An Experiment on Time Preference and Misprediction in Unpleasant Tasks*

Ned Augenblick[†] and Matthew Rabin[‡]

Forthcoming, Review of Economic Studies

March 26, 2018

Abstract

We experimentally investigate the time-inconsistent taste for immediate gratification and future-preference misprediction. Across seven weeks, 100 participants choose the number of unpleasant transcription tasks given various wages to complete immediately and at different future dates. Participants preferred 10-12% fewer tasks in the present compared to any future date, leading to an estimated β of 0.83. Comparing predictions with actual immediate-work choices provides evidence against substantial sophistication, with estimates implying that participants understand no more than 24% of their present bias. Finally, we find evidence of "projection bias": participants wished to complete 4-12% fewer tasks when decisions were elicited right after completing tasks rather than before.

*We have benefited from discussions with Aaron Bodoh-Creed, Stefano DellaVigna, David Laibson, Tarso Mori Madeira, Ted O'Donoghue, Frank Schilbach, Joshua Schwartzstein, Avner Shlain, Charles Spenger, Alex Steiny, and Séverine Toussaert, as well as seminar audiences at Universities of California Berkeley and Santa Barbara. Tarso Mori Madeira, Dan Svirsky, Wei Wu provided wonderful research assistance. This research was approved by the UC Berkeley IRB.

†Haas School of Business, University of California, Berkeley. ned@haas.berkeley.edu

‡Department of Economics and Business School, Harvard University. matthewrabin@fas.harvard.edu

1 Introduction

In this paper, we report findings from an experiment investigating the presence and scale of the time-inconsistent taste for immediate gratification, the self-awareness of this taste, and the effects of contemporaneous work burden on attitudes towards later work. The experiment involved choices and predictions about performing the unpleasant task of transcribing blurry foreign letters: over the course of seven weeks, participants specified the number of tasks they were willing to complete for different piece-rate wages on different days. Choices were elicited from participants for both immediate work and future work, and both right before and right after exerting their most recent effort on the tasks. Participants were also asked to make incentivized predictions about their future willingness to work.

Participants choose 10-12% less immediate work than future work. They appear to have no time preference among future dates between four and 30 days away (the minimum and maximum delays permitted in the experiment for future work), so that we estimate the exponential daily discounting parameter δ consistently very close to 1. Using different approaches and specifications, we generally estimate the standard now-vs.-future present-bias parameter β to be between .81 and .84. Participants seemed relatively naive about their self-control problems: the predictions they make of how much immediate work they will choose in the future largely match their current preferences about how much work they want to do in the future. We estimate the present-bias sophistication parameter $\widehat{\beta}$ from O'Donoghue and Rabin (1999a, 2001) – which we label β_h to distinguish it from the empirical estimate of β – to be very close to 1, but a positive correlation between participants' severity of present bias and their predictions suggests they understood 10-24% of their present bias. These conclusions about sophistication are complicated by participants' unexpectedly strong preference to behave consistently with their earlier predictions. Participants also appeared to exhibit "projection bias" as defined in Loewenstein, O'Donoghue, and Rabin (2003): they commit to and predict about 4-12% more work when asked before having exerted a small amount of effort than when asked after, when their current distaste for the task was presumably higher.

Section 2 presents our experimental design. We recruited 100 UC Berkeley Xlab participants, who were required to participate for seven *dates* over a subsequent six-week period: the first in the experimental laboratory, and the following six outside the laboratory using an online interface. Each date, participants chose how many unpleasant transcription tasks (between 0 and 100) they wished to complete for five randomly chosen piece-rate wages (between \$0.01 and \$0.31) in the present and on a selection of future dates. On each date after the initial one, participants were also required to complete the tasks from one previous decision that was chosen at random from questions answered on past dates or immediately before. Participants were also asked to make predictions about their future task decisions and were rewarded with bonuses if their bonuses were accurate within five tasks. Finally, participants were required to complete ten mandatory tasks each date regardless of decisions, which removed any fixed costs associated with completing a non-zero number of tasks and gave participants experience with the task. From here on in, we refer to non-mandatory tasks simply as "tasks". Participants who completed the entire experiment received \$50 on top of the average of \$60 performance-based earnings, all paid on a fixed date at the end of the experimental period. If a participant did not complete the required components on a date, she was immediately removed from the experiment, receiving previously-made earnings at the same fixed date, but not the \$50 completion payment. All participants, whether they completed the full 6 weeks or not, were paid their earnings one week after the full 6-week experiment ended.

We outline the main model and empirical strategy in Section 3. Although all estimations in the paper control for and include estimations for projection bias, both to streamline the presentation and because of its conceptual independence from present bias, we separately discuss the design and results for projection bias in Section 6. For time discounting, we use the standard two-parameter model of present bias of Laibson (1997), where immediate gratification β and long-run discounting δ are incorporated into intertemporal utility $U^t = u_t + \sum_{\tau=t+1}^{\infty} \beta\delta^{\tau} u_{\tau}$. We model sophistication about future present bias with the parameter $\beta_h \in [\beta, 1]$ whereby the person believes she will discount by β_h in the future; when β_h is close to β, the agent is relatively sophisticated about her self-control problems. The experiment is designed to identify each of these parameters based on willingness to work at different wages from different perspectives. Given a convex effort-cost curve, the participant will choose the number of tasks that equalizes the perceived marginal disutility of effort against the later monetary payment. Present bias can be identified by systematic differences in committed decisions made for immediate work vs. future work. If we give people a *small* reward for accurate predictions of future work, β_h can be identified from comparing predictions about future immediate-work choices to actual immediate-work choices. By asking these questions for multiple wages, it is possible to simultaneously estimate effort-cost curves and the parameters β, δ, and β_h.

An important feature of our design is the variation in bonuses for accurate predictions, ranging from \$0.25 to \$8.25. While very small prediction-accuracy bonuses should in principle reveal participants' true beliefs without distortion, one might worry that they provide little incentive for accuracy. Yet large payments introduce at least two potential difficulties to estimating β and β_h. The first is obvious: the incentive to match their immediate-work decisions to earlier predictions to receive the bonus complicates the estimation of β. To get estimates independent of this complication, 83% of immediate-work decisions were in situations for which they had not previously made a prediction. More challenging is estimating β_h: when the bonus is large, sophisticated present-biased participants recognize that prediction-accuracy payments can be used as partial commitments, causing the would-be prediction to be higher than the "straightforward prediction" given no bonus payment. Furthermore, because bonuses were given if chosen tasks were within five of predicted tasks, the person would choose five tasks above this desired amount. By assuming that participants treat prediction-accuracy bonuses in the same way as wages, we structurally estimate β_h based on how participants used the predictions for both accuracy and as soft commitment devices. However, we suspected, based on intuition outside of formal models of present bias and commitment, that even people who understand their future present bias might make straightforward predictions that do not optimally exploit the potential for commitment. Consequently, after testing for this possibility and showing evidence for it, we also present results as if people make straightforward predictions.[1,2]

We analyze our results using simple reduced-form tests in Section 4 and (with very similar results) using a structural model in Section 5. In both sections, we vary the set of parameters that are restricted to be common across individuals, starting with a fully "aggregate" analysis and culminating in an "individual" analysis, in which all parameters are allowed to vary across participants. The individual analysis is very demanding, and allows estimates for only 72 of the 100 participants. As an example of estimation difficulties, ten participants were impossible to study on an individual level because they rarely or never varied their choices as a function of wages. Six of these ten either consistently chose 0 or consistently chose 100 (with occasional other choices), which might mean that their disutility of effort is very low or very high; but four of the ten consistently chose the same effort between these extremes, which is hard to reconcile with a plausible cost-curve.[3] For consistency across analyses, we use the same sample of 72 participants throughout the main paper. In Appendix 9.8, we use the full sample of 100 participants to replicate all of the main figures and analyses that do not require estimation of individual parameters, leading to very similar results.

Section 4 provides reduced-form evidence of our main findings. There is (strongly statistically significant) present bias, with participants choosing on average 5.7 fewer immediate tasks than future tasks (43.5 vs. 49.1), 5.3 fewer when controlling for fixed effects for each wage, and 4.9 fewer when focusing only on decisions from the third date on, where the data suggest

[1] This design feature and resulting estimation also help alleviate our concern that our structural estimation would push some functional-form assumptions too far. In the event, our results were similar using the basic identification and the bonus-variation estimation techniques.

[2] Because our goal was to see if sophistication could be identified separately from willingness to pay for commitment, we designed the experiment to *mask* the commitment potential of predictions from participants. In doing so our design clearly does not provide a generalizable test of people's willingness to use commitment devices.

[3] We removed participants from analysis if the maximum-likelihood routine could not converge or, in two cases, led to very extreme estimates. In addition to those with little variation discussed in the text, some participants left the experiment too early to produce enough data for estimation, while others made decisions that are difficult to rationalize with any model of rational or irrational behavior that we are familiar with.

there is no more learning about the unpleasantness of the task. While the experiment does not include future decisions that are closer than four days from work, the very small, insignificant, and unpatterned discounting beyond four days provides support for a "quasi-hyperbolic" discounting over these time frames. There is evidence for near-complete naivete in the aggregate: participants predict a (statistically insignificant) average of 0.7 fewer immediate-work choices in the future than their current future-work choices, and a (statistically insignificant) average of 0.2 *more* immediate-work tasks when analyzed using wage fixed effects or focusing on later dates. Although there is significant variation in the data arising from facing different wages in different decision sets, different constraints on different days, and changing information about these constraints, we find similar patterns on an individual level: 76% of the time individuals choose less immediate than future work while 53% of the time they make lower predictions than their future-work choices. We find a significant positive correlation (0.246) between the present-biased taste for fewer immediate tasks and beliefs about such a taste in the future, suggesting some level of sophistication.

Section 5 reports on the structural identification of the parameters β, δ, and β_h by assuming a power cost function, a linear monetary-utility function, and a normally distributed error term added to the effort decision. These findings largely mirror the reduced-form findings. Although we report a large array of regressions, given different fixed effects, including and excluding the possibility that predictions were used as commitment devices, and focusing on later decisions, our primary conclusion is quite stable: the null of $\beta = 1$ is always strongly rejected, with the estimate of the present-bias parameter β ranging from 0.81 to 0.84 in the aggregate analysis, and the average estimate of β ranging from 0.79 to 0.87 in the individual analysis.[4] Aggregate estimates of β_h range from 1.00 to 1.01, and average individual estimates range from .98 to .99.[5] In each case, the null hypothesis of no perceived present bias ($\beta_h = 1$) cannot be rejected, and full awareness of the present bias ($\beta_h = \beta$) is strongly rejected. But we again find significant correlation between individual estimates of β and β_h indicating some degree of sophistication: regressions of estimated perceptions about present bias $1 - \beta_h$ on estimates of true present bias $1 - \beta$ yields positive and significant coefficients ranging from 0.10 to 0.24 across seven specifications including various robust regressions and attenuation-bias corrections.

In Section 6, we report on all features of the design and results pertaining to projection bias, whereby a person tends to project current tastes onto future states where tastes will differ. To

[4] In five specifications, we find a daily discount factor δ of between 1.003-1.005, rejecting the null hypothesis of $\delta = 1$ in one case with marginal rejection in two. The daily discount rates imply a weekly discount factor of 1.02-1.03, suggesting they would care on the order of 400% more about utility delivered at some point than a year earlier. Our ex-post intuition of the positive discount factor is that, due to uncertainty, participants are less willing to commit to large amounts of work farther in the future; it also of course hints at mispecification. As we discussed and estimate in in Appendix Section 9.3, uncertainty would bias our estimate of β *upward*.

[5] We estimate $\beta < 1$ for 69-78% and $\beta_h < 1$ for 54-60% of participants.

investigate projection bias, we varied whether decisions were made before or after mandatory tasks, positing that the distaste for the tasks would be higher immediately following these tasks. Participants chose an average of between 2.1 and 3.3 fewer tasks after completing the mandatory work than before, depending on the chosen controls. Although the differences in these basic analyses are not statistically significant, the estimates gain significance and increase to between 3.8 and 5.8 when we account for the censored nature of the data. Finally, the estimate of the difference in the structural model increases further to between 4.1 and 7.3, with consistently high statistical significance.

We discuss related literature in detail at the start of Section 7. There are perhaps four design components that distinguish our study from most of the other literature on present bias and sophistication. First, following a few recent papers, we attempt to better account with the theory of time discounting, which concerns tradeoffs of *precisely-timed utility* flows. We attempt to control for timing and the relationship between outcomes and utility by (1) focusing on the completion of time-specific non-fungible tasks that try to minimize any effects on future utility except the money earned, and (2) eliciting task preferences across many wages to estimate precisely the mapping from the number of tasks into that period's disutility-of-effort.[6] This is best contrasted with the traditional practice of identifying time preferences from choices over time-dated monetary payments. Given that money is fungible, inferences about time discounting are complicated because there is little reason to expect a tight connection between the timing of monetary payments and the timing of utility derived from consumption purchased with the money.[7]

Second, we believe our experimental design better avoids the concern, voiced by many authors, that the choices suggesting present bias in standard "money-now-or-later" experiments might actually be driven by either simplifying heuristics employed by participants, or to other failures to maximize utility, such as proportional thinking or sub-additivity.[8] The influence of heuristics (applied to either money or effort) in our experiment seems unlikely, given that we repeatedly present multiple tradeoffs over time between money and effort in a way that is less conducive to the sort of arithmetic comparisons that might be at play in these experiments.[9]

[6]Augenblick, Niederle, and Spenger (2015) also focus on choices to complete similar time-dated tasks, but use a different methodology – the convex-time-budget methodology developed by Andreoni and Spenger (2012) – to estimate the mapping from task completion to disutility.

[7]More recent studies in the monetary domain, such as Andersen et al. (2008) and Andreoni and Spenger (2012), attempt to measure and control for curvature in the monetary utility function, but do not address the fungibility issue. For discussion of this issue, see Cubitt and Read (2007), Chabris, Laibson and Schuldt (2008), Augenblick, Niederle, and Sprenger (2015), and O'Donoghue and Rabin (2015).

[8]These choice heuristics are reviewed in Ericson et al (2015), who also introduce the intertemporal choice heuristic (ITCH). Some of the more well-known heuristics include subadditive discounting (Read, 2001), similarity effects (Rubinstein, 1988; Rubinstein, 2003), and the DRIFT model (Read, Frederick & Scholten, 2013).

[9]Another design choice—forcing participants to always show up and express work preferences on each date prior to completing work—avoids two additional potential confounds. Ericson (2016) notes that present bias will

Third, we use participants' predictions of future behavior, rather than their taste for commitment, to measure sophistication. This helps with precision, particularly given our use of a large number of predictions at different accuracy incentives and different wages, and the ability to observe how these evolve over time as the person gains experience. It is difficult to imagine tightly identifying β_h from commitment choices, at least with the standard practice of binary choice of whether to commit. In such cases, finding a desire to commit implies *some* degree of sophistication, while aversion to commitment does not imply zero sophistication, given the benefit of flexibility in the face of uncertainty. Although the use of predictions has its own limitations, our results suggest that accuracy incentives seem not to affect predictions. Our results suggest that future studies of sophistication might simply provide varied accuracy payments for predictions, verify that the predictions do not change with the magnitude of the payment, and then treat the predictions as straightforward.

Fourth, testing choices between (socially useless) effort and money lets us avoid one concern that applies in many domains where present bias is studied: that the choices are laden with normative meaning to participants that interferes with interpreting the results. Experimentally observing levels of (and predictions about) activities such as savings, eating, exercise, or drug use, creates a potential for participants to be influenced by their own or others' judgment about character. In our experiment, it is more difficult to imagine that participants start the experiment with a view of the "appropriate" number of tasks to complete for a given wage. Furthermore, randomly varying the wage across different decisions makes it difficult for subjects to use past decisions as a normative anchor.[10] The use of predictions rather than commitments likely also has advantages in inoculating the measurement of sophistication from similar concerns: a participant may view commitment decisions as a reflection of her character, either as an explicit recognition that she has self-control problems or as a signal of the desire to engage in undesirable activities. Even the direction of the bias from commitment is unclear because people may commit readily to indicate they have no interest in bad behavior, or they may avoid commitment to signal that they have the willpower to resist bad behavior.

Some previous studies (especially Augenblick, Niederle, and Sprenger (2015)) share some of these features. At the start of Section 7, we discuss in detail the relatively few papers that share some of these design choices, such as Read and van Leeuwen (1998), Badger *et al.* (2007), McClure *et al.* (2007), Brown *et al.* (2009), Acland and Levy (2015), and Kaur, Kremer, and

be incorrectly estimated if no-show decisions are mistaken as a statement of preference rather than forgetting; and Kaufmann (2017) suggests that projection bias can be mistaken as present bias if future task planning is usually made outside of the workday (a "cold" state) but immediate-work decisions are made by dynamically choosing to stop working during the completion of tasks (a "hot" state).

[10] The fact that decisions are made outside the lab (presumably at home) likely further reduces the sense of experimenter oversight. Of course, insofar as internal or external judgment *does* influence behavior in most domains of interest, excluding such forces also may miss an important aspect of interest.

Mullainathan (2015). The most dramatic departure in terms of experimental design is probably the use of predictions to identify and analyze sophistication. In terms of measuring projection bias, our largest departure is the use of experimental variation in a domain lacking some of the more “visceral” explanations for situation-dependent variation in choices, such as hunger, drug craving, and sexual compulsion.

Although our methods may differ, our conclusions for the most part sit comfortably with previous findings. As such, we see our main contribution as simply providing additional evidence measuring present bias, sophistication, and projection bias, using methods that address some of the major concerns about past experiments. Our finding of near-complete naivete seems consistent with, but more clearly identified than, the previous literature, which commonly finds little willingness to pay for commitment, with the significant exception of Schilbach (2017)’s finding of sizable commitment demand by habitual alcohol abusers in India. Our study also provides some of the first measured evidence showing clear present bias over the near term, but none at all from the medium-term (here, around 1-4 weeks). This form of present bias (as opposed to smoothly decreasing impatience reflecting hyperbolic discounting) captures both a widespread intuition and a common theoretical assumption that has not been cleanly identified. Augenblick (2017) provides even more fine-tuned evidence over shorter periods of time, finding a drop – similar in magnitude to that in our study – in the desire to complete tasks over a week as the work time approaches, with two-thirds of the drop occuring a day from work and one-third occuring a few hours from work.

At the end of Section 7, we discuss some of the potential problems with our experiment, and what those problems suggest about potential follow-on studies. While we aimed to find a source of utility that is minimally fungible across time, some of the disutility of effort may reflect the opportunity cost of other tasks that bring both present and future benefits. In this case, the actual utility consequences from tasks would be spread out over time rather than all occurring immediately, leading us to underestimate present bias. Better calendar isolation of utility might be possible, but difficult, especially for experiments conducted over horizons like ours and outside the lab. In addition, participants presumably face large degrees of day-to-day uncertainty in their effort costs. We argue in Appendix 9.3 that allowing for uncertainty (that participants take into account) will also bias the estimation of the severity of present bias downward. But the direction of misestimation can change if participants systematically underpredict their expected costs—due, for example, to a general planning fallacy. And the effects of uncertainty on the estimation of sophistication (and projection bias) is ambiguous. Our conclusions about sophistication are also clouded by unexpected behavior in the experiment: participants who were reminded of earlier predictions exhibited a strong motivation to behave consistently with those predictions — either exactly matching the prediction or performing within five tasks — in a way that cannot be explained by their pecuniary incentives.[11] This complicates the interpretation of sophistication as, in principle, participants who predicted this consistency taste may have been employing commitment devices after all. But we delineate various reasons why we think such sophisticated interpretation is unlikely.

Appendix Sections 9.7-9.17 report on various robustness checks of our primary analyses, including the use of different subsamples of the data, different samples of participants, and many alternative econometric specifications. These are referenced throughout the paper when relevant.

2 Experimental Design

Our study examines participants' work decisions about the completion of unpleasant transcription tasks over time. The experiment took place over seven *participation dates* across six weeks, with the first date occurring in the laboratory and the later dates requiring logging into a website accessible from any computer. Each date, participants stated work preferences, made predictions, and completed mandatory work and agreed-to tasks based on past decisions. Section 2.1 describes the task in more detail, Section 2.2 discusses choice of dates and payment details, Section 2.3 walks through each of the components of participation dates, Section 2.4 describes our randomization procedures, and Section 2.5 describes our participant sample, including information about attrition.

2.1 Experimental Task

Each task consisted of a simple transcription of 35 blurry Greek letters through a computer interface, using the mouse to point and click on the corresponding letters. The top panel of Figure 1 is a screenshot from the interface. In order to maintain the full attention of the participant throughout the task, an auditory "beep" sounded randomly every 5-15 seconds throughout the transcription process, and the participant was required to press a button at the bottom left of the screen after hearing this noise. If the participant did not press the button within five seconds of the beeping noise, or pressed it when there was no beeping noise, all work for the current transcription was erased.[12] For a transcription to be accepted as completed, it was required to be 80% accurate (defined as requiring fewer than 7 insertions, deletions, or

[11] Indeed, the patterns we observe are hard to explain even by simple dissonance stories we might have guessed; reminding a participant of their predictions for close-by wages, which might be expected to arouse participants' sense that they are being inconsistent or misbehaving, seemed not to influence behavior.

[12] We are confident we succeeded in making the task unpleasant.

character changes to match the target text).[13] If the transcription was not accurate enough, the participant was informed and could immediately correct the text.

2.2 Participation Dates and Payment Logistics

When participants were recruited, they were told that the experiment required six additional days of participation over the ensuing six weeks, and were asked to choose these dates prior to the first day of the experiment (with no restrictions on the time-of-day of participation). The dates were required to be between four and ten days apart, with the last date required to be completed within six weeks of the start of the experiment. Participants were allowed (for a fee of $.25) to modify the dates over the course of the experiment. We allowed this modification to insure participants against large scheduling shocks, such as learning that a midterm occurred on a previously chosen date. To allow for such date changes without permitting last-minute procrastination that would have confounded our results, participants were allowed to modify a date only up to 5 p.m. on the prior day.[14]

Participants were paid $50 for completing the minimum requirements of the experiment and paid additional amounts depending on their choices. If participants did not complete all required components for some date, they were immediately removed from the study and forfeited the $50 completion payment, but still received any previous earnings. Participants, including those who were removed, received *all* payments associated with the experiment by check exactly seven weeks from the start of the experiment. This later payment date was chosen to avoid any present bias associated with positive consumption from the payment. Average earnings were about $110 for those who completed the entire experiment.

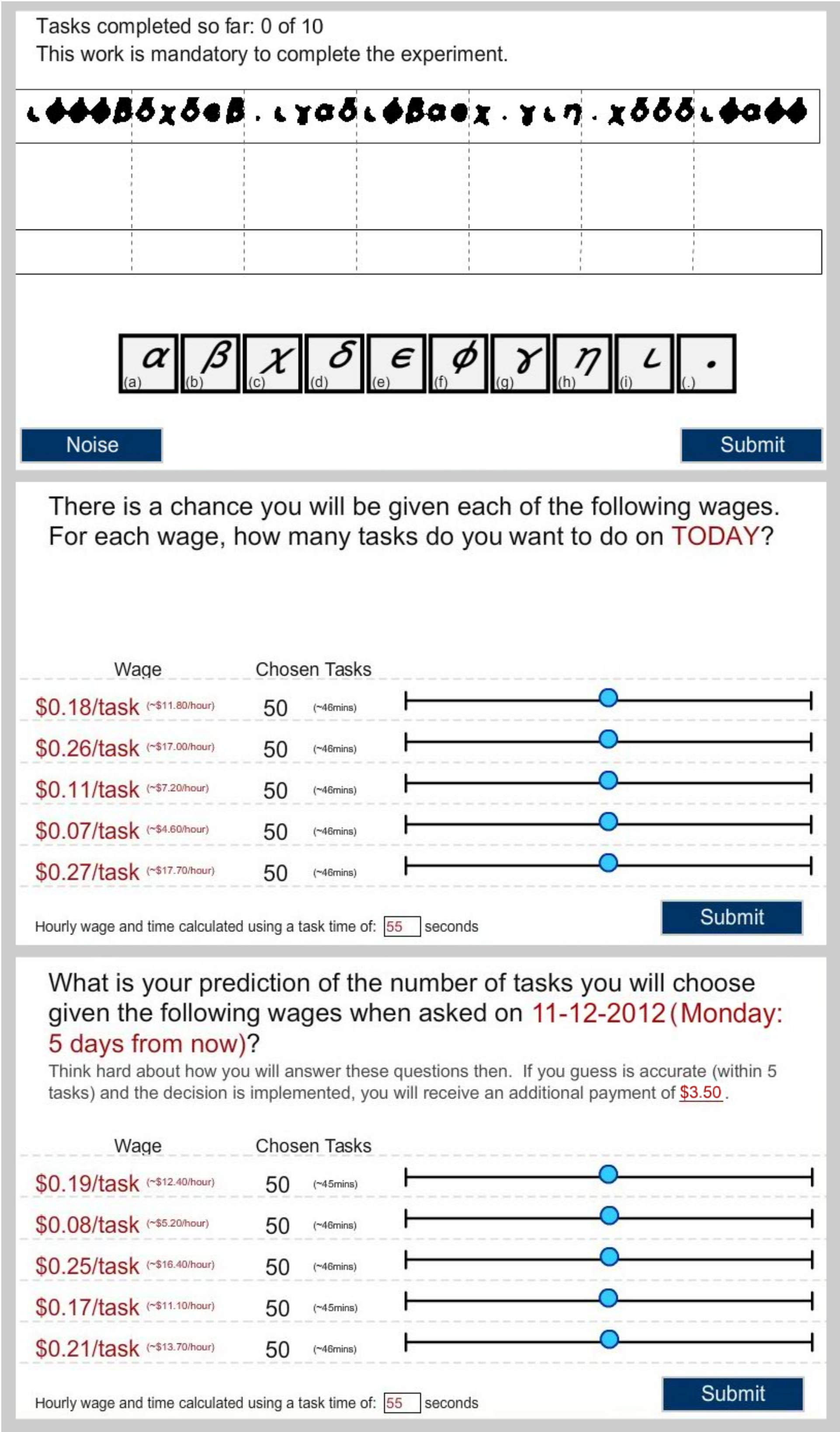
2.3 Experimental Components

Each participation date involved a set of six experimental components that we describe in detail below: completing mandatory work, stating preferences over future work, stating preferences over immediate work, stating predictions about future decisions, observing the one supplemental work decision that would be implemented for that date, and completing supplemental work in that decision. All of these components were completed on each date except for the initial

[13] Although we did not record the number of (failed) submitted transcriptions which did not achieve 80% accuracy on first attempt, more than 95% of transcriptions were more than 90% accurate when accepted, suggesting that participants were not minimizing effort to tightly match the accuracy cutoff.

[14] 12 participants modified dates once and 2 did so more than once. Ex ante, we did not intend to analyze anything about participants' choices and modifications of dates. Ex post, we found no evidence of differences in choices between participants who did and did not modify dates, but we are very reluctant to draw any conclusions given our lack of statistical power.

Figure 1: Screenshots of the transcription task (top), the decision interface for decisions about present work (middle) and predictions about future work (bottom).



laboratory visit (when there was not enough time to complete non-mandatory tasks) and the final date (when we could not ask about future tasks).

On the first day of the experiment, participants were given instructions about the entire experiment prior to making any decisions, including the logistics for all of the components of the experiment and payment. Participants were informed of all points of randomization—including payments, supplemental work determination, and ordering of components—although in cases where the exact distributions were complicated (discussed in the next subsection) they were not told the distributions in full detail. We explicitly told participants that all randomizations had been previously chosen by a random number generator and therefore could not be affected by any of their decisions.

After the first day, participants received reminder emails with a link to the experimental website both on the night prior to and the middle of each date. Once they clicked on the experimental link, they saw an experimental timeline with all components of the experiment to be completed during that day. While the instructions were always presented first, the ordering of the rest of the components for online dates were determined by block randomization, which allows us to test for ordering effects. After each component was completed, participants were again reminded of the timeline of the experimental day.

2.3.1 Completion of Mandatory Work

Each date, participants were required to complete mandatory work. These mandatory tasks (1) gave the participants experience with the task, (2) required that the participants allocated at least ten minutes to each date, eliminating any fixed cost associated with completing some rather than none of the tasks, and (3) allowed us to study whether, per projection bias, contemporaneous distaste for the task affected people's willingness to do the task in the future.

2.3.2 Decision Type 1: Current Work Decisions

Each date, participants were asked a sequence of questions concerning preferences about completing additional supplementary transcription tasks for five different wages on the present date. The middle panel of Figure 1 shows the computer interface used to elicit preferences for immediate work. For example, on the first line, a participant is asked for the number of tasks she would like to complete immediately if given a wage of $.18/task, using the slider bar to choose a number between 0 and 100.[15] Participants were required to make a decision for each wage.

[15] In order to minimize the extent to which the interface pointed a participant towards a default work decision, all slider points started in the middle position, slightly elevated from the slider bar with no assigned number. Once the slider point was clicked, it dropped to the slider bar and could be dragged.

The five wages in each decision set were all between \$0.01/task and \$0.31/task. This implies hourly wages of between \$0.80/hour and \$24.80/hour given the average empirical completion rate of 45 seconds per task. The wide range of wages was used to induce enough variation in participants' responses for estimation of effort parameters. To ensure that participants paid full attention to the exact wage in each decision, and (for better or worse) to avoid encouraging consistency across wages, the order of the five wage decisions was random. The interface displayed hourly wage estimates and time-to-completion estimates using a default task completion time of 55 seconds, with the option to enter different completion times to generate different estimates.[16] Each of the decisions had the potential to be randomly chosen as the one *decision-that-counts*, in which case the participants were required to complete the stated work for the specific wage on the relevant date.

2.3.3 Decision Type 2: Future Work Decisions

Except on the last participation date, participants were also asked questions about five wages for work on future dates. On each date, they were asked about at most two future dates. The interface for future-work decisions was identical to that for present-work decisions, with the red part (saying "today") of the initial text replaced with the relevant future date.

2.3.4 Decision Type 3: Predictions of Future Work Decisions

In addition to making decisions about future work, participants were also asked to make predictions about potential future immediate-work decisions given five randomly chosen wages. For example, on November 12th, a participant might have been asked to predict the number of tasks that they thought they would choose on November 18th when facing a wage of \$0.15. For each set of five wages, the participants were presented with a bonus that they would receive on the future date if the given wage appeared in an immediate work decision chosen as the decision-that-counts and they choose a number of tasks within five tasks of their prediction. The bonuses were randomly chosen from 14 bonus payments from \$0.25-\$8.25.[17] The bonuses were varied to later identify and control for the use of the accuracy payment to incentivize future decisions. The computer interface that elicited these decisions is shown in the bottom panel of Figure 1.

As participants were aware, if they faced a work decision later for which they had previously made a prediction, they were reminded of the prediction with a visual cue on the work-decision slider bar. Although the choice to provide prediction reminders potentially anchored the par-

[16] The difference between the displayed per-task completion time on the interface (55 seconds) and the actual empirical completion rate (45 seconds) represents our imperfect ex-ante prediction of this rate.

[17] The potential bonuses were .25, .40, .65, .85, 1.00, 1.25, 1.75, 2.25, 2.75, 3.25, 4.25, 5.25, 6.75, and 8.25.

ticipants on their prediction, we felt that not reminding them would create a potentially severe identification problem, since we could not know how likely it was that those participants interested in using predictions as commitments thought they would remember their predictions.

2.3.5 The "Decision-That-Counts" and Work Decisions

Participants were asked questions about their preferred number of tasks to complete on future dates and the current date given different wages. Therefore, at the time of determining the one decision-that-counts for a given date, participants would have made many past and present decisions about work on that date.[18] These decisions were all collected and displayed to participants. Then, one decision was randomly chosen as the decision-that-counts.

Once the decision-that-counts was chosen, the participant was required to complete the exact number of supplementary tasks in that decision for the associated wage. For example, if the decision-that-counts involved a wage of $0.18/task, and the participant previously chose to complete 40 tasks for that wage, the participant was then required to complete exactly 40 supplementary tasks for a supplementary payment of $7.20. If a participant did not complete these tasks, she was immediately removed from the experiment and forfeited the $50 completion payment.

2.4 Randomization

The order, number, and types of decisions varied across dates and between participants. There are a variety of ways in which the ordering and randomization deviated from purely independent uniform draws. First, while the five wages in each decision set were drawn from a uniform distribution across possible wages, we constrained sets to contain at least one wage below $.15 or above $.20. This was done to reduce the number of decision sets with too little variation in decisions to identify effort costs. Second, wages for work decisions were chosen to match either all, two, or none of the wages from previous prediction sets, when prediction and decision concerned the same date. This was done to ensure that participants' predictions would potentially affect their payoff, while also allowing for immediate work decisions that were uninfluenced by past predictions. On average, 11% of predictions were for wages that later appeared for actual decisions. Third, to ensure enough variation to identify the effects of bonuses, no participant received the same bonus for different prediction sets. Finally, to ensure that participants' predictions and future work decisions were evenly spread across future dates, the relevant date of participant's predictions or decisions were not randomly chosen. For example, on date two,

[18] Recall that participants could modify the timing of participation dates. Participants knew that if a date was modified, then the decisions about that partictipation date were transfered to the new day.

participants either made decisions about future work on dates 3 and 5 or decisions about future work on dates 4 and 5. No participant made a work decision for more than three dates in the future or for a prediction decision more than four dates in the future. On the first and sixth dates, participants made one set of future work decisions and one set of future predictions. From the second to the fifth date, participants made two sets of future work decisions and two sets of future predictions.

2.5 Sample

100 participants from the UC Berkeley Xlab subject pool were recruited into the experiment across four experimental sessions on October 17-19, 2012.[19] 79 completed all seven weeks of the experiment and received the $50 completion payment.[20] Participants who completed the experiment made 130 decisions each. In total, participants made 11,405 decisions: 2,750 about immediate work, 4,235 about future work, and 4,420 predictions.

In our analysis, we report a standard parametric aggregate estimation which assumes that all participants have the same time-discounting parameters. Additionally, we conduct a more demanding exercise in which all of the parameters are estimated for each participant. Given the demanding nature, this individual estimation fails for a significant number of participants. Some participants' behavior interferes with estimation due to lack of variation in decisions. For example, three participants completed the experiment but did not vary their task decisions once across their 130 decisions, always choosing either zero or one hundred tasks. These participants likely have extreme costs parameters which makes it impossible to draw any conclusions about their time discounting functions in our experiment.[21] Similarly, estimation is difficult for some participants with few observations due to attrition, although we are still able to estimate parameters for 15 of the attritors. As in any experiment, some participants made decisions that are difficult to rationalize with any standard economic theory. For example, six participants had non-monotonicities in more than a third of their decision sets. Other participants have some combination of these three issues. Finally, the irregular decisions of two participants lead the maximum likelihood routine to converge with estimated parameters that are very clear outliers from all other participants.[22] In total, we cannot estimate individual parameters for

[19] We placed no restriction on participation except that we disallowed those who were subjects in Augenblick, Niederle, and Sprenger (2015).

[20] Of the 21 participants that did not complete the experiment, 7, 7, 1, 1, 4, and 1 participant(s) dropped out on date 2, 3, 4, 5, 6, and 7, respectively.

[21] It might be tempting to conclude that a participant who always chose 0 tasks or the maximum 100 tasks regardless of wages is inherently time consistent. But of course these participants may be more willing to work in the future than present, but either perform 0 tasks because no wage is high enough to induce positive future work or because no wage is low enough to induce fewer than 100 immediate tasks.

[22] For example, the estimates of the sophistication-about-present-bias parameter β_h for these participants are

28 participants. We focus on the remaining 72 participants throughout the paper so that the sample is consistent across the individual and aggregate estimations.

Importantly, all of the non-parametric figures in the paper can be reproduced using all of the participants with little change, as shown in Appendix 9.8. Furthermore, the aggregate structural analysis can be performed using the entire sample when the cost curve is not individually estimated for the 28 dropped participants, leading to broadly similar parameter estimates (with a slightly lower estimate of β).

Appendices 9.8-9.15 contain further robustness checks on specification choices, participants sample, and choices sample, as well as information about the attritors. We first replicate our main aggregate estimation with different samples, such as the entire sample minus attritors. Similarly, we run the analysis given three different choice samples—such as removing decisions that were very similar to recent decisions—and under a variety of alternative specifications using different assumptions about the form of decision error. All of these robustness checks support the qualitative conclusions from the main text.

The average completion time for a single transcription task for was 56 seconds for the 1st of the seven dates, 48 for the 2nd, declining to 44 and 43 for the 3rd and 4th, and then stabilizing at 42 seconds for each of the last three dates. Because this suggests that participant's effort costs might be changing over time, we explicitly estimate and control for changes in the cost function over time. Furthermore, to remove any concerns about learning, our analysis and discussion below differentiates early and late behavior when conceptually relevant. In Appendix 9.1, we demonstrate that weekly parameter estimates remain stable over the course of the experiment.

3 Model and Identification Strategy

In this section, we outline a model of the experimental decisions and discuss the consequent identification of cost, time-preference, and sophistication parameters. Our design is geared mainly towards the identification of the present-bias parameter, β, and the sophistication-about-present-bias parameter, β_h. Loosely, comparing decisions about future work with those about immediate work allows for the identification of β and the other cost and discounting parameters. Similarly, comparing the decisions about future work with *predictions* about future behavior given different bonus payments allow for the identification of β_h. The identification of β_h depends on our assumptions about how sophisticated participants think about the incentive effects of predictions: we first discuss the model and identification of parameters under the assumption that participants make "straightforward" predictions of future work

both greater than 2, which are both identifiable as outliers (A Grubbs test identifies them as outliers with a p-value less than 10^{-5}) and are hard to interpret psychologically.

without any distortion from prediction-accuracy payments, and then assume that participants use prediction-accuracy payments as soft-commitment devices.

3.1 Theoretical Model

3.1.1 Baseline

We present a simple model in which an agent trades off disutility from effort with consumption utility derived from the consequent payment.[23] Specifically, at some *decision time* k, an agent chooses a number of tasks (an effort level) e to complete at *work time* t for piecewise wage w, which will be received on *payment time* T, where $k \leq t < T$.

We parameterize effort costs (as the instantaneous disutility) as a separable, increasing, and convex function $C(e)$. We assume that the agent discounts utility using quasi-hyperbolic present bias, discounting costs in time t from the perspective of time k by discount factor 1 when $t = k$ and $\beta \cdot \delta^{t-k}$ when $t > k$.

The parameterization of the utility and discounting of payments is more complicated, because it represents the total indirect utility arising from a consumption stream derived from a future monetary payment.[24] For simplicity, we model the payments as generating discounted utility of $\beta \cdot \delta^{T-k} \cdot (e \cdot w)$, due to our expectation that nearly all consumption derived from the payment occurs in the future and that there is relatively little diminishing marginal utility in consumption from the payments in the experiment.[25,26]

[23] We do not include uncertainty about future preferences and circumstances here, but do include this possibility in Appendix 9.3, and provide an intuition why such uncertainty likely works against the finding of present bias. The model—in line with both existing literature and our own ex-ante assumptions—ignores the apparent taste for following predictions we see in the data.

[24] It should be noted that many of the complications end up not mattering at all because no matter our specifications or assumptions we estimate $\delta \approx 1$. The complications arise from all the unobservable contingencies that affect how and when payments influence utility. With no liquidity constraints, utility is derived from the consumption stream associated with the optimal plan to distribute the net-present value of the future payment (which depends on the interest rate). This plan would adjust over time and as information is revealed, with some small part of that change occuring immediately (and hence be part of immediate gratification). With liquidity constraints, the consumption stream would all occur after the first date of liquidity, but may still be spread over time and therefore possibly subject to different discounting. Finally—and as we model in the paper—people might follow a heuristic that does not match their actual realized discounted utility, in which they simply act as if the delivery date of money is the delivery date of utility.

[25] Appendix 9.7 models distinct long-term discount factors δ_m and δ_e for money and effort. Empirically, both of these parameters are identified: variation across decisions in the time-to-payment and time-to-effort identify δ_m and δ_e,and then we separately estimate the two parameters. We also estimate δ_e holding $\delta_m = 1$, all under multiple fixed-effects specifications. The two discount factors are all estimated between 1.000 and 1.004, and none are different by anything approaching statistical significance. The estimates of other parameters are unchanged given the separation.

[26] In Appendix 9.9 we use the random receipt of bonus payments to show that there is little evidence of decreasing marginal utility from small monetary payments. But we also show that assuming significant payment-utility curvature has very little impact on the discount parameters estimates.

Given these assumptions, when the decision is for future effort ($k < t$), the preference of the agent is the solution to

$$e^*_{k<t} = \arg\max_e \; \beta \cdot \delta^{T-k} \cdot (e \cdot w) - \beta \cdot \delta^{t-k} \cdot C(e), \tag{1}$$

where we explicitly do not simplify the δ terms to visually separate effort and monetary discounting.[27] For immediate effort decisions ($k = t$), the agent's preferences are:

$$e^*_{k=t} = \arg\max_e \; \beta \cdot \delta^{T-k} \cdot (e \cdot w) - C(e), \tag{2}$$

where the β parameter no longer applies to work because it occurs in the present. The agent's prospective *prediction* of their solution to the present-work problem (2) replaces the true present-bias parameter with β with the perceived parameter β_h:

$$e^*_p = \arg\max_e \; \beta_h \cdot \delta^{T-k} \cdot (e \cdot w) - C(e). \tag{3}$$

We label this the *straightforward prediction.* Simple renormalizations allows these three equations to be written as one equation that emphasizes the relative effect of timing and prediction on the effort cost function:

$$e^* = \arg\max_e \delta^{T-k} \cdot (e \cdot w) - \frac{1}{\beta^{\mathbf{1}(k=t)}} \cdot \frac{1}{\beta_h^{\mathbf{1}(p=1)}} \cdot \delta^{t-k} \cdot C(e). \tag{4}$$

where $\mathbf{1}(k = t)$ is an indicator function that the decision occurs in the same period as the expenditure of effort and $\mathbf{1}(p = 1)$ is an indicator function that the decision is a prediction.

3.1.2 Predictions-as-Commitment

In our experiment, the prediction decision is more complicated because we incentivize participants' predictions with an accuracy-bonus payment b received at time T. If $\beta_h < 1$, the agent making a prospective prediction has preferences over future effort represented in (1) and desires to complete $e^*_{k<t}$ tasks, but believes that when the future arrives she will have preferences to follow (3), preferring to complete $e^*_p < e^*_{k<t}$ tasks. This time inconsistency could lead an agent to try to compel herself in the future to complete more tasks. The bonus payment can be used

[27] For the agent to truthfully report this preference in the experiment, she must believe that when the work time arrives, she will complete these tasks rather than exiting the experiment and receiving a penalty P. We chose a relatively large penalty ($50) with the presumption that it will force participants to complete previously chosen effort levels. Given the estimated parameters in our main specification, an agent would need to have $\beta_h < 0.1$ to believe that she would prefer to exit. In fact, seven participants were removed from the experiment for not completing the tasks they agreed to do. Reassuringly, six of these removals occured on the first at-home date, suggesting these participants discovered something about the experiment rather than indicating any time inconsistency.

as a soft commitment device to this end. To see this, consider the case of a very large bonus. When the future arrives and the effort decision is chosen, the bonus will compel the agent to match the previous prediction in order to receive the bonus. Given that the bonus effectively gives her control of the future effort decision, the agent will choose a prediction $e^*_{k<t}$ in line with her current preferences in (1). However, this strategy will not work as the bonus becomes smaller: she will choose her new optimal level e^*_p, preferring to lose the small bonus rather than getting the bonus by matching the prediction. Therefore, the optimal prediction given a bonus is the level of effort closest to that in (1) such that the agent believes this deviation will not occur:

$$\begin{aligned} & e^{**}_p = \arg\max_e \ \ \beta \cdot \delta^{T-k} \cdot (e \cdot w) - \beta \cdot \delta^{t-k} \cdot C(e) \\ & \text{such that:} \\ & \beta_h \cdot \delta^{T-k} \cdot (e \cdot w + b) - C(e) \geq \ \ \beta_h \cdot \delta^{T-k} \cdot (e^*_p \cdot w) - C(e^*_p). \end{aligned} \tag{5}$$

Recall that we equally rewarded predictions that were within five tasks of the chosen effort, leading to the second distortion. Given $\beta_h < 1$, the agent perceives that she will prefer a lower effort when the work time arrives than when work is in the future. As she will be able to choose five fewer tasks than the prediction and still receive the bonus when future arrives, the agent must adjust her prediction to be five higher than her target level to incentivize herself to complete these tasks, leading to the adjusted decision

$$e^{***}_p = e^{**}_p + 5. \tag{6}$$ [28,29,30]

Given these two effects, an agent with $\beta_h < 1$ who recognizes the incentive effects of prediction-accuracy payments will make predictions that rise with the bonus amount and level out at future effort choices plus five tasks. Conversely, straightforward predictions do not vary with the bonus level b. When $\beta_h = 1$, there is no difference across predictions with different prediction-accuracy payments.

[28] Similarly, when $\beta_h > 1$, the agent must adjust her prediction downward. When $\beta_h = 1$, the agent is technically indifferent between all choices within 5 of $e^*_{k<t}$—this agent believes that she is time-consistent and will choose $e^*_{k<t}$ in the future regardless of the bonus.

[29] Of course, the suggestion that the agent's prediction will be exactly five tasks above e^{**}_p ignores uncertainty. Unfortunately, the precise effect of uncertainty on the prediction depends heavily on the location and shape of the error term.

[30] For example, consider a simple parameterization where $\beta = .6$, $\beta_h = .8, w = 1, \delta = 1$, $C(e) = \frac{1}{200}e^2$. In this case, the agent prefers to complete $e^*_{k<t} = 100$ future tasks, but only prefers $e^*_{k=t} = 60$ immediate tasks. She perceives that, in the future, she will prefer to complete $e^*_p = 80$ tasks, but that a bonus can compel her to choose at most $80 + 4\sqrt{10b}$) tasks. Given her current preference of $e^*_{k<t} = 100$, $e^{**}_p = \min(80 + 4\sqrt{10b}, 100)$ and $e^{***}_p = e^{**}_p + 5$. For example, if $b = \frac{5}{8}$, then $e^{**}_p = 90$ and $e^{***}_p = 95$.

3.2 Empirical Identification

In order to identify the parameters β, δ,and β_h, we make the further structural assumption that the cost function takes a power form with two parameters, such that $C(e) = \frac{1}{\varphi \cdot \gamma}(e+10)^{\gamma}$, where the 10 represents the mandatory work and the multiplier $\frac{1}{\varphi \cdot \gamma}$ is chosen such that the first-order condition takes a simple form.[31] We discuss identification given straightforward predictions and then separately address identification given that participants use predictions as soft commitments. Throughout the structural estimation in Section 5, we vary the number of parameters assumed to be common across the participants.

3.2.1 Baseline

Adding in the zero-wage ten mandatory tasks required in the experiment and rewriting (4) given the structural assumptions yields:

$$e^* = \arg\max_e \quad \delta^{T-k} \cdot (e \cdot w) - \frac{1}{\beta^{\mathbf{1}(k=t)}} \cdot \frac{1}{\beta_h^{\mathbf{1}(p=1)}} \cdot \delta^{t-k} \cdot \frac{1}{\varphi \cdot \gamma}(e+10)^{\gamma}. \tag{7}$$

Taking the first-order condition of (7) with respect to e and solving for e yields the predicted choice e^* given parameters β, β_h, δ, γ, φ and experimental variation in w, k, and the type of decision:

$$e^* = \left(\frac{\delta^{T-k} \cdot \varphi \cdot w}{\frac{1}{\beta^{\mathbf{1}(k=t)}} \cdot \frac{1}{\beta_h^{\mathbf{1}(p=1)}} \cdot \delta^{t-k}}\right)^{\frac{1}{\gamma-1}} - 10. \tag{8}$$

Assuming that observed effort is distributed around this predicted level of effort with a normal error term ε with mean 0 and standard deviation σ yields a likelihood of observing work decision e_j of

$$L(e_j) = \phi\left(\frac{e_j^* - e_j}{\sigma}\right), \tag{9}$$

where ϕ is the standard normal probability density function.[32] To deal with the presence of

[31] The parameter φ is necessary and represents the "exchange rate" between effort and the payment amount. If instead $C(e) = \frac{1}{\gamma}e^{\gamma}$, a requirement such as linear marginal costs (which necessitates $\gamma = 2$), would also imply that the marginal cost of e tasks is exactly e monetary units, regardless of the task type or the payment currency.

[32] Loosely, each parameter is identified by variation in the experimental variables: whether the decision occurs on the day of effort ($k = t$) identifies β; whether the decision is a prediction ($p = 1$) identifies β_h; wage variation w identifies φ and γ; variation in e_j identifies σ; and the timing of the decision (T, t, k) identifies δ. For the latter, the timing is somewhat endogenous because participants choose the exact participation dates within a set of constraints. Appendix 9.17 shows that there is virtually no change in our estimates when we instead use the average distance between participation dates in our estimation.

corner solutions in effort decisions (where participants are required to choose between 0 and 100), we follow the correction in a Tobit regression and adjust the likelihood to account for the possibility that the tangency condition implied by (8) does not hold with equality:

$$L^{tobit}(e_j) = \mathbf{1}(e_j < 100)\phi\left(\frac{e_j^* - e_j}{\sigma}\right) + \mathbf{1}(e_j = 100)\Phi\left(\frac{e_j^* - 100}{\sigma}\right), \tag{10}$$

where Φ is the standard normal cumulative density function. For estimation of the parameters, we maximize the sum of the logarithms of L^{tobit} using standard maximum-likelihood routines.

3.2.2 Predictions-as-Commitment

Assuming that people are using predictions as soft commitments, as in equation (6), the optimal prediction changes to

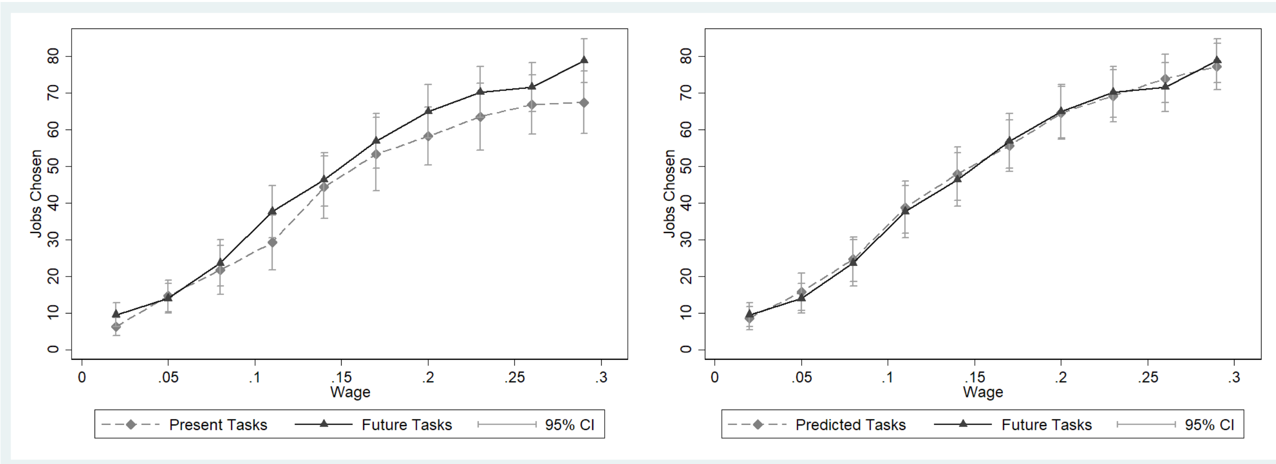
$$\begin{aligned}
&5 + \arg\max_e \;\; \beta \cdot \delta^{T-k} \cdot (e \cdot w) - \beta \cdot \delta^{t-k} \cdot \tfrac{1}{\varphi \cdot \gamma}(e+10)^{\gamma} \\
&\text{such that:} \\
&\;\; \beta_h \cdot \delta^{(T-k)} \cdot (e \cdot w + b) - \tfrac{1}{\varphi \cdot \gamma}(e+10)^{\gamma} \geq \\
&\;\; \beta_h \cdot \delta^{(T-k)} \cdot (e^* \cdot w) - \tfrac{1}{\varphi \cdot \gamma}(e^*+10)^{\gamma}
\end{aligned} \tag{11}$$

when $\beta_h < 1$ (with an analogous adjustment of -5 when $\beta_h > 1$) and e^* is the straightforward prediction of the optimal action when work time t arrives. To avoid a discontinuity at $\beta_h = 1$ (where there is no adjustment of five tasks), we locally smooth the adjustment around $\beta_h = 1$. There is a unique solution to this maximization problem as the objective function is concave under the assumption that $\gamma > 1$, but this solution must be determined numerically.

4 Reduced-Form Analysis

Before estimating the parameters in our structural model, this section describes the data and performs a set of simple non-parametric analyses on the aggregate and individual level. Recall that the interpretation of participants' predictions about future work decisions depends on participants' recognition that predictions can be used as soft commitment devices. We begin by abstracting from this issue, and present data as if participants make straightforward predictions about expected effort. We then show that there is evidence against the two potential distortions that would be expected if participants were using predictions as commitment devices. In Section 5.3, we show that allowing for such use has little effect on the structural results. Throughout the paper, all reported standard errors are clustered at the participant level unless otherwise noted.

Figure 2: Present vs. future decisions (left) and predictions vs. future decisions (right)



Note: Left graph: Comparison between decisions made about work in the future and decisions made about work in the present given different wages. The difference is a reduced-form measure of present bias. Right graph: Comparison between decisions made about work in the future and predictions made about work in the future given different wages. The difference is a reduced-form measure of perception about present bias. For readability, wages are grouped into 10 bins. Confidence intervals are created using standard errors clustered at the individual level.

4.1 Present Bias and Perceptions of Present Bias: Aggregate

To visualize the effect of wages on choices, the two graphs in Figure 2 show the average task decisions given different wages for work decisions and predictions. For visual ease, observations from the 31 wages are placed in ten bins.[33] Note that, for each type of decision, average task decisions monotonically rise with offered wages, suggesting that the 72 participants in our primary sample understood the tradeoff between effort costs and monetary payoffs.[34,35,36]

Recall that present bias is intuitively identified by the comparison of decisions about future work and decisions about immediate work. The left-hand graph compares these decisions for different wages. The task decisions for immediate work appear consistently lower than the

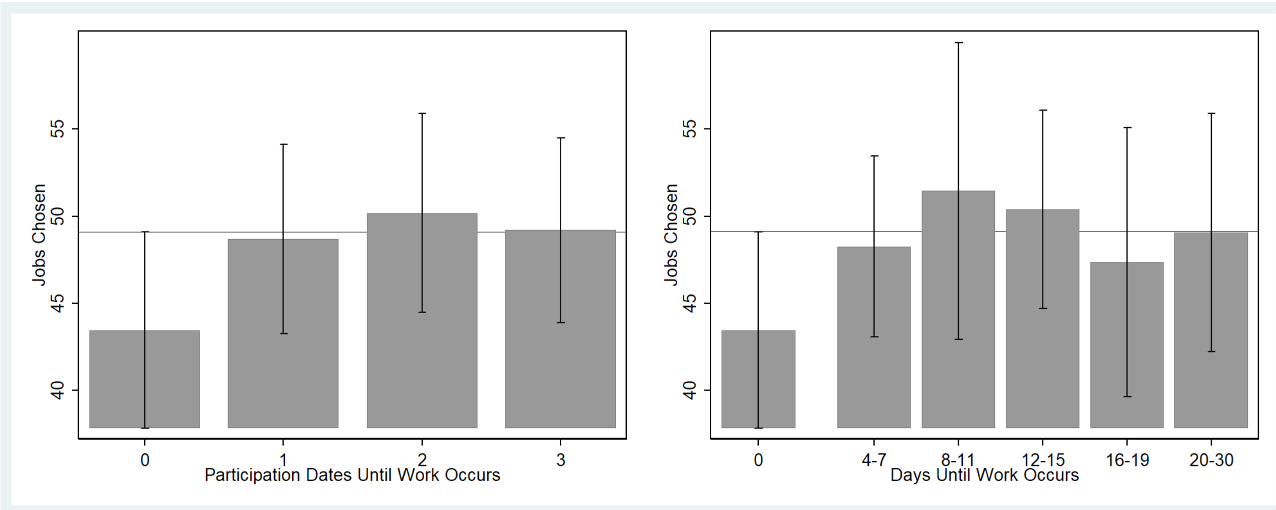
[33] Appendix 9.19 contains the equivalent graph with no wage-bin aggregation.

[34] Individually, 66 of the 72 participants have fewer than 5 total non-monotonicities as wages rise within a decision set (given a total of 104 violation opportunities in adjacent decisions for the full experiment). Given the random ordering of wages, we take this as evidence that the participants in our primary sample understood the main tradeoff in the experiment.

[35] Large monotonicity violations would likely cause our individual estimation strategy to fail and lead to the removal of the participant from the primary sample. Appendix section 9.8 plots the labor supply curve of the removed participants: seventeen are upward sloping, ten are nearly flat (including seven who cluster at one corner option of 0 or 100 tasks), and only one clearly slopes downward.

[36] In analysis suggested by a referee, we find no evidence that choices for a given wage are affected by the magnitude of the other four wages in a decision set. Specifically, decisions for a given wage appear unchanged by the average, maximum, and minimum of the other four wages and are similarly unaffected by the relative rank of the wage in the decision set.

Figure 3: Work decisions given different delays to the time of work



Note: Left graph: Work decisions given the number of participation dates until work occurs. Right graph: Work decisions given the number of calendar days until work occurs. The 2 percent of decisions about work in 22 or more days are added to the final column. Confidence intervals are created using standard errors clustered at the individual level.

decisions for future work, particularly for higher wage levels, which matches the predictions of a time-inconsistent multiplicative discount function. When making decisions about immediate work, participants choose an average of 5.7 ($Z = 4.10$, $p < 0.001$) fewer tasks than when making decisions about future work.[37,38]

Our evidence speaks fairly loudly about the quasi-hyperbolic specification of present bias, where the departure from exponential discounting occurs only between "now" and "later." The two graphs in Figure 3 show the average task decisions across all wages depending on the number of participation dates (left graph) or calendar days (right graph) until the work must be completed, with zero days representing decisions about immediate work. Because no participant made a decision more than three dates into the future and dates were required to lie between four and ten days apart, future dates are all between four and 30 days into the future. For readability, calendar days are combined into four-day groups. The visual evidence that people are present-biased but do not differentiate among any of the days beyond four days is confirmed by statistical tests. Task decisions about immediate work are 5.2, 6.7, and 5.7

[37] The difference remains similar at 5.3 ($Z = 4.52$, $p < 0.001$) when controlling for fixed effects for each wage and 4.9 ($Z = 3.05$, $p = 0.003$) when focusing only on decisions after the second participation date.

[38] Appendix 9.18 presents the distributions of the residuals of a regression of task decisions on participant and wage fixed effects. The distributions are well approximated by a normal distribution and the distribution from future decisions first-order stochastically dominates that from present decisions. In constrast, the distributions from future decisions and prediction decisions are nearly identical.

lower than decisions for work in the following three future dates, respectively, with $p < 0.01$.[39] None of the decisions for work on future dates are pairwise statistically significantly different (the highest statistic is $Z = 0.89$, $p = 0.38$). The same basic results hold with calendar days. Our conclusions are limited in one obvious way: the closest "later" decision is four days from work and the vast majority are greater than a week away.[40]

Assuming that participants are making straightforward predictions, the perception of present bias is identified by the comparison of predictions with work decisions. Intuitively, a completely sophisticated participant asked to predict immediate-work decisions in the future will predict in line with her immediate-work preferences and a completely naive participant will predict in line with her future-work preferences.

The right panel of Figure 2 compares predictions and future work decisions for different wages. It provides clear visual evidence for naivete: predictions about future work are largely in line with future-work decisions. There is an average difference of 0.7 ($Z = 1.11$, $p = 0.27$) fewer tasks in predictions compared to future-work decisions.[41] When comparing to immediate-work decisions, predictions are 4.9 higher ($Z = 3.48$, $p = 0.001$) on average.[42] At the aggregate level, participants appear to have little sophistication about their own present bias.

4.2 Present Bias and Perceptions of Present Bias: Individual

Figure 2 is based on aggregate data. To provide an assessment of decisions on an individual level, Figure 4 presents a histogram of the measures discussed above within each individual. Fixed effects for ten wage bins are added due to the small number of observations for each individual, although there is still presumably a large amount of noise in the data. The left-hand side presents the average difference between decisions about immediate work and future work for each individual, a reduced-form measure of individual present bias. Following the results above, the distribution is centered to the left of zero, suggesting a preference for less work in the present, with 76% of individuals choosing fewer tasks for immediate decisions.

The middle panel presents the difference between predictions and decisions about future

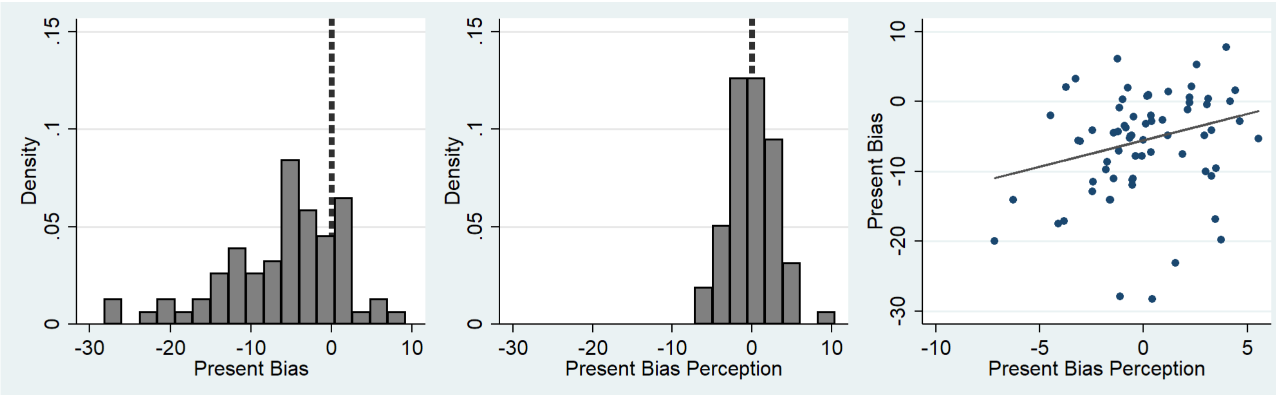
[39] The differences remain similar at 5.0, 5.5, and 5.5 ($Z = 3.93$, $p < 0.001$; $Z = 3.92$, $p < 0.001$; $Z = 3.72$, $p < 0.001$) when controlling for fixed effects for each wage and 4.8, 4.4, and 5.2 ($Z = 2.77$, $p = 0.007$; $Z = 2.02$, $p = 0.048$; $Z = 2.30$, $p = 0.025$) when focusing only on decisions after the second participation date.

[40] Kaur, Kremer, and Mullainathan (2015) argue that the patterns of work ahead of deadlines suggest that people do differentiate between one-day and 3-day delays. Augenblick (2017) uses a similar methodology to study changes in task decisions throughout the week approaching work, finding large changes in the final 24 hours before work.

[41] This difference remains statistically insignificant regardless of controls, with participants choosing an average of 0.2 ($Z = 0.51$, $p = 0.61$) more tasks for predictions when controlling for fixed effects for each wage and 0.2 ($Z = 0.20$, $p = 0.84$) more when focusing only on decisions after the second participation date.

[42] The difference remains stable at 5.0 ($Z = 3.27$, $p = 0.002$) with wage fixed effects and 5.4 ($Z = 4.78$, $p < 0.001$) focusing on later decisions.

Figure 4: Participant-level non-parametric measures of actual and perceived present bias



Note: Left graph: Participant-level differences between decisions made about work in the future and decisions made about work in the present given different wages, a measure of present bias. Center graph: Participant-level differences between work decisions in the future and predictions made about work in the future given different wages, a measure of perception about present bias. These graphs include a vertical line at zero. Right graph: Scatterplot of the individual measures (removing three outliers for visual ease) with the robust linear regression line from all participants. The scatterplot including the three outliers is shown in Appendix Section 9.6.

work, a reduced-form measure of individual sophistication. Following the results above, the distribution is centered near zero, with 53% of individuals to the left of zero. Note that there is less variance in the individual differences in the right-hand graph than in the left-hand graph.

Individual predictions about future work decisions are related to individual outcomes: the correlation coefficient between the participants' distances in the left and middle panels of Figure 4 is 0.25, which is statistically significant ($Z = 2.13$, $p < 0.04$).[43] Given the concern that outliers might be driving the result, it is reassuring that the significance of the relationship rises when using the robust regression techniques of M-estimation ($p = 0.01$), MM-estimation ($p < 0.03$), and S-estimation ($p < 0.001$).[44]

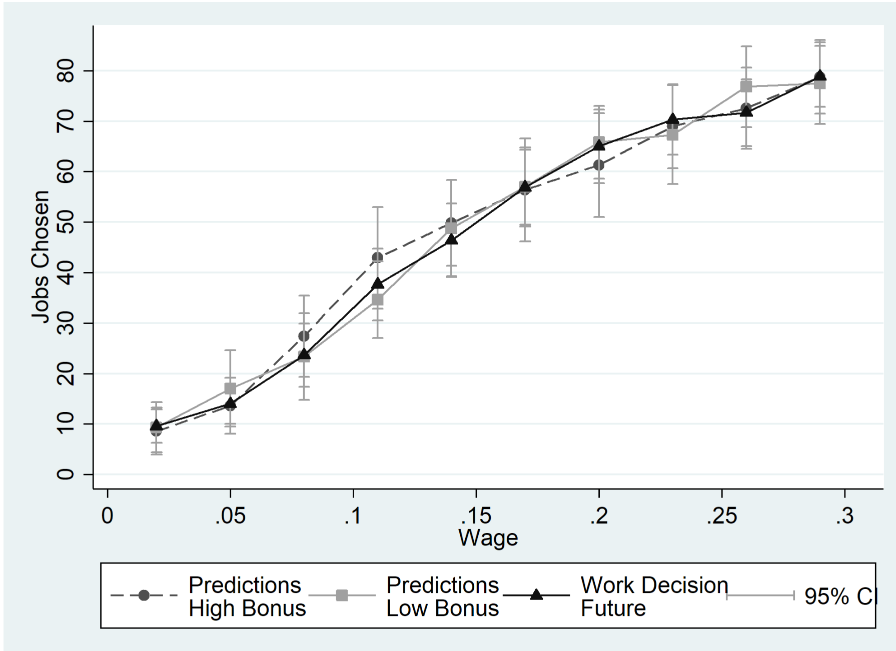
4.3 Prediction as Commitment

So far we have analyzed the data as if participants are not distorting their predictions as an attempt to modify future incentives. This assumption would be valid insofar as participants are fully naive about their present bias, or if they are (at least partially) sophisticated but fail to think through the incentive effects of prediction-accuracy payments. As discussed in

[43] All correlations in the paper refer to Pearson's Correlation. The results are broadly similar using alternative measures of correlation, such as the Spearman's Correlation.

[44] Robust regressions require the specification of additional tuning parameters. All of our results use the default choices in the stata command *robreg*.

Figure 5: The variation of predictions given different bonus amounts



Note: Comparison between predictions made about work in the future and decisions made about work in the future. Predictions are shown when bonuses are relatively small or large. For readability, wages are grouped into 10 bins. Confidence intervals are created using standard errors clustered at the individual level.

Section 3, when this assumption is not satisfied, participants will predict five tasks above their current preferences for high bonus amounts and predict fewer tasks for lower bonus amounts. We demonstrate that, for whatever reason, we do not find evidence for these behaviors in our data.[45]

Figure 5 shows the average prediction given different wages under low and high bonus levels, as well as future decisions about work, with medium bonuses omitted for visual ease.[46] First, there does not appear to be a connection between the bonus level and predictions at the aggregate level, a conclusion that is confirmed with a non-parametric analysis. Participants facing low, medium, and high bonuses predicted an average of 0.1, 0.7, and 1.2 tasks fewer than future work decisions. These differences are not statistically significant ($Z = 0.11$, $p = 0.92$; $Z = 0.45$, $p = 0.66$; $Z = 0.88$, $p = 0.38$) from zero and not statistically different from each other (for high and low bonuses, $Z = .054$, $p = 0.59$). Controlling for wage fixed effects and

[45] As planned, we explore changes in predictions given different bonus levels only at the aggregate level given a lack of power at the individual level: each participant faces 10 prediction sets and each set is associated with the same bonus level.

[46] The cutoffs were determined to attempt to equalize the number of observations in three bins of low, medium, and high bonuses. Low and high bonuses lie between \$0.25-\$1.00 and \$4.25-\$8.25, respectively. The conclusions are robust to the number or borders of the bonus bins.

focusing on later decisions, the effect of bonuses on predictions are non-monotonic and similarly not statistically significant.[47] Second, while the theory predicts that agents using predictions for commitment purposes will choose five tasks over preferences for future work for relatively large bonuses, we find no evidence for this behavior for any bonus size. Therefore, in aggregate, there is little evidence that participants are using the prediction to manipulate the incentives of their future selves.

5 Structural Analysis

In this section, we estimate time-discounting, sophistication, cost, and projection-bias parameters under our structural assumptions and identification strategy outlined in Section 3. The qualitative conclusions of this section largely follow the conclusions of the non-parametric analysis. We first estimate the parameters under the assumption of straightforward predictions, which we relax with little effect on the estimates. As in the reduced-form section, all standard errors are clustered at the participant level when applicable.

5.1 Present Bias and Perceptions of Present Bias: Aggregate

Table 1 estimates the main structural parameters through the maximization of the likelihood in (10) for our primary sample of 72 participants. The first four columns present the estimation results assuming that participants are making straightforward predictions. Column (1) presents the estimation under the assumption of common parameters for each participant. Column (2) adds participant fixed effects, effectively allowing the slope of the cost curve to vary arbitrarily for each participant.[48] Column (3) adds additional fixed effects for the date of decision, which allows for the slope and curvature of the cost curve to change across time.[49] Column (4) focuses only on decisions made on or after the third date. We return shortly to Column (5), which

[47] The predictions for low, medium, and high bonuses are 0.6 higher ($Z = 0.62$, $p = 0.54$), 0.2 lower ($Z = 0.19$, $p = 0.85$), and 0.4 higher ($Z = 0.32$, $p = 0.75$) than future work decisions when controlling for fixed effects for each wage, and are 0.8 higher ($Z = 0.53$, $p = 0.60$), 1.0 higher ($Z = 0.51$, $p = 0.61$), and 1.6 lower ($Z = 1.00$, $p = 0.32$) when focusing only on decisions after the second participation date, with all pairs remaining statistically indistinguishable (for high and low bonuses: $Z = 0.23$, $p = 0.82$; $Z = 1.00$, $p = 0.32$).

[48] Originally, we planned on using participant fixed effects on both the cost slope φ and curvature γ parameters. However, this specification does not converge when decision-date fixed effects are added (in this estimation, there are 144 subject fixed effects and 14 date fixed effects). In Appendix Section 9.13, we discuss the effect of different combinations of date and participant fixed effects and show there is little impact on the results.

[49] The fixed effects control for consistent changes in the cost curve as the *date-of-decision* changes, capturing learning about the task over time. They do not control for consistent changes in the cost curve as the *date-of-work* in the decision changes. These changes could occur if, for example, all participants face a midterm on a specific participation date, which would lead them to choose fewer tasks to be completed on that date *regardless* of when the decision was made. In Appendix 9.14, we account for these effects by adding work-date fixed effects and find no change in our conclusions.

includes the assumption that participants are manipulating predictions for incentive purposes, briefly noting that it is largely similar to the other columns. When a parameter is estimated with fixed effects, we report the average parameter across these fixed effects. All specifications also include the estimation of a projection bias parameter $\widetilde{\alpha}$, which is discussed separately in Section 6.[50]

The aggregate estimate of the present-bias parameter β ranges from 0.812 to 0.835. In each case, the null hypothesis of no present bias ($\beta = 1$) is firmly rejected. Estimates of the perceived present-bias parameter β_h range from 0.999 to 1.014. In each case, the null hypothesis of no perceived present bias ($\beta_h = 1$) cannot be rejected. Estimates of the standard daily time discounting parameter estimates δ range from 1.003 to 1.005. The null hypothesis of no standard discounting ($\delta = 1$) is only rejected in specification (2) under classic criteria, although the test statistic is borderline significant in all but column (1). A parameter δ greater than one suggests that people prefer to complete less work as the temporal distance to work increases. One explanation for this finding is that participants do not want to commit to more work farther in the future because there is greater uncertainty about other obligations on these dates.

In all specifications, the aggregate effort-cost parameter estimates hover near two, suggesting near quadratic costs. Quadratic costs imply that the marginal cost of completing a task rises linearly with the number of tasks completed. The slope parameter φ is around 700 in the specifications, which allows costs to be converted into dollar amounts. For example, in the first specification—where cost curves are assumed to be stable over time—the marginal costs of the 25th, 50th, 75th and 100th future-work task are \$0.055, \$0.122, \$0.194, and \$0.270, respectively. Although not shown in the table, the fixed effects specification in Column (3) produces estimates of the cost curve for each date. For example, the marginal costs of the 50th tasks for the seven dates are estimated at \$0.110, \$0.110, \$0.116, \$0.123, \$0.126, \$0.129, and \$0.129, respectively. Although task time-to-completion drops over time, this finding suggests that the task causes slightly higher disutility as the experiment progresses. Reassuringly, controlling for this learning does not noticeably change the estimates of β, β_h, δ, implying that it is not spuriously driving our results. In Appendix 9.1, we further discuss changes in the cost curve and show that the main parameter estimates are stable over time.

Following the discussion of the appropriateness of the quasi-hyperbolic discounting model in the reduced-form analysis, we separately estimate the relative weight placed on the disutility from tasks that must be completed on future dates in comparison to tasks that must be completely immediately. We find that the weights for tasks that must be completed one, two, and three dates away are 17%, 20%, and 17% higher than the weight placed on immediate tasks.

[50]The entire table is replicated in Appendix 9.5 without the projection bias parameter – all other parameters are virtually unchanged.

Table 1: Primary aggregate structural estimation

	(1) Initial Estimation	(2) Participant FE	(3) Decision Day FE	(4) Later Decisions	(5) Pred. Soph.
Present Bias β	0.835	0.812	0.833	0.833	0.825
	(0.038)	(0.042)	(0.040)	(0.041)	(0.041)
Naive Pres. Bias β_h	0.999	1.014	1.006	1.003	1.004
	(0.011)	(0.011)	(0.010)	(0.009)	(0.003)
Discount Factor δ	1.003	1.005	1.003	1.003	1.003
	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
Cost Curvature γ	2.145	2.142	2.118	1.971	2.126
	(0.070)	(0.084)	(0.081)	(0.079)	(0.082)
Cost Slope φ	724	710	687	367	720
	(251)	(268)	(244)	(127)	(259)
Proj Task Reduction $\tilde{\alpha}$	7.302	5.257	5.269	4.066	5.207
	(2.597)	(1.278)	(1.290)	(1.262)	(1.269)
Participant FE		X	X	X	X
Day FE			X	X	X
Prediction Soph.					X
Later Decisions				X	
Observations	8049	8049	8049	5539	8049
Participants	72	72	72	64	72
Log Likelihood	-28412	-25079	-24838	-16522	-24837
$H_0(\hat{\beta}=1)$	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
$H_0(\widehat{\beta_h}=1)$	p= 0.93	p= 0.23	p= 0.58	p= 0.74	p= 0.11
$H_0(\widehat{\alpha}=0)$	p= 0.005	p<0.001	p<0.001	p= 0.001	p<0.001
$H_0(\widehat{\delta}=1)$	p= 0.37	p= 0.01	p= 0.06	p= 0.08	p= 0.07

Note: Our main aggregate structural estimations for our primary sample of 72 participants. Columns (1),(2),(3),(4) assume straightforward predictions. Column (1) presents the baseline estimation. Column (2) adds fixed effects for participants applied to the slope parameter. Column (3) adds fixed effects for decision dates applied to the effort cost and slope parameters. Column (4) matches column (3) but focuses on participation dates three and later. Column (5) matches column (3) but assumes that participants use predictions-as-commitments. When fixed effects are added, the parameter presented is the average of the fixed effects. In all specifications, standard errors are clustered at the participant level.

All of these estimates are significantly different from 0 ($\chi^2(1)$=13.66, p<0.001; $\chi^2(1)$=13.69, p<0.001; $\chi^2(1)$=9.85, $p = 0.002$).[51]

5.2 Present Bias and Perceptions of Present Bias: Individual

To analyze individual heterogeneity, we estimate participant-specific parameters using a variety of specifications. We initially assume that participants make straightforward predictions and have no projection bias. The mean, median, and standard deviation of the main individual parameters for all decisions, early decisions (on or before the third date), and late decisions (on or after the fourth date) are shown in Columns (1)-(3) in Table 2. Note that, as we did to create our primary sample of 72 participants, we remove participants when the maximum-likelihood routine does not converge or produces estimates that are transparent outliers for each specification.[52] The results largely mirror the aggregate results, with the average estimate of β ranging from 0.794-0.874, the average estimate of β_h ranging from 0.978-0.988, the average estimate of δ ranging from 1.006-1.016, and the average estimate of γ ranging from 2.044-2.160.

Figure 6 contains histograms of the individual estimates of β (left panel) and β_h (right panel) from the main specification in Column (1). Note that the histograms largely match the corresponding histograms of reduced-form measures of present bias and sophistication calculated in Section 4.2 and shown in Figure 4. The correlation between the individual estimates of β (β_h) and the measure of present bias (sophistication) is 0.679 (0.594), providing in-sample validation of the structural parameter estimates. The correlation between the estimates of β and the estimates of β_h across specifications ranges from 0.244-0.284, which is always statistically significant ($p = 0.016$, $p = 0.039$, $p = 0.044$). The significance of the relationship almost universally rises when using the robust regression techniques of M-estimation ($p = 0.004, p = 0.191, p = 0.023$), MM-estimation ($p < 0.001, p = 0.033, p = 0.033$), and S-estimation ($p < 0.001, p < 0.001, p < 0.001$).

However, while the correlation suggests that participants are (on average) at least partially aware of their own present bias, it does not quantify the level of awareness. As a quantification

[51] The weights remain similar when adding participant fixed effects at 17%, 20%, 23% ($\chi^2(1)$=10.41, p=0.001; $\chi^2(1)$=11.81, p<0.001; $\chi^2(1)$=11.95, p<0.001) and both participant and decision-date fixed effects at 17%, 17%, 14% ($\chi^2(1)$=10.61, p=0.001; $\chi^2(1)$=9.74, p=0.002; $\chi^2(1)$=6.56, p=0.01). In all of these specifications, none of the weights on future participation dates are pairwise statistically significantly different from each other (the highest statistic is $\chi^2(1)$=2.38, p=0.12 comparing one and three dates away in the second specification).

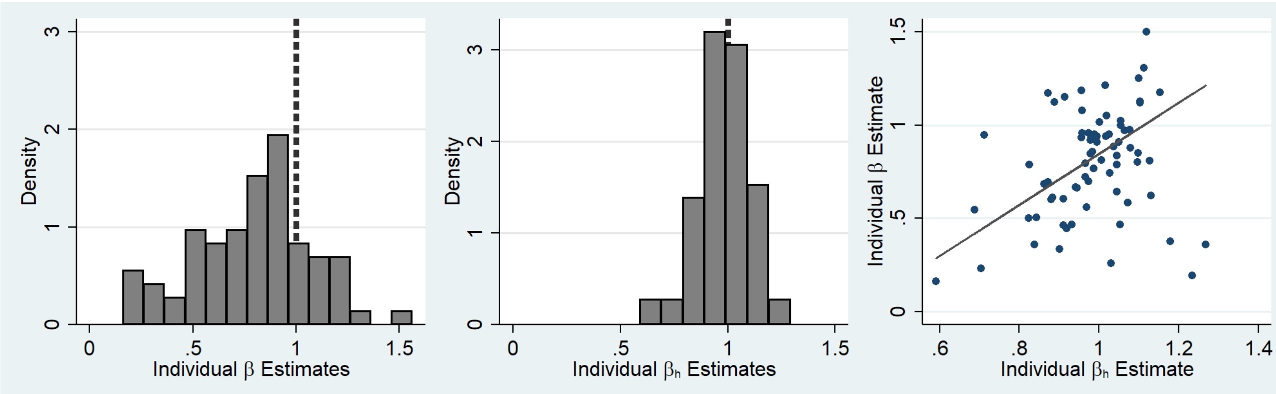
[52] We remove outliers whose estimates are rejected by a Grubb's outlier test with 99.99% confidence. Specifically, 6, 1, 7, and 1 outliers are removed from the "Early Decisions", "Later Decisions", "Proj. Bias", and "Pred. Soph." specifications, respectively. The largest source of outliers are very large β_h parameters ($\beta_h \in [2, 33]$) . When these outliers are included, the means of β_h in Table 2 change to .984, 1.60, 0.97, 1.11, and 1.23, respectively, although the medians remain largely unchanged. Removing more outliers (by changing the confidence level to 99% or 95%) does not meaningfully change the means, except in the case of the estimates of β_h in the estimation with sophistication, which swings between 0.87 and 0.97 depending on the cutoff.

Table 2: Summary statistics for individual structural estimates

	(1) Primary Estimation	(2) Early Decisions	(3) Later Decisions	(4) Proj. Bias	(5) Pred. Soph.
mean($\hat{\beta}_i$)	0.794	0.853	0.874	0.772	0.826
median($\hat{\beta}_i$)	0.824	0.873	0.901	0.807	0.832
sd($\hat{\beta}_i$)	(0.286)	(0.338)	(0.244)	(0.280)	(0.247)
mean($\hat{\beta}_{h,i}$)	0.984	0.989	0.978	0.985	1.150
median($\hat{\beta}_{h,i}$)	0.988	0.991	0.984	0.997	1.002
sd($\hat{\beta}_{h,i}$)	(0.120)	(0.153)	(0.101)	(0.118)	(0.945)
mean($\hat{\delta}_i$)	1.016	1.010	1.006	1.016	1.007
median($\hat{\delta}_i$)	1.008	1.005	1.004	1.008	1.005
sd($\hat{\delta}_i$)	(0.035)	(0.024)	(0.024)	(0.035)	(0.020)
mean($\hat{\gamma}_i$)	2.138	2.144	2.044	2.245	2.063
median($\hat{\gamma}_i$)	1.930	1.957	1.905	2.050	1.981
sd($\hat{\gamma}_i$)	(0.692)	(0.762)	(0.636)	(0.727)	(0.407)
mean($\hat{\alpha}_i$)				6.413	
median($\hat{\alpha}_i$)				3.238	
sd($\hat{\alpha}_i$)				(11.288)	
P[$\hat{\beta}_i$]<1	0.78	0.76	0.69	0.79	0.78
P[$\hat{\beta}_{h,i}$]<1	0.54	0.56	0.60	0.51	0.48
r($\hat{\beta}_i$, $\hat{\beta}_{h,i}$)	0.284	0.244	0.273	0.237	0.060
p-value r($\hat{\beta}_i$, $\hat{\beta}_{h,i}$)	0.016	0.039	0.044	0.052	0.654
Observations	72	72	55	68	58

Note: Summary statistics of the individual parameter estimates. Columns (1)-(3) show our main estimation on the entire sample, early decisions (on or before the third date), and late decisions (on or after the fourth date), respectively. Column (4) uses the entire sample and includes the projection bias parameter estimate, while Column (5) assumes that subjects fully appreciate the use of accuracy payments as commitment devices. Participants for whom the estimation does not converge or creates strongly-outlying estimates are removed. sd($\hat{x}_i$) is the standard deviation of the distribution of individual estimates (not the average standard error). P[$\hat{x}_i$]<1 is the proportion of estimates below 1. r($\hat{\beta}_i$, $\hat{\beta}_{h,i}$) is the correlation coefficient between $\hat{\beta}_i$ and $\hat{\beta}_{h,i}$, and p-value (r($\hat{\beta}_i$, $\hat{\beta}_{h,i}$)) is the p-value of the test that the coefficient is zero.

Figure 6: Individual estimates $\hat{\beta}_i$ and $\hat{\beta}_{h,i}$



Note: Left graph: Distribution of individual estimates $\hat{\beta}_i$. Center graph: Distribution of individual estimates $\hat{\beta}_{h,i}$. These graphs include a vertical line at one. Right graph: Scatterplot of the individual measures with the robust linear regression line. These can be compared to the non-parametric measures in Figure 4.

across individuals, we report on an intuitive measure of *relative sophistication* λ as the ratio of a person i's belief about her level of present bias, $(1-\beta_{h,i})$, vs. her actual level of present bias $(1-\beta_i)$:

$$\lambda = \frac{(1-\beta_{h,i})}{(1-\beta_i)}. \tag{12}$$

That is, a person with $\beta = .8$ and $\beta_h = .9$ and a person with $\beta = .9$ and $\beta_h = .95$ would both be classified as having relative-sophistication of .5.[53]

The estimates of λ – calculated by regressing the individual estimations $(1-\widehat{\beta}_{h,i})$ on $(1-\widehat{\beta}_i)$ – are displayed in Table 3. Row 1 presents the OLS regression with robust standard errors. Row 2 uses generalized least squares (GLS) with weights equal to the inverse of the individually-estimated variances of the dependant variable $(1-\widehat{\beta}_{h,i})$ derived from the maximum-likelihood routine. Rows 3-5 report estimates using three robust regression techniques to circumvent the influence of outliers. Rows 6-7 attempt to correct for the attenuation bias introduced by using noisy estimates of the independent variable through an errors-in-variable regression, first by using an arbitrary conservative estimate of the reliability coefficient of .5 and then by using a data-driven estimate of the reliability coefficient using the ratio of the average individually-estimated variance of the independent variable $(1-\widehat{\beta}_i)$ to the total variance of this variable (with these estimates displayed in italics below the estimation). Focusing on our

[53]There is little discussion of degree-of-sophistication measures in the literature that try to relate different combinations of β and β_h, presumably because there are so few estimations of individual present-bias and sophistication parameters. An interesting exception is Acland and Levy (2015), who are unable to separately estimate β and β_h, but are able to estimate a degree-of-sophistication measure $\omega = 1-\lambda$.

Table 3: Relationship between individual estimates $\hat{\beta}_i$ and $\hat{\beta_{h,i}}$

	(1) Primary Estimation	(2) Early Decisions	(3) Later Decisions	(4) Projection Bias	(5) With Sophistication
OLS	0.104*	0.105*	0.125**	0.085	-0.137
	(0.061)	(0.062)	(0.054)	(0.062)	(0.401)
GLS	0.133***	0.145**	0.142***	0.106**	0.225
	(0.043)	(0.059)	(0.050)	(0.046)	(0.161)
MM-estimation	0.189***	0.189***	0.078	0.145	-0.196
	(0.057)	(0.065)	(0.049)	(0.116)	(0.263)
M-estimation	0.147**	0.144**	0.098**	0.109	-0.153
	(0.066)	(0.067)	(0.049)	(0.076)	(0.293)
S-estimation	0.238***	0.192***	0.048	0.237***	-0.135
	(0.036)	(0.035)	(0.060)	(0.057)	(0.279)
EIV $\rho_{xx'}$*:=0.5*	0.238***	0.221**	0.225**	0.199**	0.461
	(0.092)	(0.102)	(0.105)	(0.097)	(1.020)
EIV w/ est $\rho_{xx'}$	0.135**	0.204**	0.138**	0.121**	0.298
	(0.054)	(0.094)	(0.066)	(0.061)	(0.659)
(est $\rho_{xx'}$*)*	*0.88*	*0.54*	*0.82*	*0.82*	*0.77*
Observations	72	72	55	68	58

Note: These are the coefficients of the regression of (1-$\hat{\beta}_i$) on (1-$\hat{\beta_{h,i}}$). *GLS* uses in the inverse of the variance estimates for $\hat{\beta_{h,i}}$ provided by the individual estimation as weights. *MM-*, *M-*, and *S-estimations* are robust regressions. *EIV estimated* $\rho_{xx'}$*:=0.5* is an errors-in-variables regression using a conservative fixed reliability coefficient of .5. *EIV estimated* $\rho_{xx'}$ is an errors-in-variables regression using a reliability coefficient calculated using the relative portion of the variance in $\hat{\beta}_i$ not explained by the variance in estimates of $\hat{\beta}_i$. The estimated reliability coefficient appears on the next line. All specifications use robust standard errors. * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

primary estimation in column (1), the estimated coefficients $\widehat{\lambda}$ are statistically significant in all specifications and very significant in the majority of specifications. The coefficient varies from 0.104 to 0.238, leading us to conclude that participants understood between 10-24% of their present bias. Restricting the sample to early or later decisions in columns (2) and (3) largely mirrors this conclusion, with the exception of two of the robust regression results for later decisions. Reassuringly, the analogous relative-sophistication measure calculated using the previously-introduced non-parameteric measures of actual and perceived present bias are also broadly similar and all statistically significant at 0.072, 0.067, 0.086, 0.104, 0.245, 0.144, and 0.096, respectively.

5.3 Prediction-as-Commitment

The parametric estimates above are calculated under the assumption that participants make straightforward predictions. Column (5) in Table 1 mirrors the main aggregate estimation under the assumption that participants optimally use predictions as soft commitment devices. The results are virtually unchanged, which is not surprising given that $\beta_h \approx 1$ in the estimates with straightforward predictions.

We also perform the estimation for each individual under the prediction-as-commitment assumption. However, we fear that this complicated estimation is too demanding given the relatively few observations and small amount of bonus-payment variation for each participant. When running the estimation, we face many more problems with convergence and observe many clear outliers when convergence does occur, leading to estimates for only 58 participants even with liberal inclusion rules. Therefore, we present the results in column (5) of Tables 2 and 3 largely for completeness and are reluctant to draw many conclusions given the results. In Table 2, the means and medians largely mirror previous estimations, although the variance of the individual estimations is much greater and the means are particularly dependant on the rule to remove outliers. The effects of the estimation issues are particularly clear in Table 3, where the results are very unstable across estimations and the standard errors are much larger, leading to a lack of statistical significance even when the estimated coefficients are relatively large.

6 Projection Bias

The experiment was also designed to examine *projection bias*, defined in Loewenstein, O'Donoghue, and Rabin (2003) as the tendency for a person to make decisions about the future as if her tastes in the future will reflect her current tastes rather than predictable future tastes. To examine this bias, we randomly vary if participants express work preferences before or after completing the ten mandatory tasks, which induces different levels of current distaste for the task when making the same decision. In Appendix 9.4, we present a simple model in which the agent (mistakenly) projects her current marginal disutility of work onto her future marginal costs, leading to a prediction that participants will desire to complete fewer tasks when the decision is made after completing mandatory tasks.

To examine this prediction, Figure 7 compares the average of task decisions for different wages given the timing of mandatory tasks. While it does visually appear that participants generally choose less work after completion of mandatory work, the difference is not dramatic. Participants reduce task decisions by an average of 2.1 tasks after completing mandatory work, which is not statistically significant ($Z = 1.28$, $p = 0.20$).[54] The difference rises to 2.5 ($Z = 1.55$, $p = 0.13$) when controlling for fixed effects for each wage and rises to a marginally statistically significant 3.3 ($Z = 1.71$, $p = 0.09$) when focusing only on decisions after the second participation date. This simple analysis does not correct for the fact that 40% of the data is censored, as participants must choose between 0 and 100 tasks. Using a Tobit to account for this censoring, the statistics above rise to 3.8 ($Z = 1.55$, $p = 0.12$), 4.2 ($Z = 1.78$, $p = 0.08$) and 5.8 ($Z = 1.94$, $p = 0.052$).[55,56] We can also estimate this change in task decisions while including all of our main structural assumptions. In the aggregate estimations (Table 1), the change – labeled as $\widetilde{\alpha}$ – ranges from 4.1 to 7.3 and is always statistically significant ($p < 0.005$), while the average individual estimate is 6.4 (Column (4) of Table 2). Therefore, at least when taking into account the censored nature of the data, there is evidence that participants experience projection bias.[57]

7 Discussion and Conclusion

7.1 Related Literature on Present Bias and Projection Bias

Building from millennia of folk wisdom, recent theoretical and empirical research (now too expansive to list fully) has incorporated the idea that people have a taste for immediate gratification. Formal theory and economic applications, such as Strotz (1956), Laibson (1997), O'Donoghue and Rabin (1999a, 2001), have directly considered the relevant components of present bias to be the flow of utility over time rather than the timing of receiving money. Yet a long tradition of experimental research had primarily examined preferred timing of money. Our design follows such papers as Read and van Leeuwen (1998), Badger *et al.* (2007), McClure *et al.* (2007), Brown *et al.* (2009), and Augenblick, Niederle, and Sprenger (2015) in moving away

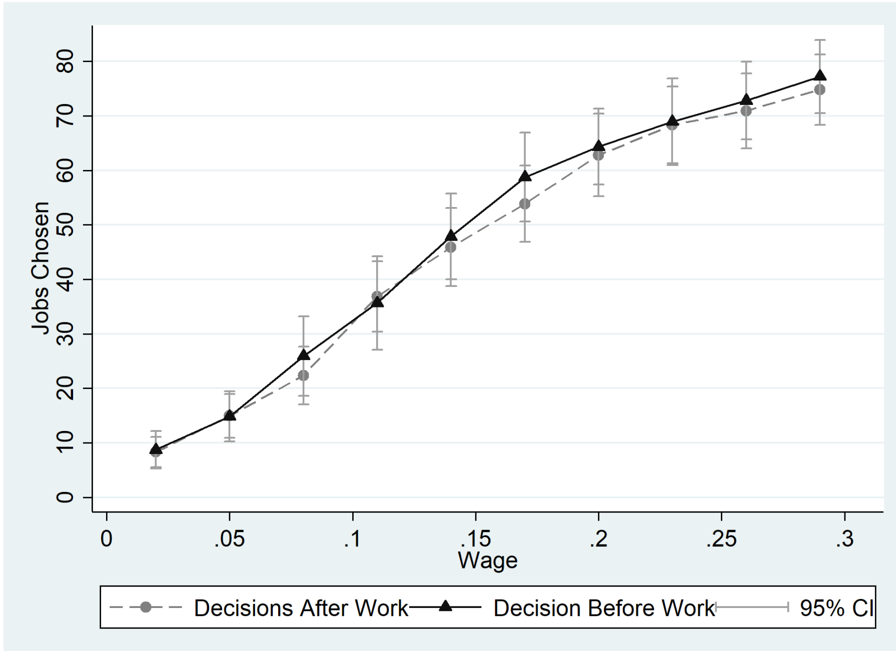
[54] Projection bias applies in all three types of situations we study—choice of immediate work, choice of future work, and prediction about future work—and we had no *a priori* expectation of differences across the situations, and did not plan to analyze them separately. Ex post we did, however, notice strong (and perhaps surprising) differences: participants reduce present work by an average of 5.4 tasks ($Z = 2.04$, $p = 0.045$) after completing mandatory work, while reducing future work by an average of 1.8 tasks ($Z = 0.88$, $p = 0.381$) and predictions by 1.5 tasks ($Z = 0.70$, $p = 0.463$). These p-values are unadjusted for the ex-post nature of estimates.

[55] Indeed, the effects of projection bias seem to show up clearly in the extreme choices: 18.6% of participants chose 0 tasks before the mandatory work, and 20.2% did after, whereas 24.0% chose 100 tasks before the mandatory work and 20.2% after. The differences would appear significant, with ($Z = 2.05$, $p = 0.04$) and ($Z = 2.91$, $p < 0.01$), although these were not ex-ante planned comparisons.

[56] As throughout the paper, the statistical significance reported for these 6 comparisons are from two-sided tests, which does not take our ex ante, projection-bias-based directional prediction into account.

[57] In Appendix 9.4, we structurally estimate the parameter in the projection-bias model, finding a statistically significant estimate that lies within a reasonable range for multiple specifications. However, we are skeptical of its meaning beyond its sign, largely because the estimate is identified off of just two points on the cost curve and leans heavily on the functional form of the cost curve. We therefore believe that the estimation of $\widetilde{\alpha}$ largely captures what we can identify in the data without additional parameteric assumptions.

Figure 7: Comparison of decisions made before and after 10 mandatory tasks.



Note: For readability, wages are grouped into 10 bins. Confidence intervals are created using standard errors clustered at the individual level.

from this money-immediacy paradigm to an approach that more directly tests present bias as employed in economic theory and applications.[58]

There are a range of studies measuring β in both ecological and experimental data using consumption choices. The closest to our experiment is Augenblick, Niederle, and Sprenger

[58] For discussions of the problems and confusions associated with using monetary experiments to investigate the theory of present bias, see Chabris, Laibson and Schuldt (2008), Cubitt and Read (2007), Augenblick, Niederle, and Sprenger (2015), and O'Donoghue and Rabin (2015). O'Donoghue and Rabin (1999b) illustrates the difference in stark form: they invoke present bias to explain *delays* in wealth accumulation: because the effort associated with switching money (or traveling on a tight schedule to pick up and then cash experimental earnings) generates immediate *dis*utility, whereas the increase in wealth is a *delayed* benefit of increased future consumption, present-bias theory predicts delayed wealth accumulation. Although many researchers (including O'Donoghue and Rabin (1999b), citing evidence in contradiction to their own approach) invoked money-immediacy evidence in support of the theory of present bias, the theory itself does not predict that people want money sooner in general. In particular (and important) circumstances, the timing of money receipt can coincide with the timing of utility. This is the case, for instance, when there are plausible liquidity constraints.

(2015), using a very similar task and with participants drawn from the same pool.[59,60,61] Using our different elicitation structure—whereby participants always trade off money vs. work, without ever making any now vs. later choices within the work domain—we obtain a very similar range of estimates for β. By asking participants to make a large number of work choices and specific predictions over many weeks, we are better equipped to estimate each parameter for all individuals, analyze learning over time, and identify if preferences across different future work dates are consistent with the quasi-hyperbolic discounting model. Kaur, Kremer, and Mullainathan (2015) do not estimate a β parameter, but do find that Indian data-entry workers escalate their effort consistently over the week as payday approaches, which they interpret as evidence in favor of smoothly declining discount factors, as opposed to the seeming now-vs-later, quasi-hyperbolic discounting we find. The time horizon in our study unfortunately has little overlap with this study (we have no observations of future-work decisions made less than four days in advance, and very few less than seven days), but the two studies may seem to be at odds. We speculate that the difference might be related to the choice environment (for example, credit-constrained factory workers might be trying to meet weekly earnings targets).[62] Finally, the dynamic nature of the experiment builds on Halevy (2015), who highlights the theoretical issues in separating static and dynamic preference reversals, although his experiment is in the monetary domain and does not share the goal of estimating preference parameters.

There exist few studies measuring β_h from ecological data, and we know of no previous

[59] The tasks were identical, except that—by requiring participants to concurrently identify a random noise through headphones—we feel (we are proud to say) we made ours more irritating. While participants were drawn from the same pool (Berkeley's xlab), we asked the recruitment organizers not to permit the same specific students to sign up. Much of the preliminary analysis in ANS was completed before all the details of our experiment were finalized and the experiment was implemented. We do not believe (based on memory, records, and the nature of the design and adjustment issues) that results from ANS influenced any of the main components of our design.

[60] ANS is primarily concerned with comparing "present bias" in monetary payments—as traditionally studied in experiments—to present bias in real effort—which is meant to capture discounting over utility as the theory is developed for and as most applications assume. To this end, they use Andreoni and Sprenger's (2012) convex-budget-set design, who show no taste for immediate money delivery when asked to trade off money at one time versus another under different exchange rates. They replicate Andreoni and Sprenger's (2012) finding of no taste for the immediate delivery of money, but find present bias using identical procedures where participants trade off effort at different times.

[61] Besides the papers discussed in the text, the only other paper we are familiar with that attempts to experimentally measure present bias in the context of unpleasant effort is Bisin and Hyndman (2014). They require students to complete either one or three word-sorting tasks over time, allowing some of the students to create self-imposed deadlines. By assuming full sophistication and using a structural model of optimal stopping-time choice, the authors estimate $\beta = .44$ in one treatment and $\beta = 1$ in another.

[62] Augenblick (2017) uses a similar methodology to ours, but focuses on work decisions made a week or less from the time of work. He finds a similar drop in the desire to complete tasks across the week as the work time approaches, with two-thirds of the drop occuring within a day of work and one-third occuring within a few hours of work.

studies attempting to use an experimental manipulation to directly estimate β_h.[63] Much of the research on sophistication explores the binary choice of whether or not to self-commit—where self-commitment suggests at least partial awareness of present bias. To take just some of the examples, Ariely and Wertenbroch (2002) document significant student demand for deadlines to complete classroom assignments—and find that these deadlines improve performance. Ashraf, Karlan, and Yin (2006) show that in the Philippines, where credit constraints create a tighter link between money and immediate consumption, 30% of bank clients demand a savings commitment product.[64] Kaur, Kremer, and Mullainathan (2017) find that 36% of the time workers were willing to set a positive (but relatively small) minimal target for which they would be penalized for non-completion. Augenblick, Niederle, and Sprenger (2015) find little willingness to pay for commitment. But around half of the participants in their study choose to commit when it is free, and they find that such commitment is correlated with individual structural estimates of present bias. This seems relatively consistent with our finding that, although there is little evidence of full sophistication, around half of our participants are estimated to have $\beta_h < 1$ (and might therefore demand commitment at a very low price when there is little uncertainty) and that this measure is correlated with the structural estimates of present bias. Perhaps the most dramatic finding of a taste for commitment is Schilbach (2017), who studies rickshaw drivers in India with drinking problems and finds that one-third of the drivers were willing to sacrifice money to receive incentives to be sober. Acland and Levy (2015), who follow Charness and Gneezy (2009) in paying students to attend the gym and examining the development of habits, is the closest antecedent to a structural estimate of β_h using experimental data. While they cannot separately identify β and β_h, an earlier version of the paper (Acland and Levy (2013)) structurally estimates $(1 - \beta_h) = 0.33 \cdot (1 - \beta)$, using the demand for commitment implied by valuations of future-behavior-contingent contracts.

We differ from these studies by attempting to identify sophistication from exploring how closely predictions on future behavior align with the behavior exhibited for immediate choice. In

[63] There are a handful of ecological studies using structural models to estimate various measures of sophistication. For example, Mahajan and Tarozzi (2011) use a dynamic, discrete-choice model to segment the population into naive time-inconsistent agents, sophisticated time-inconsistent agents, and time-consistent agents, and estimate time-preference parameters for each group. Using adoption decisions about bed nets in India, they estimate 40% of the population are fully time-consistent with parameters $\beta_h = \beta = 1$, 50% are naive with parameters $\beta \approx .97, \beta_h = 1$, and 10% are sophisticated with parameters $\beta \approx .55, \beta_h \approx .55$, generating an average $\beta \approx .94$ and $\beta_h \approx .96$. Relatedly, Yang and Wang (2015) use a dynamic, discrete-choice model to structurally estimate time discounting and sophistication parameters with data about adult women's decisions to undertake mammography, finding average estimates of $\delta \approx .78$, $\beta \approx .66$, and $\beta_h \approx 1$. Skiba and Tobacman (2008) use initial borrowing and default timing from a large sample of payday loan borrowers—who are presumably highly credit-constrained—to structurally estimate parameters of $\beta \approx .5$ and β_h of between .9 and 1.

[64] They also show that this demand is predicted (at the 10% level) by time-inconsistent behavior in hypothetical monetary choices, although not (as is more strongly predicted by theory) with hypothetical decisions about rice or ice cream.

addition to helping more finely identify the level of sophistication, focusing on predictions rather than demand for commitment devices can potentially separate sophistication from the taste for commitment. One way to frame this separation is in the language of O'Donoghue and Rabin (1999a), who characterize sophistication as influencing behavior in two ways: a "pessimism effect", whereby a person predicts her own future misbehavior and takes actions now to mitigate the cost of that misbehavior, and an "incentive effect", whereby she tries to manipulate future circumstances to minimize that misbehavior. Although the present-bias model predicts both effects inseparably, our speculation was that the pessimism effect might be operative without the incentive effect, at least in our context where the incentives they could provide themselves (via prediction-accuracy bonuses) might be opaque to the participants. If our finding of little sophisticated pessimism replicates, and studies such as Schilbach (2015) continue to find an apparent strong taste for self-commitment, it might suggest the confusing result that people demand commitment without necessarily being pessimistic about future behavior. That said, these differences might be due to sophistication differing in different situations. Given the low value that Augenblick, Niederle, and Sprenger (2015) find participants place on committing for the same (unfamiliar) tasks we study, one obvious possibility is that the participants in Schilbach's (2015) experiment may have developed a greater sophistication about an activity for which they have years of experience and which is a major problem in their lives.

Finally, we know of relatively few papers with experimental or ecological estimates of the degree of projection bias. We note three exceptions from ecological data. Conlin, O'Donoghue, and Vogelsang (2007) estimate a moderate amount of projection bias from consumers who misorder weather-related clothing based on the idiosyncratic weather conditions at the time they order. Levy (2010) estimates similar level from the failure of cigarette smokers to predict how habit forming their smoking will be. Chang, Huang, and Wang (2017) find that consumers in China are more likely to purchase health insurance on days with high pollution and are more likely to reverse their purchase decision when pollution drops during a cost-free cancellation period. Some of the original (and best) evidence of projection bias comes from experiments—Read and van Leeuwen (1998) find people order future food by current hunger state, and Badger *et al.* (2007) find addicts pay considerably more for *future* delivery of a heroin substitute when their current craving state is unusually high vs. when their current craving state is lower.[65] More recently, Acland and Levy (2013) estimate a large amount of projection bias based on participants' underappreciation of habit formation in their exercise routines. We do not know of other studies of projection bias in the domain of effort over unpleasant tasks. This domain may be valuable because, first, projection bias may be an important force in shaping planning

[65] Both studies were also cited above for being seminal in identifiying present bias; each demonstrated both projection bias and present bias in a way that separately identified the two. Neither study measured the scales of the biases, nor studied naivete about present bias.

decisions for school and work (Kaufmann (2017)); second, work preferences are less "visceral" than those commonly used to study projection bias (such as hunger, addiction, and sex drive), providing some justification for the expanded projection bias model proposed in Loewenstein, O'Donoghue, and Rabin (2003); and third, it allows for a more continuous quantification than many other studies that examine binary choices.

7.2 Potential Limits to the Analysis

There of course remain some potential worries about our identification strategy for present bias. The first concerns the timing of utility. In our analysis and estimation, we assume that all costs associated with completing tasks—captured in the function $C(e)$—correspond to immediate disutility. Insofar as some of the cost function is something not associated with immediate disutility, our estimates of β would be too high, leading us to underestimate the bias.[66] Our measures of projection bias could also potentially be affected: if the source of the *convexity* of the cost function is in part due to displacement of some alternative activity that has concave benefits and is not subject to taste misprediction, then we would also be underestimating the degree of projection bias.

Second, our results could be biased in the presence of uncertainty.[67] In Appendix 9.3, we model a participant who faces random shocks to the convexity of her cost curve, showing that foreseeable uncertainty would tend to make participants want to commit to fewer tasks in the future than under certainty, leading us to *underestimate* the severity of present bias.[68] Given our experimental design, the large number of participant choices allows us to both estimate the level of this uncertainty (essentially comparing the variation in future decisions with the variation in present decisions) and control for it. However, this analysis assumes that participants have rational expectations about the nature of uncertainty, which is potentially incorrect. Furthermore, the effect of uncertainty on the estimates of sophistication *do* depend more on the precise assumptions about the distribution of uncertainty.

The third issue arises as a result of an unanticipated empirical observation about participant decisions given past predictions. To understand the concern, consider a participant who—following our estimates of $\beta_h \approx 1$ and $\beta \approx .83$—is largely naive about her moderate present

[66] Given that much of the experiment occurs online, another unobservable aspect in our design is the possibility that participants "outsourced" the tasks to other people. We cannot rule out the possibility of outsourcing, but are skeptical that it occured in any significant amount. If it did occur, the effects on our parameter estimates are not clear, although our suspicion is that it would mute the finding of present bias.

[67] In an experiment in which participants must choose when to complete an unpleasant take (listen to tone for 20 minutes), Casari and Dragone (2015) conclude that uncertainty about future utility plays an important role in dynamic decision making.

[68] Due to the more complicated relationship between the bet-induced incentives and the wage incentives, the effect of uncertainty on the estimate of β_h is both more complex and more ambiguous.

bias. As we observe in the data, this participant would make predictions that are an average of five tasks above her immediate-work preferences, with the deviation rising in wages. When the work date arrives, our *a priori* assumption that participants only cared about money and effort has some basic implications about the participant's behavior when facing the previous prediction and the associated accuracy payment. Three of the theoretical implications of naivete appear in the data: the participants are most likely to choose at the bottom end of the prediction interval; the average effort is significantly lower than the associated prediction; and the average difference between choice and prediction rises with wages. Two observations, however, are importantly inconsistent with our *a priori* model that participants would treat money from prediction-accuracy bonuses the same way they treat money from wages. First, for predictions strictly between 0 and 100 tasks, the percentage of participants choosing to match their *exact prediction* is nearly 30%, a behavior the model predicts is rare. Second, participants *never* choose outside the prediction interval, although rough simulations given the model suggest that this should occur nearly 20% of the time, with the large majority (80%) of these below the interval. Therefore, it appears that many participants prefer to exactly match their previous predictions, while many others have a preference to choose an effort level that prevents them from losing a prediction-accuracy bonus, even when that is too small to be explained by their monetary preferences. We did not anticipate such preferences, and have not incorporated them into our formal analysis. One barrier to doing so is that we don't, even in hindsight, really understand either of these two different ways that participants dislike behaving inconsistently with earlier "statements." Some natural psychological underpinnings do not seem to be driving behavior: participants' work decisions are not effected by reminders of previous predictions for *other* similar wages, so that it seems that the substantive admission of inconsistency in motives is not at play. Nor is it easy to understand those who choose five tasks fewer than predicted simply to keep a negligible bonus. Such a don't-lose-money-because-of-misprediction motive could potentially be explained by a model such as Eyster (2002) in which people don't like taking actions that make some of their past actions a mistake. If something like this turned out to be the explanation, it would suggest that such a "mechanical" notion of inconsistency aversion (where people are unbothered by seeming inconsistency of beliefs so long as it does not render past choices costly) has power relative to some of the more self-assessment notions of inconsistency.

The observed consistency preferences are in any event potentially problematic for our conclusions about β_h: sophisticated participants could use their unobservable internal-consistency rewards as a commitment device, which would contaminate our estimate. Yet this alternative does not explain away the three above observations indicating evidence of naivete about present bias. Furthermore, we suspect the fine-tuned sophistication required to recognize and use these very particular consistency rewards as commitment devices is unlikely.[69] We discuss the reaction to bonuses further in Appendix 9.2 and supply some additional data. All said, we don't see how this finding suggests that there is more sophistication about present bias than our tests indicate, and our best guess is that participants are not anticipating such behavior any more than we did. Such conjectures are only speculative, however, so that this unexpected and hard-to-interpret finding might reasonably temper confidence in our conclusions against extensive sophistication.

Other issues appear less concerning given further analysis presented in the Appendix. For instance, because participants make multiple decision sets each day, it is possible that the answers on earlier sets influence answers on later sets. And participants making immediate-work decisions for a given wage might be affected when reminded of past predictions about decisions for a *different* wage. In Appendix 9.11, we show that the results are robust to eliminating later decision sets or decision sets with any prediction reminder. Another concern is about selection effects in our main analysis because we focus on only 72 of the 100 participants whose decisions allow estimation of individual parameters. While we have relatively little to say about the individual estimates of the dropped participants (they were removed exactly because of the lack of identification), we show in Appendix 9.8 that their average reduced-form statistics are very similar to our primary sample, and demonstrate in Appendix 9.11 that our aggregate results are consistent when including the data from all participants. Another concern arises because of the artificial and unfamiliar nature of the task we study. Results may be biased if participants learn about the unpleasantness of the task or become more proficient at the task over time. While we observe evidence of both of these effects—the immediate work choices and task completion time both drop significantly in the first few dates—there is little change from the third decision date onward. Our conclusions remain unchanged when focusing only on these later dates or allowing the cost function to vary over decision dates. Furthermore, in Appendix 9.1 we allow the parameters β, β_h, and $\widetilde{\alpha}$ to vary on each decision date, and find that for the most part they remain stable over time.[70] Finally, given the potential number of degrees-of-freedom in an analysis like this, one might be concerned that results are not robust to changes in the exact approach we take. Appendix Sections 9.7-9.17 report on various changes to the analysis and show that our conclusions continue to hold.

[69] This interpretation is in fact consistent with an aspect of the approach of Eyster (2002), where he allows for agents to be naive about their future dislike of inconsistency.

[70] The one exception is hard to interpret: significant movement in the estimate of the parameter $\widetilde{\alpha}$ during the last two participation dates, dropping from 7.5 to 0.3, but then rising back to 11.5.

7.3 Final Thoughts

We conclude with three thoughts on directions of research in this area that are provoked by this experiment. The first is clearest: to better understand the curious results about the aversion to different forms of inconsistency with prior predictions. Further research elaborating how previous predictions affect behavior, and the degree to which these effects are anticipated by participants, is important. The second is to better measure (and gain greater statistical confidence in) our finding confirming our hypothesis that recent effort affects future-oriented choice, so that misprediction of future tastes may indeed be important in (the very classical) domain where people face convex effort costs. We wonder if the effects we found comparing before and after ten initial tasks would be changed if the workload was much greater, the decisions took place over longer horizons, or the type of work was more naturalistic. Finally, the experiment points to potential situations in which present bias and projection bias are confounded. For example, had we in our experiment elicited immediate-work preferences by having people complete tasks for a given wage and freely choose when to stop working, both projection bias and present bias would lead to fewer immediate task completions. While most studies that we are familiar with do not face this confound, there is a potential for researchers focused on discounting to misinterpret degrees of present bias by ignoring projection bias.

8 References

Acland, Daniel and Matthew Levy, "Naivete, Projection Bias, and Habit Formation in Gym Attendance," Management Science 2015, 61(1): 146-160.

Acland, Daniel and Matthew Levy, "Naivete, Projection Bias, and Habit Formation in Gym Attendance," Working Paper, March 2013.

Andersen, Steffan, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutstrom, "Eliciting Risk and Time Preferences," Econometrica, 2008, 76(3), 583–618.

Andreoni, James and Charles Sprenger, "Estimating Time Preferences with Convex Budgets," American Economic Review, 2012, 102 (7), 3333-3356.

Ariely, Dan and Wertenbroch, Klaus, "Procrastination, Deadlines, and Performance: Self-Control by Precommitment, Psychological Science, 2002,13 (May), 219-224

Ashraf, Nava, Dean Karlan, and Wesley Yin, "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," Quarterly Journal of Economics, 2006, 121 (1), 635-672.

Augenblick, Ned. "Short-Term Discounting in Real-Effort Tasks," Working Paper 2017.

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

Badger, Gary, Warren K. Bickel, Louis A. Giordano, Eric A. Jacobs, George F. Loewenstein, and Lisa Marsch, “Altered states: The impact of immediate craving on the valuation of current and future opioids.” Journal of Health Economics, 2007, 26: 865-876.

Bisin, Alberto and Kyle Hyndman, “Present-bias, Procrastination and Deadlines.” Working Paper, 2014.

Brown, Alexander L., Zhikang Eric Chua, and Colin F. Camerer, “Learning and Visceral Temptation in Dynamic Saving Experiments,” Quarterly Journal of Economics, 2009, 124 (1), 197-231.

Casari, Marco and Davide Dragone, “Choice reversal without temptation: A dynamic experiment on time preferences,” Journal of Risk and Uncertainty, 2015, 50 (2), 119-140

Chabris, Christopher F., David Laibson, and Jonathon P. Schuldt, “Intertemporal Choice,” in Steven N. Durlauf and Larry Blume, eds., The New Palgrave Dictionary of Economics, London: Palgrave Macmillan, 2008.

Chang, Tom Y., Wei Huang, and Yongxiang Wang, “Something in the Air: Pollution and the Demand for Health Insurance,” Review of Economic Studies (forthcoming)

Charness, Gary and Uri Gneezy, “Incentives to Exercise,” Econometrica, May 2009, 77 (3), 909–931.

Conlin, Michael, Ted O’Donoghue, and Timothy J. Vogelsang. 2007. “Projection Bias in Catalog Orders.” American Economic Review, 97(4): 1217-1249.

Cubitt, Robin P. and Daniel Read, “Can Intertemporal Choice Experiments Elicit Preferences for Consumption,” Experimental Economics, 2007, 10 (4), 369-389.

Ericson, Keith M. “On the Interaction of Memory and Procrastination: Implications for Reminders, Deadlines, and Empirical Estimation.” Journal of the European Economic Association, 2017, 15(3), 692-719.

Eyster, Erik, “Rationalizing the Past: A Taste for Consistency,” London School of Ecnomics mimeo, 2002.

Fang, Hanming and Yang Wang, “Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions,” International Economic Review, 2015, 56: 565–596.

Halevy, Yoram, “Time Consistency: Stationarity and Time Invariance,” Econometrica, 2015, 83(1): 335-352.

Kaufmann, Marc. “Projection Bias in Effort Choices”, Working Paper 2017.

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

Laibson, David, “Golden Eggs and Hyperbolic Discounting,” Quarterly Journal of Economics, 1997, 112 (2), 443-477.

Levy, Matthew, “An Empirical Analysis of Biases in Cigarette Addiction,” Working Paper, 2010.

Loewenstein, George, Ted O’Donoghue, and Matthew Rabin. (2003): “Projection Bias in Predicting Future Utility,” Quarterly Journal of Economics, 118 (4): 1209-1248.

Mahajan, Aprajit and Alessandro Tarozzi, “Time Inconsistency, Expectations and Technology Adoption,” Working Paper (2011).

McClure, Samuel, Keith Ericson, David Laibson, George Loewenstein, and Jonathan Cohen, “Time discounting for primary rewards,” Journal of Neuroscience, 2007, 27 (21), 5796–5804.

O’Donoghue, Ted and Matthew Rabin. 1999a. “Doing It Now or Later,” American Economic Review, 89(1): 103-124.

O’Donoghue, Ted and Matthew Rabin. 1999b. “Incentives for Procrastinators”, Quarterly Journal of Economics, 114(3), 769-816.

O’Donoghue, Ted and Matthew Rabin. 2001. “Choice and Procrastination,” Quarterly Journal of Economics, 116(1): 121-160.

O’Donoghue, Ted and Matthew Rabin, “Present Bias: Lessons Learned, and To Be Learned,” American Economics Review, Papers and Proceedings, forthcoming, June 2015.

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

Rabin, Matthew. “A Perspective on Psychology and Economics,” European Economic Review, 2002, 46, 657-685.

Read, Daniel and Barbara van Leeuwen, “Predicting Hunger: The Effects of Appetite and Delay on Choice,” Organizational Behavior and Human Decision Processes, 1998, 76 (2), 189-205.

Schilbach, Frank, “Alcohol and Self-Control: A Field Experiment in India”, mimeo, 2017.

Shapiro, Jesse. “Is there a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle,” Journal of Public Economics, 2005, 89, 303-325.

Skiba, Paige and Jeremy Tobacman, “Payday Loans, Uncertainty and Discounting: Explaining Patterns of Borrowing, Repayment, and Default,” Vanderbilt Law and Economics Research Paper No. 08-33, August 2008.

Strotz, Robert H., “Myopia and Inconsistency in Dynamic Utility Maximization,” Review of Economic Studies, 1956, 23, 165-180.

Toussaert, Séverine, “Eliciting temptation and self-control through menu choices: a lab experiment,” Working Paper (2015).