices. The randomly assigned shock that subjects face is the price at which they are offered commitment (i.e. the line of the price list which is implemented). Empirically, a one-unit increase in the price of commitment is associated with a subject needing to complete an additional two-thirds of a task. Each task took about 1.3 minutes for the median participant and 2.2 minutes for the mean participant in the study, so the treatment is equal to about 1 to 1.5 minutes of induced experimental time. This randomization allows for a series of tests on realized time use. While these are interesting reduced-form exercises based on random variation, they are not theoretically founded and results should be interpreted cautiously.

We estimate equations of the form

$$\Delta\text{Time}_i = \alpha_0 + \alpha_1 \Delta\text{Time Shock}_i + \mathbf{x}_i'\alpha_2 + \upsilon_i \tag{7}$$

In the above equation $\Delta\,Time_i$ is the change in time spent on a given activity between weeks 1 and 2, $\Delta\,Time\ Shock_i$ is the change in commitment price between the two weeks, υ_i is the stochastic error term associated with this regression, and all other variables are the same as in equation (5).

Table 7 reports the estimated effects of a marginal increase in commitment price on time use, measured in minutes per day. Within this table, we correct for multiple testing using the procedure of Romano and Wolf (2005) with 10,000 bootstrap replications. None of the results are statistically significant at conventional levels, but point estimates are nonetheless instructive. When the price of commitment increases

Table 7: Time use and commitment price

	Class	Exercising	Other	Sleeping
Time Shock diff.	-0.95	0.38	4.94	-1.28
	(1.32)	(0.54)	(2.31)	(1.45)
Observations	209	209	209	209
Romano-Wolf p-value	0.85	0.85	0.22	0.85

	Socializing	Studying	TV	Working
Time Shock diff.	1.45	-5.04	2.25	-1.75
	(1.95)	(2.54)	(1.22)	(0.91)
Observations	209	209	209	209
Romano-Wolf p-value	0.85	0.28	0.28	0.28

Notes: The table shows results from estimating equation 7. Each column is a separate regression. Time use changes are in minutes. All regressions contain the following controls: an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). (Romano and Wolf, 2005) multiple hypothesis correction procedure is carried out on all tests and the p-values are based on 10,000 bootstrap replications. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

by one task (roughly 1 to 1.5 minutes), studying decreases by roughly five minutes. "Other time" and television time increase by roughly 5 and 2 minutes, respectively; the latter result is consistent with TV being a time use luxury, as found in Aguiar et al. (2017). This pattern of results is potentially consistent with present-biased preferences, as agents substitute toward immediately pleasurable time uses and away from a time use with largely deferred payoffs.

7 Conclusion

This paper models agents whose dynamic inconsistency potentially arises from two sources: discounting and beliefs. Agents with optimistic beliefs about future time shocks will exhibit dynamically-inconsistent choices over effort that are observationally equivalent to those driven by present bias. An informational intervention that tells agents about their past time inconsistency, however, will yield different behavior for these two biases. Optimistic agents will change effort allocations, but agents with present bias will not. Present biased agents will increase commitment demand, while optimistic agents will not necessarily do so.

We test these different predictions experimentally and find that both preferences and beliefs matter for time inconsistency. The results help explain puzzlingly low take-up of costly commitment and offer an alternative policy prescription to help overcome time inconsistent behavior—providing information on agents' own past execution of their plans.

The welfare effects of an informational treatment on behaviorally biased agents are unclear. To the extent that information pushes a naïvely present-biased decision maker towards sophistication, she will be better able to make plans that account for her present-biased future self. On the other hand, if optimistic beliefs are directly valuable to decision makers, giving them clear evidence their beliefs are biased could make them worse off. One avenue for future research is to identify situations in which subjects demand this information, and how it can be structured to reduce time inconsistency with minimal associated welfare losses.

References

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

Aguiar, M., M. Bils, K. K. Charles, and E. Hurst (2017). Leisure luxuries and the labor supply of young men. *NBER Working Paper No. 23552*.

Andreoni, J. and C. Sprenger (2012a). Estimating time preferences from convex budgets. *American Economic Review 102*(7), 3333–56.

Andreoni, J. and C. Sprenger (2012b). Risk Preferences Are Not Time Preferences. *American Economic Review 102*(7), 3357–3376.

Ariely, D. and K. Wertenbroch (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science 13*(3), 219–224.

Ashraf, N., D. Karlan, and W. Yin (2006). Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines. *Quarterly Journal of Economics 121*(2), 635–672.

Augenblick, N., M. Niederle, and C. Sprenger (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics* (2001), 1067–1115.

Augenblick, N. and M. Rabin (2018). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*.

Avery, M., O. Giuntella, and P. Jiao (2019). Why don't we sleep enough? A field experiment among college students.

Barro, R. J. (1999). Ramsey meets Laibson in the neoclassical growth model. *Quarterly Journal of Economics 114*(4), 1125–1152.

Bénabou, R. and J. Tirole (2002). Self-confidence and personal potivation. *Quarterly Journal of Economics 117*(3), 871–915.

Bessone, P., G. Rao, F. Schilbach, H. Schofield, and M. Toma (2019). Sleepless in Chennai: The consequences of increasing sleep among the urban poor. Technical report.

Börsch-Supan, A. H., T. Bucher-Koenen, M. D. Hurd, and S. Rohwedder (2018). Saving regret. Technical report, National Bureau of Economic Research.

Browning, M. and J. Tobacman (2015). Discounting and optimism equivalences.

Brunnermeier, M. K., F. Papakonstantinou, and J. A. Parker (2016). Optimal time-inconsistent beliefs: Misplanning, procrastination, and commitment. *Management Science 63*(5), 1318–1340.

Brunnermeier, M. K. and J. A. Parker (2005). Optimal expectations. *American Economic Review 95*(4), 1092–1118.

Camerer, C. and D. Lovallo (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review 89*(1), 306–318.

Carrera, M., H. Royer, M. Stehr, J. Sydnor, and D. Taubinsky (2019). How are preferences for commitment revealed? Technical report, National Bureau of Economic Research.

Casaburi, L. and R. Macchiavello (2019). Demand and supply of infrequent payments as a commitment device: Evidence from Kenya. *American Economic Review 109*(2), 523–555.

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

Eil, D. and J. M. Rao (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics 3*(2), 114–38.

Falk, A., D. Huffman, and U. Sunde (2006). Self-confidence and search. *IZA Working Paper No. 2525*.

Frazis, H. and J. Stewart (2012). How to Think About Time-Use Data: What Inferences Can We Make About Long- and Short-Run Time Use from Time Diaries? *Annals of Economics and Statistics 105/106*, 231–245.

Frederick, S., G. Loewenstein, and T. O'Donoghue (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature 40*(2), 351–401.

Gabaix, X. (2014). A Sparsity-Based Model of Bounded Rationality. *Quarterly Journal of Economics 129*(4), 1661–1710.

Gabaix, X. and D. Laibson (2017). Myopia and Discounting. *NBER Working Paper No. 23254*, 37.

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

Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review 102*(1), 469–503.

Halevy, Y. (2008). Strotz meets allais: Diminishing impatience and the certainty effect. *American Economic Review 98*(3), 1145–62.

Hamermesh, D. S., H. Frazis, and J. Stewart (2005). Data watch: The American Time Use Survey. *Journal of Economic Perspectives 19*(1), 221–232.

Kahneman, D. and A. Tversky (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases*, pp. 414–421. Cambridge University Press.

Kroese, F. M., D. T. De Ridder, C. Evers, and M. A. Adriaanse (2014). Bedtime procrastination: introducing a new area of procrastination. *Frontiers in Psychology 5*, 611.

Krugman, P. (2015). Wearables and self-awareness. `https://krugman.blogs.nytimes.com/2015/03/09/wearables-and-self-awareness-personal/`.

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

Laibson, D. (2015). Why Don't Present-Biased Agents Make Commitments? *American Economic Review 105*(5), 267–272.

Lam, T. C. and K. A. Small (2001). The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review 37*(2-3), 231–251.

Larwood, L. and W. Whittaker (1977). Managerial myopia: Self-serving biases in organizational planning. *Journal of Applied Psychology 62*(2), 194.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics 22*(4), 773–793.

Malmendier, U. and G. Tate (2005). CEO overconfidence and corporate investment. *The Journal of Finance 60*(6), 2661–2700.

Malmendier, U. and G. Tate (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics 89*(1), 20–43.

Mueller, A. I., J. Spinnewijn, and G. Topa (2018). Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence and Bias. *NBER Working Paper No. 25294*, 70.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics 122*(3), 1067–1101.

O'Donoghue, T. and M. Rabin (1999). Doing It Now or Later. *American Economic Review 89*(1), 103–124.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Roy, M. M., N. J. Christenfeld, and C. R. McKenzie (2005). Underestimating the duration of future events: Memory incorrectly used or memory bias? *Psychological Bulletin 131*(5), 738.

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

Small, K. A., C. Winston, and J. Yan (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica 73*(4), 1367–1382.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies 23*(3), 165.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica 47*(2), 143–148.

Taylor, S. E. and J. D. Brown (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin 103*(2), 193.

Thaler, R. H. and S. Benartzi (2004). Save more tomorrow$^{\text{TM}}$: Using behavioral economics to increase employee saving. *Journal of political Economy 112*(S1), S164–S187.

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

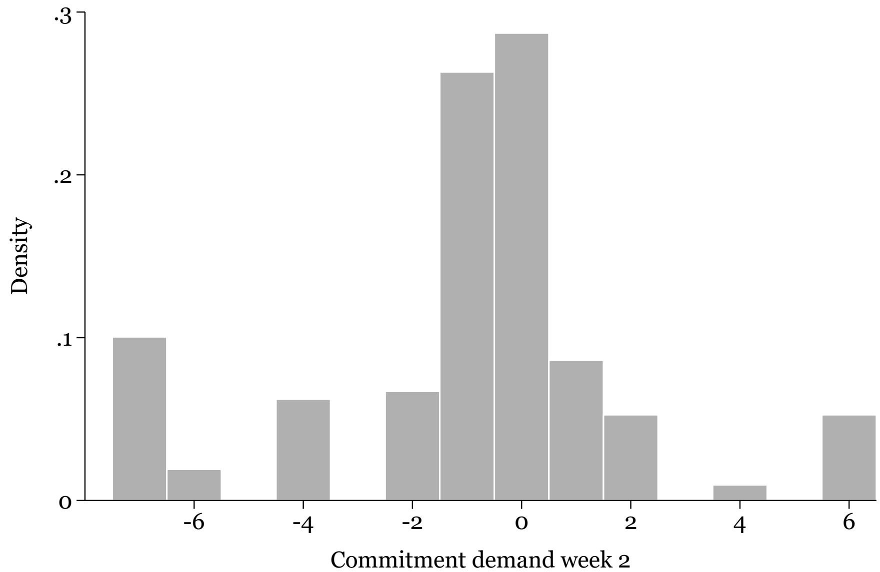
Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

Appendix for online publication

A Appendix figures and tables

Figure 6: Commitment demand week in 2



Notes: The figures show commitment demand for subjects in week 2, after treatment. The x-axis shows the price paid for commitment in terms of extra tasks.

Table 8: Regression of study completion dummy on observables

	Finished study
Treat	0.052
	(0.053)
Age	0.018
	(0.013)
GPA	0.070
	(0.055)
Female (indicator)	0.035
	(0.054)
Study wave	0.024
	(0.023)
F	1.21
p-value	0.30
Observations	273

Notes: Sample includes all subjects who completed our baseline survey instrument: 64 who did not complete the study and 209 who did. Estimates are from a regression of a study completion dummy on the listed variables. No other variables are included. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Treatment interactions with self-reported naïvete measures

	(1) Change in commitment demand	(2) Change in commitment demand	(3) Work deferred week 2
Treat	-0.71	-0.69	0.14
	(0.69)	(0.59)	(0.49)
Treat × Tend to procrastinate	0.77		
	(0.67)		
Treat × Work deferred week 1	0.79***	0.88***	-0.67**
	(0.13)	(0.29)	(0.33)
Work deferred week 1			0.33**
			(0.14)
Treat × Work deferred week 1 × Tend to procrastinate	-0.54**		
	(0.27)		
Treat × Bedtime regret		0.18	
		(0.14)	
Treat × Work deferred week 1 × Bedtime regret		-0.18*	
		(0.095)	
Expect busy day			-0.11
			(0.56)
Treat × Expect busy day			0.35
			(0.75)
Treat × Work deferred week 1 × Expect busy day			0.39
			(0.33)
Wave controls	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes
Observations	209	209	209

Notes: The table shows results from estimating modified versions of equations 5 and 6 that include interactions with self-reported measures of procrastination and busyness. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Effect of treatment on planned tasks

	(1) Morning tasks week 2	(2) Morning tasks week 2	(3) Evening tasks week 2	(4) Evening tasks week 2
Treat	-0.13	-0.13	-0.45	-0.49
	(0.43)	(0.44)	(0.50)	(0.52)
Work deferred week 1	0.15	0.15	-0.18	-0.18
	(0.11)	(0.11)	(0.15)	(0.15)
Treat × Work deferred week 1	-0.29	-0.30	0.13	0.13
	(0.22)	(0.22)	(0.27)	(0.27)
Wave controls	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	No	Yes
Observations	209	209	209	209

Notes: The table shows results from estimating equation 5, but using planned tasks or executed tasks as the dependent variable. Each column shows the results of a separate regression. All regressions contain the following controls: an indicator for gender, indicators for study wave, age, age squared, GPA, GPA squared, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Finite Mixture Model

	Class 1	Class 2	Class 3	Class 4
Work put off week 2				
Treat	0.98	1.43	-0.08	-0.05
	(1.08)	(1.34)	(0.55)	(0.54)
Work put off week 1	-0.00	0.44***	-0.05	1.04***
	(0.03)	(0.07)	(0.05)	(0.07)
Treat × Work put off week 1	-0.14	-1.91***	1.20***	-1.98***
	(0.12)	(0.21)	(0.16)	(0.21)
Change in commitment demand				
Treat	4.16**	-0.71	0.22	-1.13
	(1.86)	(0.66)	(0.85)	(0.69)
Treat × Work put off week 1	1.13***	-0.31**	0.31	-0.08
	(0.15)	(0.14)	(0.20)	(0.08)
Latent Class Proportion	0.097	0.074	0.418	0.411

Notes: The table shows results from estimating equations 5 and 6 in a seemingly unrelated regression with Gaussian errors using maximum likelihood. Sets of coefficients are estimated for four classes. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Effect of treatment on work deferred, by gender

	(1) Work deferred week 2	(2) Work deferred week 2	(3) Work deferred week 2
Treat	1.08	1.10	1.13
	(0.74)	(0.75)	(0.78)
Work deferred week 1	0.57**	0.57**	0.58**
	(0.24)	(0.24)	(0.24)
Treat × Work deferred week 1	-0.72***	-0.72***	-0.64**
	(0.26)	(0.26)	(0.28)
Female × Treat	-1.12	-1.15	-1.30
	(0.83)	(0.84)	(0.85)
Female × Work deferred week 1	-0.42	-0.42	-0.41
	(0.26)	(0.26)	(0.26)
Female × Treat × Work deferred week 1	0.47	0.48	0.34
	(0.43)	(0.43)	(0.44)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Adj. R^2	0.048	0.039	0.065
Observations	209	209	209

Notes: The table shows results from estimating equation 5. Each column shows the results of a separate regression. All regressions include study wave indicators. Demographic controls are gender, age, and age squared. Additional controls are GPA, GPA squared, self-reported busyness during week 1, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 13: Effect of treatment on commitment demand, by gender

	(1) Change in commitment demand	(2) Change in commitment demand	(3) Change in commitment demand
Treat	-0.47	-0.55	-0.63
	(0.67)	(0.82)	(0.81)
Female × Treat	0.48	0.65	0.61
	(0.63)	(0.95)	(0.94)
Treat × Work deferred week 1	0.54**	0.54**	0.53**
	(0.23)	(0.23)	(0.24)
Female × Treat × Work deferred week 1	-0.18	-0.18	-0.14
	(0.32)	(0.32)	(0.33)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Adj. R^2	0.02	0.007	-0.008
Observations	209	209	209

Notes: The table shows results from estimating equation 6. Each column is a separate regression. Demographic controls are an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. Additional controls are GPA, GPA squared, and self-reported busyness in week 1. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 14: Effect of treatment on work deferred, Winsorized data

	(1) Work deferred week 2	(2) Work deferred week 2	(3) Work deferred week 2
Treat	0.29	0.31	0.23
	(0.32)	(0.32)	(0.33)
Work deferred week 1	0.40***	0.41***	0.43***
	(0.13)	(0.13)	(0.13)
Treat × Work deferred week 1	-0.48**	-0.49**	-0.50**
	(0.22)	(0.22)	(0.22)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Adj. R^2	0.042	0.029	0.053
Observations	209	209	209

Notes: Variables for tasks deferred are Winsorized at the 2.5% level in both weeks. The table shows results from estimating equation 5. Each column shows the results of a separate regression. All regressions include study wave indicators. Demographic controls are gender, age, and age squared. Additional controls are GPA, GPA squared, self-reported busyness during week 1, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 15: Effect of treatment on commitment demand, Winsorized data

	(1) Change in commitment demand	(2) Change in commitment demand	(3) Change in commitment demand
Treat	-0.26	-0.24	-0.32
	(0.37)	(0.39)	(0.40)
Treat × Work deferred week 1	0.41**	0.40**	0.40**
	(0.19)	(0.19)	(0.18)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Adj. R^2	0.018	0.010	-0.0040
Observations	209	209	209

Notes: Variables for commitment demand are Winsorized at the 2.5% level in both weeks. The table shows results from estimating equation 6. Each column is a separate regression. Demographic controls are an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. Additional controls are GPA, GPA squared, and self-reported busyness in week 1. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 16: Assessing experimenter demand using treatment message consistency

	(1) Work deferred week 2	(2) Commitment demand diff.
Treat	0.21	-0.20
	(0.43)	(0.46)
Work deferred week 1	0.33**	
	(0.14)	
Treat × Work deferred week 1	-0.53**	0.42*
	(0.27)	(0.21)
Treat × Bedtime error 1	0.20	0.11
	(0.20)	(0.18)
Treat × Work deferred week 1 × Bedtime error 1	0.20	0.052
	(0.39)	(0.20)
Wave controls	Yes	Yes
Demographic controls	Yes	Yes
Observations	209	209

Notes: The table shows results from estimating equations 5 and 6. "Bedtime error 1" is the bedtime prediction error reported (which was reported to treatment group subjects), divided by its own standard deviation. Controls and sample are the same as in Tables 3 and 4. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 17: Effect of treatment on bedtime forecast error, raw planned bedtimes

	(1) Discrete forecast error week 2	(2) Discrete forecast error week 2
Treat	-0.67**	-0.79**
	(0.31)	(0.33)
Forecast error week 1	1.08**	1.19**
	(0.46)	(0.48)
Treat × Forecast error week 1	-1.22**	-1.31**
	(0.55)	(0.57)
Wave controls	Yes	Yes
Demographic controls	Yes	Yes
Observations	202	202

Notes: Forecast error based on raw self-reported planned bedtime, without adjusting planned bedtimes 10 a.m.–3 p.m. for likely a.m.-p.m. entry error. Each column of the table shows results from a separate regression. All regressions contain the following controls: an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

B Tables Without Multiple Switchers

Table 18: Treatment-control balance without multiple switchers

	Mean/(SD)	Mean/(SD)	Diff./(SE)
Commitment demand week 1	-1.03	-0.24	-0.79
	(3.20)	(3.43)	(0.52)
Commitment demand week 2	-1.63	-0.71	-0.92**
	(2.79)	(3.04)	(0.46)
Work deferred week 1	-0.082	-0.011	-0.071
	(3.15)	(1.92)	(0.40)
Work deferred week 2	0.22	0.56	-0.34
	(2.74)	(2.86)	(0.44)
Bedtime difference from plan (minutes)	42.9	44.6	-1.74
	(67.7)	(53.4)	(9.46)
GPA	3.26	3.32	-0.062
	(0.47)	(0.45)	(0.072)
Female (indicator)	0.58	0.68	-0.11
	(0.50)	(0.47)	(0.076)
Study wave	2.48	2.55	-0.070
	(1.16)	(1.11)	(0.18)
Observations	73	91	

Notes: The significance of the differences is assessed using a t-test. Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 19: Effect of treatment on work deferred without multiple switchers

	(1) Work deferred week 2	(2) Work deferred week 2	(3) Work deferred week 2
Treat	0.32	0.40	0.26
	(0.43)	(0.43)	(0.44)
Work deferred week 1	0.27*	0.28*	0.27*
	(0.15)	(0.15)	(0.15)
Treat × Work deferred week 1	-0.40*	-0.43**	-0.38*
	(0.21)	(0.21)	(0.21)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Adj. R^2	0.0	0.0	0.0
Observations	164	164	164

Notes: The table shows results from estimating equation 5 estimated on the sample that excludes multiple switchers. Each column shows the results of a separate regression. Controls are indicated at the bottom of each regression. Demographic controls are an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. Additional controls are GPA, GPA squared, and self-reported busyness during week 1. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 20: Effect of treatment and task procrastination on commitment demand without multiple switchers

	(1) Change in commitment demand	(2) Change in commitment demand	(3) Change in commitment demand
Treat	0.14	0.13	-0.081
	(0.47)	(0.47)	(0.45)
Treat $\times$ Work deferred week 1	0.47***	0.48***	0.47**
	(0.18)	(0.18)	(0.18)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Adj. R^2	0.03	0.02	-0.0006
Observations	164	164	164

Notes: The table shows results from estimating equation 6 estimated on the sample that excludes multiple switchers. Each column is a separate regression. Demographic controls are an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. Additional controls are GPA, GPA squared, and self-reported busyness in week 1. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 21: Effect of treatment on planned tasks without multiple switchers

	(1) Morning tasks week 2	(2) Morning tasks week 2	(3) Evening tasks week 2	(4) Evening tasks week 2
Treat	0.13	0.15	-0.18	-0.24
	(0.50)	(0.51)	(0.59)	(0.59)
Work deferred week 1	0.16	0.16	-0.11	-0.11
	(0.11)	(0.12)	(0.16)	(0.16)
Treat × Work deferred week 1	-0.26	-0.26	0.14	0.17
	(0.24)	(0.25)	(0.28)	(0.29)
Wave controls	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	No	Yes
Observations	164	164	164	164

Notes: The table shows results from estimating equation 5, but using planned tasks or executed tasks as the dependent variable, on the sample without multiple switchers. Each column shows the results of a separate regression. All regressions contain the following controls: an indicator for gender, indicators for study wave, age, age squared, GPA, GPA squared, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 22: Effect of treatment on bedtime forecast error without multiple switchers

	(1) Discrete forecast error week 2	(2) Discrete forecast error week 2	(3) Discrete forecast error week 2
Treat	-0.61*	-0.68*	-0.89**
	(0.35)	(0.36)	(0.41)
Forecast error (normalized)	0.21	0.16	0.21
	(0.28)	(0.27)	(0.29)
Treat × Forecast error (normalized)	-0.48	-0.46	-0.51
	(0.36)	(0.35)	(0.38)
Wave controls	Yes	Yes	Yes
Demographic controls	No	Yes	Yes
Additional controls	No	No	Yes
Observations	158	158	158

Notes: Each column of the table shows results from a separate regression estimated on the sample that excludes multiple switchers. All regressions contain the following controls: an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.

Table 23: Time use and commitment price without multiple switchers

	Class	Exercising	Other	Sleeping
Time Shock diff.	-0.89	0.53	4.49*	-1.17
	(1.44)	(0.57)	(2.54)	(1.52)
Observations	164	164	164	164

	Socializing	Studying	TV	Working
Time Shock diff.	2.46	-6.70**	2.73**	-1.45
	(2.30)	(2.68)	(1.36)	(0.91)
Observations	164	164	164	164

Notes: The table shows results from estimating equation 7, without multiple switchers. Each column is a separate regression. All regressions contain the following controls: an indicator for gender, indicators for study wave, age, age squared, and an indicator for whether the subject was employed at the time of the study. In parentheses are heteroskedasticity-robust standard errors (White, 1980). Significance indicated by: *** p<0.01, ** p<0.05, * p<0.1.