Short-Term Time Discounting of Unpleasant Tasks*

Ned Augenblick†

July 20, 2018

Abstract

Experimental subjects repeatedly state their preferences about the completion of unpleasant tasks for different monetary wages over the course of a week. Subjects desire to complete 86% fewer tasks when work is imminent versus one week away. Preferences change steeply close to work: one-third of the change occurs within a few hours and another third within a day of work. The full weekly discount curve is structurally estimated at .94, .91, and .87 a few hours, one day and one week from work. Common discount functions have difficulty accounting for the initial steepness and later flatness of the curve.

*This experiment was born out of similar work with Matthew Rabin, who also provided comments. Stefano DellaVigna provided extremely valuable early encouragement. I have benefited from discussions with Aaron Bodoh-Creed, Supreet Kaur, David Laibson, Ted O'Donoghue, Charles Spenger, Alex Steiny, Daniel Svirsky, and Séverine Toussaert, as well as audiences at the University of California, Berkeley and the SITE experimental conference. This experiment was preregistered on the AEA Registry as #1061. This research was approved by the UC Berkeley IRB.

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

1 Introduction

This paper presents an experiment designed to elicit a detailed estimate of the short-term time-discounting curve. Throughout a week, subjects are repeatedly asked how many unpleasant transcription tasks they want to complete at the end of the week given different piecewise wages. When the work is a week away, subjects choose to complete an average of 42 tasks. These tasks preferences drop gradually throughout the week to 40, 38, and 36 tasks when work is one day away, one hour away, and imminent, respectively. A structural estimation yields an estimated discount function across the week that follows a similar pattern to the raw task data: the function is around 0.87 a week away from work, with around a third of the drop occurring in the first few hours from work and another third in the first day. The data is then used to structurally estimate and test the fit of a variety of common parametric discounting curves, which commonly have difficulty matching the large change in choices over the week. Finally, appending data from Augenblick & Rabin (2018) (henceforth referred to as "AR"), a similar experiment including comparable work decisions between 1-3.5 weeks away, suggests that there is little additional change in the discount factor after one week.

The main contribution of this paper is a precise estimate of the short-term discount function, which is beneficial for four reasons. First, the findings can discipline and test the myriad theoretical proposals of the shape of the discount function, particularly given that the differences across theories are often more pronounced as the time of consumption approaches. Second, the result can inform a vein of literature that looks for specific heuristic, evolutionary, uncertainty-based, and perceptual explanations for the origin of discount functions that might explain decisions in intertemporal tradeoffs, particularly as these commonly focus on the standard hyperbolic form that appears at odds with the experimental results.[1] Third, the findings might help refine theoretical behavioral predictions given non-exponential discounting – such as procrastination and commitment choices – which commonly focus on the quasi-hyperbolic model due to its simplicity. Broadly though, the experiment suggests that this assumption is not so far from reality, particularly when the timing of decisions occurs on a daily or weekly scale. Finally, the shape of the curve provides rough guidance on the empirical effect of using different time definitions for "now" and "later" to estimate the quasi-hyperbolic parameter β. For example, the estimates suggest that providing consumption coded as "now" in a few hours rather than immediately will decrease the estimated parameter by about a third, and that providing "later" consumption more than a few days away will largely remove concerns of contamination by a present-bias effect.

Interestingly, although there appears to be a large amount of movement in the discount function

[1] Examples of this type of analysis include Sozou (1998), Azfar (1999), Weitzman (2001), Read (2001), Rubinstein (2003), Halevy (2004), Dasgupta & Maskin (2005), Farmer & Geanakoplos (2009), Read, Frederick & Scholten (2013), Ericson *et al* (2015), and Gabaix & Laibson (2017).

within a week, there are few studies that attempt to characterize the function in this interval: Read *et al* (1999) study movie preferences on the order of days, finding dynamically-inconsistent preferences for viewing "high-brow" movies on future evenings but not on the present evening; McClure *et al* (2007) and Brown *et al* (2009) examine preferences of thirsty subjects for juice delivered a points within 25 minutes, both finding that subjects exhibit non-exponential discounting driven by a significant preference for instantaneous consumption[2]; Kaur, Kremer, & Mullainathan (2015) study the effort of Indian data-entry workers, finding that they generally escalate effort over the week as payday approaches[3]; and finally, Balakrishnan, Haushofer, & Jakiela (2016) show that Kenyan subjects exhibit present bias for immediate payments via a mobile phone, which they contrast with Andreoni & Spenger (2012), who use the same methodology to find that American students exhibit little present bias given a delay of a few hours.[4] The paucity of research might be explained by the standard use of time-dated monetary payments to infer time preferences. Although this methodology has been criticized due to the unknown mapping from the payment timing to the timing of primary consumption (Cubitt & Read (2007) and Chabris, Laibson & Schuldt (2008)), others argue that "narrow bracketing" might cause subjects to treat money like consumption (Halevy (2015) and Balakrishnan, Haushofer, & Jakiela (2016)). However, the criticism becomes particularly sharp when focusing on very short time frames: it is challenging to imagine a monetary-choice experiment with non-credit constrained subjects that could credibly identify time preferences on the order of hours (if not days).

The paper proceeds with the experimental design (Section 2), a model (Section 3), the results (Section 4) and a final discussion (Section 5).

2 Experimental Design

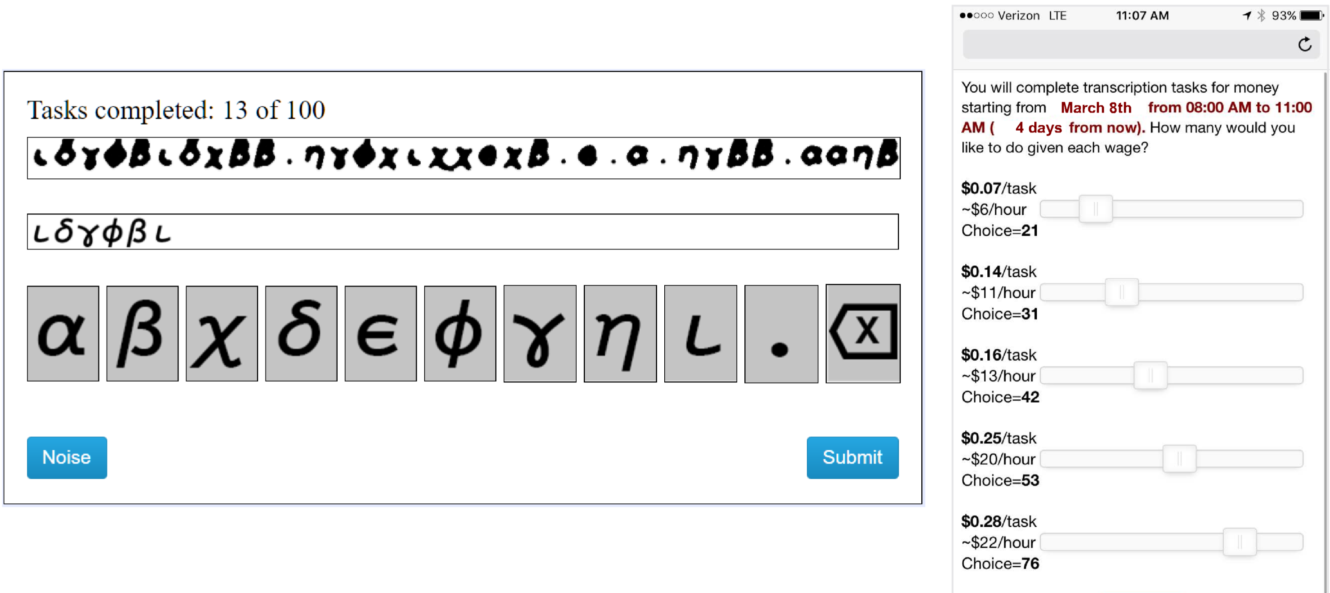
The experiment requires subjects to repeatedly state the number of tasks they wish to complete for different wages at the end of a week. The experimental task – previously employed in Augenblick, Niederle, and Spenger (2015) and AR – is the transcription of blurry Greek letters, with each

[2]Even given these very short delays, McClure *et al* (2007) and Brown *et al* (2009) estimate β at 0.5 and 0.6-0.7, respectively. These estimates are more extreme than those in this paper, potentially because very thirsty subjects display extreme present bias over liquid consumption. Another large difference is both papers assume (rather than empirically estimate) the exact curvature of the utility function.

[3]In that paper, Figure 3 plots earnings relative to payday, which is somewhat similar to the left panel of this paper's Figure 3, although the effect is identified off closer distance-to-payment inducing more work rather than closer distance-to-work inducing less work. Interestingly, in contrast with this paper's estimate of a relatively convex curve with substantially increasing effects in the final 24 hours as time-to-work approaches, that paper's plot appears concave, with a (non-significant) negative effect on payday. The difference might be driven by different subject pools, different identification strategies, or the environment (for example, workers with weekly income targets facing uncertainty might optimally vary average work across the week regardless of discounting).

[4]In addition, Hayden (2016) reviews time discounting studies in animals, which commonly find indifference between small immediate rewards and a reward twice the size in a few seconds.

Figure 1: Screenshots of the Task (left) and Mobile Decision Interface (right)



transcription taking about 45 seconds to complete. a screenshot of the task interface is shown in the left panel of Figure 1.[5] The experiment was preregistered on the AEA RCT Registry (henceforth referred to as the *PD*). The PD analysis plan is reproduced in Appendix A.18 and the experimental instructions in Appendix A.19.

After receiving a recruitment email describing the timeline and expected payments, subjects attend an initial laboratory session to learn the experimental details. During this initial session, subjects choose a day of the week for four future *work dates*, with the first date used for *practice* and the last three for *chosen work*. Then, subjects choose a three-hour *time window* for each of these dates, starting between between 5am and midnight.[6] The time window is the period in which, on each work date, the subjects must log in to the experimental website to complete the tasks for that week. Finally, subjects choose a large set of *free times* from 5am to 11pm for each day of the week during which they are available to spend a few minutes entering their work preferences using their mobile phone, and agree to also be able to answer questions around their work windows.

[5]The subject uses a mouse to point and click on the corresponding letters. The transcription must be within 7 character changes of the target text (in the experiment, the average submitted transcription was within 1-2 changes) or the subject must modify their answer. To make the task more onerous, an auditory "beep" randomly occurs every 5-15 seconds throughout the transcription process and a subject's transcription is erased if they do not click on the "noise" button shortly after the noise.

[6]A three hour window was chosen to balance the desire to pinpoint the exact timing of future work with the need to allow subjects some flexibility in login times. Subjects are required to place their time windows at different times of day over the three weeks. Without this constraint, subjects might have a very uneven distance-to-work distribution: for example, subjects with all work windows between 5am-8am will never express preferences when work is 0-8 hours away. Finally, subjects are asked to identify their expected login time within each time window to study the endogeneity of the actual login time within the time window (Appendix A.9 exploits this information and suggests this endogeneity has virtually no impact on the results).

For the practice work date, subjects become familiar with the task and interface by remotely logging in to the experimental website and completing a total of 80 tasks, which amounts to about one hour of work. Subjects are removed from the experiment if they do not log in during their chosen time window or do not complete the tasks within two hours of logging in (the consequences of removal are discussed shortly).

After the practice work date is completed, subjects repeat the following protocol for the next three weeks. Throughout each week, subjects receive around nine text messages with a mobile website link to the interface in the right panel of Figure 1, which presents a set of five decisions about *supplemental* tasks to be completed at the end of the week.[7] In each decision, the subject uses a slider bar to choose the number of tasks (between 0 and 100) that she would like to complete for a given wage, which varies from \$0.01/task to \$0.31/task (around \$1/hour to \$25/hour).[8] The slider buttons are initialized on the middle of the slider bars and subjects must click and drag each slider button to make a decision before submitting. If subjects do not answer at least 80% of the decision sets within one hour of the text across the entire experiment, they are removed from the experiment.[9]

At the end of each of these three weeks, the subject logs in to the experimental website using a computer.[10] After logging in, the subjects are first directed to complete a final work decision set on their phone. The subject then returns attention to the computer, where all of the decisions from the week (around 45 decisions) are collected and displayed. Then, one is randomly chosen as the unique *decision-that-counts (DTC)* for that week. The subject must then complete the number of supplemental *tasks* in the DTC as well as ten *mandatory* tasks (which are required to ensure that at least some tasks are completed on each date, eliminating any fixed cost associated with completing more than zero tasks). As above, if the subject does not log in or does not complete all tasks within the two hours of logging in, the subject is removed from the experiment.

Subjects that complete the experiment are paid for their participation in three separate payments following the experiment. Each payment contains a completion payment of \$20 plus the supplemental earnings associated with the respective DTC. The payment is by *virtual Visa* sent through email, which works like a Prepaid Visa card and can be used immediately at any online retailer. Each payment is made exactly 4 weeks from the time that the decisions in the associated

[7] The histogram of message timing relative to work time is shown in the right panel of Figure 3. The timing was chosen to have (1) higher frequency as the work window approaches, (2) no two messages occurring within an hour, and (3) diversity in distance from the work window within and across subjects, and (4) no violations of the subject's stated free times.

[8] Sets of task decisions were presented in order of increasing wage (rather than randomly, as in AR) to reduce the cognitive cost of completing many decisions sets throughout the week.

[9] Subjects answered 89% of the text messages and only one subject was removed. There was no significant change in the percentage of texts answered over the three weeks.

[10] Before all work dates, subjects receive both a reminder email and text message both 24 hours and one hour prior to the start of their chosen time window.

Table 1: Summary of the Experiment

	Timing	Method	Description
Initial Lab Session	Day t<0	Laboratory	Discuss experiment, choose time windows
Week 1: Practice Tasks	Day 0 window	Website	10 mandatory tasks, 70 supplemental tasks
Week 2: Task Questions	Days 1-7	Mobile Phone	Task decisions for 5 wages (~9 times)
Week 2: Complete Tasks	Day 7 window	Website	10 mandatory tasks, suppl. tasks in *DTC*
Week 3: Task Questions	Days 8-14	Mobile Phone	[as in week 2]
Week 3: Complete Tasks	Day 14 window	Website	[as in week 2]
Week 4: Task Questions	Days 15-21	Mobile Phone	[as in week 2]
Week 4: Complete Tasks	Day 21 window	Website	[as in week 2]
Payment for week 2	Week 2 DTC+28	Virtual Visa	Week 2 total wage + $20 (if not removed)
Payment for week 3	Week 3 DTC+28	Virtual Visa	Week 3 total wage + $20 (if not removed)
Payment for week 4	Week 4 DTC+28	Virtual Visa	Week 4 total wage + $20 (if not removed)

DTC were made.[11] If subjects are removed from the experiment for any reason, they do not receive any completion payments, but do receive the supplementary earnings made prior to being removed with the exact same timing described above. A summary of the experimental structure is displayed in Table 2.

110 subjects from the UC Berkeley Xlab subject pool were recruited for the experiment, which took place across 6 sessions on February 17, 2016. As stated in the PD, this number was chosen such that the expected number of subjects who continued with the experiment once learning the details and practicing the task would be approximately 100. In fact, 11 subjects dropped out prior to entering the main three-week portion of the experiment. Of these 99 subjects, 90 received the full $60 completion payment, in line with expectations for a multi-week study.[12]

In the analysis, I first estimate individual cost parameters while assuming a shared time-discounting function, and then allow all parameters to vary arbitrarily across individuals. Following the PD (and Augenblick, Niederle, and Spenger (2015) and AR), the analysis in the main paper removes subjects for which individual estimations fail. In the most obvious example, nine of the subjects have little or no variation in decisions across wages and time. In this case, the wage variation in the experiment simply does not allow for the identification of these subjects' time discounting preferences. In other cases, subjects give answers that are hard to understand given classical models

[11]As discussed in the next section, this timing assures that the distance to the potential monetary payment is the same at every decision point the experiment, which removes concerns that decisions are affected by slight differences in the distance to payment. Of course, if people treat money as fungible, this payment-timing difference should have virtually no impact on decisions. In fact, even if people treat monetary payments as time-dated consumption, the results of this paper suggest that decisions will still be largely unchanged given that all payments are more than a few weeks away.

[12]Three and five subjects dropped out in weeks 2 and 3, respectively. One subject was removed for making very few work decisions. The average total payment for subjects that completed the experiment was around $85.

of optimization. In the end, I remove 21 subjects, each of which are discussed and categorized in Appendix A.2. Crucially, the results which can be performed with all subjects do not change in any meaningful way when all subjects are included, as demonstrated in detail in Appendix A.7.

3 Theoretical and Empirical Model

To understand the experimental decisions, consider a model similar to that in AR in which an agent with separable preferences over monetary payments and effort who, at *decision time* k, chooses a number of tasks (an effort level) e to complete at fixed *work time* t given a per-task wage w received at *payment time* $k+T$.

The agent's disutility from effort is captured by the increasing and convex function $C(e)$, which is discounted using the function $D(t)$. The agent's indirect utility from wages is more complicated, as it represents the total discounted utility derived from the consumption stream attained from the payment (see AR for a longer discussion of this issue). I model this function as two multiplicative components: a utility function from wages $U(e \cdot w)$ and a discount function D_m.[13]

Given these assumptions, the agent desires to complete effort:[14,15]

$$e^* = \arg\max_e \; D_m(T) \cdot U(e \cdot w) - D(t-k) \cdot C(e). \tag{1}$$

Identifying the effort discounting function requires further structural assumptions about the monetary utility and effort cost functions. Following the PD, the utility function is assumed to be quasi-linear in money and the cost function component is parameterized as a power function $\frac{1}{\varphi\gamma}(e+10)^\gamma$, with the particular form chosen such that the first order condition is simple.[16] Given this, and matching the experiment by requiring the agent to complete 10 mandatory tasks prior to this chosen effort, equation (1) can be rewritten as:

[13] D_m captures the discounting effect of changing payment timing, which depends on the agent's conversion of money into a stream of consumption. In the experiment, the timing to payment is constant, which will render D_m inconsequential to the estimation of D.

[14] The experiment is designed to be robust to differing levels of agent sophistication. To address the issue that a sophisticated agent might strategically choose to not express her preferences e^* in Equation (1) for fear of reneging in the future, the experiment includes a heavy penalty – the loss of the $60 completion payment – for failure to complete previously-stated preferences.

[15] This formulation carries the implicit assumption of time separability across weeks. This assumption appears to be reasonable: later effort choices do not have a significant relationship with previous effort choices instrumented by the randomly-assigned previous wage.

[16] The assumption of quasi-linear monetary utility is discussed and empirically validated in AR. For robustness, Appendix A.11 modifies the model to include monetary curvature and finds virtually no impact on the smoothed discounting curve. Additionally, Appendix A.12 models two alternative cost-curve specifications and similarly finds very little effect on the smoothed discounting curve.

$$e^* = \arg\max_e \; D_m(T) \cdot (e \cdot w) - D(t-k) \cdot \frac{1}{\varphi \cdot \gamma}(e+10)^{\gamma}. \tag{2}$$

Taking the first-order condition of (2) with respect to e and solving for e yields the predicted choice e^* given the cost parameters, the discount functions, and experimental variation in w and k:

$$e^* = (\frac{D_m(T) \cdot \varphi \cdot w}{D(t-k)})^{\frac{1}{\gamma-1}} - 10. \tag{3}$$

Recall that $D_m(T)$ is constant throughout the experiment as the monetary payment is always made T periods from the time of the decision k. Consequently, defining the constant "discounted" slope parameter $\varphi_D \equiv D_m(T) \cdot \varphi$ allows for the substitution:[17]

$$e^* = (\frac{\varphi_D \cdot w}{D(t-k)})^{\frac{1}{\gamma-1}} - 10. \tag{4}$$

To rationalize differences between the empirical data and this predicted effort, observed effort e is assumed to be distributed around this predicted level of effort with a normal error term ε with mean 0 and standard deviation σ, an error form that was specified in the PD. Then, as effort decisions are censored at 100 tasks, a Tobit correction is applied to account for the possibility that the tangency condition implied by (3) is violated, leading to a likelihood function of observing observation j of effort level e_j as:

$$L^{tobit}(e_j) = \mathbf{1}(e_j < 100)\phi\left(\frac{e_j^* - e_j}{\sigma}\right) + \mathbf{1}(e_j = 100)\Phi\left(\frac{e_j^* - 100}{\sigma}\right), \tag{5}$$

where ϕ is the standard normal probability density function and Φ the corresponding cumulative density function. Standard techniques are used to find parameters that minimize the sum of this likelihood over all observations.

It is straightforward to parameterize the function D to estimate parameters associated with common time-discounting functions, such exponential, hyperbolic, generalized hyperbolic, or quasi-hyperbolic, by setting $D(t-k)$ as δ^{t-k}, $(1+\kappa \cdot (t-k))^{-1}$, $(1+\kappa \cdot (t-k))^{-\frac{\alpha}{\kappa}}$, or $\beta^{\mathbf{1}(k=t)}\delta^{t-k}$, respectively.[18] In the latter model, no discount occurs on consumption occurring "now," while all other consumption is discounted by $\beta\delta^{t-k}$. But, how soon is "now?" The empirical section both

[17] Intuitively, φ is an "exchange rate" between dollars and effort, and D_m is the discount on that rate for dollars received in the future. Constant distance-to-payment disallows separate identification of these (nuisance) parameters, but permits cleaner identification of the main object of interest, D. Conversely, AR estimates φ through the use of a parametric assumption of the shape of D_m and experimental variation in the distance-to-payment.

[18] These definitions are mostly consistent with past literature, although "hyperbolic discount function" is unfortunately used to refer to different functions and sometimes to any function with a decreasing discount rate. The functions can be seen in Hernnstein (1961) and Mazur (1987); Harvey (1986) and Prelec & Lowenstein (1992); Phelps and Pollak (1968), Laibson (1997), and Rabin & O'Donoghue (1999a , 1999b). The PD notes a plan to estimate the fixed-cost model in Benhabib *et al* (2010), but this

estimates a model in which only the time immediately prior to work is defined as "now," and a model in which any time prior to an additionally-estimated parameter η is considered "now." Note that, given the discontinuity between "now" and "later," the parameter η can only be identified up to a small interval in a finite dataset.

4 Results

4.1 Relationship between Tasks, Wages, and Distance to Work

The main sample consists of 8,875 observations from 1,775 decision sets from 79 subjects. I first examine the basic relationships in the data between task decisions, wages, and time to work. First, the left panel of Figure 2 presents the raw relationship between task decisions and wages. The standard errors used to construct the 95% confidence intervals are clustered at the subject level, as in every graph, statistic, and table throughout the paper. The relationship is nearly monotonic but somewhat noisy, given the large amount of subject heterogeneity in task decisions.[19,20,21] When accounting for this heterogeneity by instead using the residuals from a regression of tasks decisions on subject fixed effects, the relationship becomes universally monotonic and the standard errors are reduced by around 30%. The relationship suggests that, at least at an aggregate level, subjects understood the main tradeoff between effort and wages in the experiment.[22] Interpreted through the lens of the model, the somewhat linear relationship suggests that the marginal cost from effort is somewhat linear in the number of tasks.

The focus of this paper is examining how these decisions change as the time of work approaches. The middle panel of Figure 2 separates the residuals into three categories: (1) decisions that occur immediately before work began, (2) decisions that occur prior to work but within 24 hours, and (3) decisions that occur more than 24 hours away. Although noisy, there is visual evidence of a drop in task decisions across wages as the time to work approaches. This relationship is more pronounced in the right panel of Figure 3, which plots the cumulative distribution for the three categories with wage fixed effects added to allow aggregation of observations across wages. The three distributions are generally bell-shaped and exhibit a near first-order stochastic dominance relationship. The Kolmogorov–Smirnov test strongly rejects equality in all three pairwise distributional comparisons

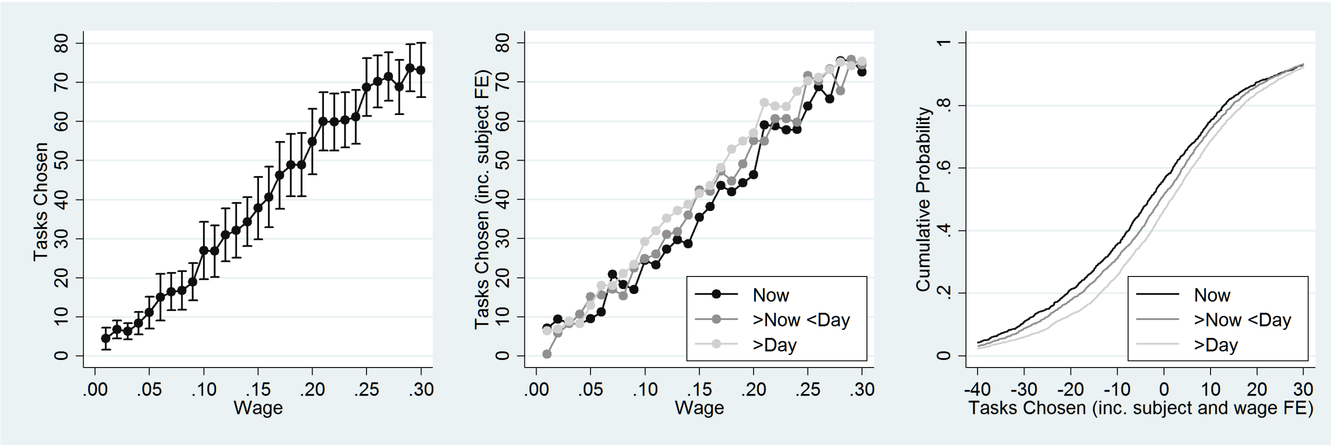
[19]Subject fixed effects account for 47% of the variation in task decisions not explained by wages.

[20]The PD specifies the use of ten wage bins for visual ease. I instead present the full graph as it provides more information with little cost. Appendix A.17 presents the smoother aggregated ten-bin graph .

[21]Subjects choose the ceiling of 100 tasks in 18% of decisions and choose to complete no tasks in 23% of decisions. This is in line with AR and is a consequence of the intentional wide range of wages used to induce choice variation across subjects with different task-money tradeoffs.

[22]On an individual level, 69 of 78 subjects (and 15 of the 21 non-included subjects) had fewer than five total non-monotonicities within a decision set (given an average of around 92 violation opportunities).

Figure 2: Task Decisions Changing with Wages and Over Time



Note: This figure presents the relationship between wages and task decisions. The left panel plots the average task decision for each wage, with the 95% confidence intervals constructed using standard errors clustered at the subject level. The middle panel plots, for three timing categories, the residuals from a regression of task decisions on subject fixed effects (with the residuals at a wage of \$0.01 normalized to equal the average task decisions at that wage). The right panel presents the cumulative distribution, for the same three timing categories, of the residuals from a regression of task decisions on subject and wage fixed effects, with the x-axis constrained to include around 90% of the data for visual ease.

(all $p \leq 0.001$).

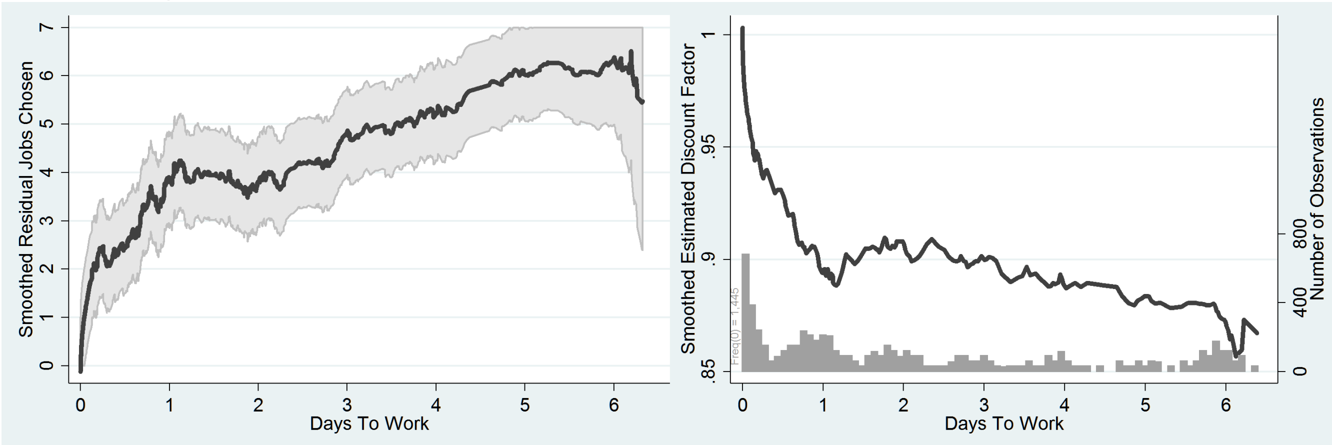
Regressing task decisions on these time periods (while including fixed effects for each wage, subject, and week) confirms these conclusions: in comparison to the first category, subjects choose to complete 2.4 more tasks in the second category and 5.0 more tasks in the third category. Both statistics are significantly different from zero ($Z = 2.91, p = 0.005; Z = 5.70, p < 0.001$) and significantly different from each other ($F(1, 77) = 20.09, p < 0.001$).[23,24] Although there is a relatively large amount of noise in the individual-level data – from decision noise, changing constraints both within and across weeks, and different wages – the same trends largely hold for the majority of subjects: in comparison to the first category, 61% choose more average tasks in the second category and 67% choose more in the third, which rise to 63% and 74% when adding an individual linear control for wages. In a univariate linear regression of task decisions on time-to-work, 78% of subjects exhibit a positive slope and 60% exhibit both a positive slope and negative quadratic relationship, which weakly rise to 81% and 60% when adding an individual linear control for wages.

The left panel of Figure 3 presents an estimate of the change in decisions over time with less

[23]The statistics do not change dramatically when focusing on the last two work dates, adjusting to 2.0 ($Z = 2.18, p = 0.033$) and 3.9 ($Z = 3.89, p < 0.001$) respectively, and remaining statistically different from each other ($F(1, 76) = 8.45, p = 0.005$).

[24]Appendix A.8 presents the results using different controls: the results are generally stable although noticeably noisier without the inclusion of subject fixed effects.

Figure 3: Smoothed Task Decisions and Estimated Discount Factors over Time



Note: The left panel shows the smoothed change over time in the residuals of a regression of task decisions on wage, subject and week fixed effects. The right panel presents the smoothed change over time of 250 discount factors estimated using the structural model in which the cost curve parameters vary arbitrarily across subjects and weeks, dropping the most extreme 5% of factors. In addition, the right panel includes a histogram showing the frequency of decisions at each time point, omitting the bar representing the 1,445 immediate decisions. In both graphs, the 0.5% of decisions after 6.2 days are omitted as the estimates become extremely noisy.

reliance on parametric assumptions. To construct this graph, the residuals of a regression on wage, subject and week fixed effects are smoothed over the distance to work using a symmetric nearest-neighbor running-line smoother implemented using the Stata *running* command (normalizing the smoothed residuals to be equal to zero when the distance to work is zero). The bandwidth is intentionally chosen to be small to capture the precise details in the curve close to work, at the cost of added noise. The curve suggests that, when work is around a week away, subjects choose an average of around six tasks more than when work is immediate. The entire curve is broadly increasing with the distance to work, with a large increase (to around 2 tasks) in the first few hours and another significant increase (to around 4 tasks) in the first 24 hours.

Given the many implicit decisions used to create the this figure, Section 4.4 outlines a large number of robustness checks discussed in the Appendix. Additionally, in Appendix A.5, I append the data from AR, which asks similar task questions one week to three-and-a-half weeks from the work time, and find little additional increase in task decisions after the first week.

4.2 Smoothed Discounting Curve

The structural model allows for a similar non-parametric estimation of the discount function from task decisions. To create this curve, I first jointly estimate 250 discount factors across the week along with the other parameters in the model, where I allow the cost curve parameters to vary across subjects and weeks.[25] Then, I drop 5% of the outliers and smooth these discount factors across the week using the same method as in the previous section. The omission of outliers is important as, given the number of discount factors estimated, a few of the factors are extremely high (greater than 10) or low (lower than 0.1). The results are shown in the right panel of Figure 3, which also includes a histogram of the number of decisions at different points of time. Broadly, the discount function looks like an inverted and rescaled version of the residual task curve discussed above, which is perhaps not surprising given the apparently linear relationship between task number and marginal cost. The curve suggests that subjects discount work that is a few hours away by 0.94, 24 hours away by 0.91, and around a week away by 0.87.

Appendix A.1 again outlines the many robustness checks for this figure in the Appendix. Additionally, as with the task decisions, Appendix A.5 contains the estimated curve after appending data from AR. Following the conclusions above, the discount function appears largely flat after one week, dropping to only .86 after three weeks.

4.3 Fit and Parameters of Common Discount Functions

Table 2 presents estimates given a variety of common parameterizations of the discount function. In each specification, the cost curvature is allowed to vary freely across work weeks and individuals and the cost slope is allowed to vary freely across weeks, with the table containing the average of these estimates.[26] The estimates are extremely stable across all specifications, with the results in column (1) implying the average marginal cost at the completion of 25, 50, 75, and 100 tasks is \$0.09, \$0.15, \$0.21, and \$0.27, respectively, with a total cost of \$1.51, \$4.59, \$9.19, and \$15.28 at these points.

Column (1) assumes an exponential parameterization of the discount function. The estimate of the daily exponential discount factor $\hat{\delta}$ is 0.983, which is strongly statistically significantly different from 1. The estimate must account for the relatively large drop in task decisions over the week, leading to a low yearly discount factor of around 0.2%. Columns (2)-(4) estimate parameters from the quasi-hyperbolic model, sequentially relaxing the constraint that the exponential component of discounting be fixed at 1 and that the definition of "now" only include decisions made at the time of work. The estimate of the present-bias parameter $\hat{\beta}$ varies from $0.919 - 0.946$, which – while marginally to highly statistically-significantly different from 1 – is certainly closer to 1 than previous estimates from similar studies.[27] This discrepancy arises from the timing of decisions used to

[25] 250 is the largest number of estimated discount factors for which all specifications converge. This arbitrary choice does not have a large impact on the results: Appendix A.14 replicates the main figure using 25, 50, 100, and 750 estimated discount factors.

[26] Appendix A.3 contains summary statistics when each individual's parameters are estimated separately.

[27] Traditional discounting experiments with monetary payments find persistent *magnitude effects* (Thaler (1981)), in which discounting apparently declines with the payment amount. One explanation (following Benhabib *et al*

estimate "now" versus "later" choices: in most studies, "later" decisions are very temporally distant from the time of consumption. In this study, "later" decisions are all less than a week away, including many decisions where the time of work is mere hours away. Given that the discount function is smooth rather than discontinuous, the estimate of the discontinuous present-bias parameter is dependent on the timing of the decisions in a study, an important point across all studies of present bias. Column (4) shows that the estimate of the optimal discontinuous division between "now" and "later" $\hat{\eta}$ is around one hour from work. However, this estimate is sensitive to the precise specification, although it is consistently less than one day from work. Finally, Columns (5)-(6) estimate a hyperbolic and generalized hyperbolic model. The single hyperbolic parameter $\hat{\kappa}$ is highly statistically significant, while neither of the two parameters in the generalized hyperbolic is close to significant. In fact, given the flexibility of the generalized hyperbolic curve, very different parameters can produce very similar curves, leading to a relatively flat mapping from parameters to likelihood.

The middle portion of the table reports the Akaike information criterion (AIC) and Bayesian information criterion (BIC) weights for each model, which are normalized measures of the relative fit of the different models that are derived from the model's log likelihood and number of free parameters.[28] The exponential model and hyperbolic model are strongly and almost equally disfavored, a problem which is exacerbated when including data from AR. This is perhaps not surprising: exponential and standard hyperbolic curves calibrated such that the discount factor at one week is 0.87 are extremely similar within the week (for example, the respective discount factors are 0.9992 and 0.9991 for work that is one hour away, and 0.9803 and 0.9791 when one day away). When the exponential component is not fixed, the quasi-hyperbolic models perform relatively well, although the BIC and AIC measures differ on the relative benefit of allowing the definition of "now" to vary. The generalized hyperbolic model performs worse than the standard model of quasi-hyperbolic discounting, although this relationship reverses when including data from AR. However, one reason to favor the quasi-hyperbolic model is its success given a relatively rigid form. Although the model is seemingly much more restrictive and less complex than the generalized hyperbolic function, both the AIC or BIC consider the models as equally flexible because complexity is quantified as the number of free parameters (for a brief discussion of this issue, see Vandekerckhove *et al* (2014)).

(2010)) is a high fixed transaction cost associated with future payments, which they estimate at \$4. Here, there is little change in discounting as effort levels increase: separate estimates of β for low, medium, and high wages are 0.93, 0.93, and 0.91, respectively (with no significant difference (highest $Z = 0.21$, $p = 0.837$)). When a fixed cost is added to the estimation, it is only \$0.30 and leads to virtually no impact on the discounting estimates. This result is potentially due to design choices made to reduce and equalize transaction costs.

[28] Burnham & Anderson (2002) note that the AIC "may be interpreted as the probability that [a given model is] the best model for the sampling situation considered." The BIC is similar, although it penalizes the number of parameters in the model differently.

Table 2: Aggregate Results

	(1) Delta Only	(2) β-δ δ:=1	(3) β-δ -	(4) β-δ-η -	(5) Hyper-bolic	(6) General Hyperb.
Discount Factor δ	0.983	1.000	0.987	0.989		
	(0.004)	(.)	(0.004)	(0.004)		
Present Bias β		0.919	0.946	0.941		
		(0.030)	(0.031)	(0.026)		
Now (Hours) η		0.0	0.0	1.1		
		(.)	(.)	(.)		
Hyperbolic κ					0.018	113.522
					(0.005)	(425.370)
Gener. Hyper. α						1.96
						(6.56)
Cost Curvature γ	1.95	1.95	1.95	1.95	1.95	1.95
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Cost Slope φ_D	318	308	309	310	318	309
	(165)	(159)	(161)	(162)	(165)	(161)
Observations	8875	8875	8875	8875	8875	8875
Subjects	78	78	78	78	78	78
Log Likelihood	-26884.6	-26889.3	-26877.2	-26874.8	-26884.4	-26877.8
Akaike IC	53939.3	53948.6	53926.5	53923.6	53938.9	53927.6
Bayesian IC	54542.0	54551.3	54536.3	54540.5	54541.6	54537.4
AIC (weight)	0.000292	0.000003	0.172453	0.727475	0.000357	0.099419
BIC (weight)	0.032074	0.000300	0.546747	0.066458	0.039221	0.315199
$H_0(\hat{\hat{\delta}}=1)$	p<0.001		p= 0.003	p= 0.010		
$H_0(\widehat{\beta}=1)$		p= 0.007	p= 0.082	p= 0.022		
$H_0(\widehat{\kappa}=0)$					p<0.001	p= 0.790
$H_0(\widehat{\alpha}=1)$						p= 0.883

This table includes the aggregate structural estimations of different discount function specifications for our primary sample of 78 subjects. The cost function parameters are allowed to vary arbitrarily by subject and/or work week - the parameter presented is the average of these estimates. In all specifications, standard errors are clustered at the subject level.

4.4 Robustness

The Appendix contains a large number of robustness checks and additional analyses. The main results are relatively stable across these checks as summarized in Appendix A.1. Appendix A.4 estimates a variety of linear splines and piecewise exponential-discounting functions to estimate the discount curve. Appendices A.6, A.7, A.8, A.9, and A.10 contain many versions of the two curves in Figure 3 and of the estimates in Table 2 given different data samples, subject samples, use of fixed effects, measures of work timing, and smoothing methods. Appendices A.11, A.12, A.13, and A.14 contain additional versions of the estimated time discounting curves in Figure 3 and the estimates in Table 2 using a variety of different structural assumptions, such as the assumption of monetary curvature, different cost-function specifications, different error-term specifications, and different treatment of outliers.

5 Conclusion

The experiment was designed to create a detailed estimate of the short-term time discounting curve through the repeated elicitation of preferences to complete unpleasant tasks for different wages over the course of a week. Work decisions and implied discount rates change significantly across the week, particularly when less than a day from work, which is difficult to reconcile with exponential discounting. Interestingly, the standard hyperbolic discounting model – which is commonly used to explain non-exponential discounting across long time periods – is similarly poor at fitting these sharp short-term changes. Not surprisingly, the data does not show the discontinuity implied by the quasi-hyperbolic model, although the model performs well given the steep change in discounting within a day from work.

Although the design makes headway on a few of the many criticisms raised about previous time-discounting studies, there are still important issues left unresolved: positive consumption is potentially discounted differently than effort; there is likely fungible out-of-the-experiment effort that could change the in-experiment cost-of-effort, biasing the discounting curve upwards; analogous to precautionary savings, uncertainty potentially leads to lower work decisions when work is farther away, biasing the curve upwards; and planning-fallacy issues likely lead to fewer imminent tasks being chosen than desired due to unanticipated time constraints, biasing the curve downwards. There are many opportunities for future results to address these complications, particularly in understanding the complicated interconnection between time preferences and uncertainty.

6 References

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

Augenblick, Ned and Matthew Rabin, "An Experiment on Time Preference and Misprediction in Unpleasant Tasks", Reveiw of Economic Studies, forthcoming 2018

Augenblick, Ned, Muriel Niederle and Charles Sprenger, "Working Over Time: Dynamic Inconsistency in Real Effort Tasks", Quarterly Journal of Economics, August 2015

Azfar, Omar, "Rationalizing Hyperbolic Discounting." Journal of Economic Behavior & Organization 1999, 38(2): 245-52

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

Balakrishnan, Uttara, Johannes Haushofer, and Pamela Jakiela, "How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments," NBER Working Paper #23558, 2017

Benhabib, Jess, Alberto Bisin, and Andrew Schotter. "Present-bias, quasi-hyperbolic discounting, and fixed costs." Games and Economic Behavior 69.2 (2010): 205-223.

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.

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.

Cohen, Jonathan, Keith Ericson, David Laibson, and John Myles White, "Measuring Time Preferences," Working Paper 2017.

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

Dasgupta, Partha, and Eric Maskin. "Uncertainty and hyperbolic discounting." The American Economic Review 2005, 95(4): 1290-1299.

Ericson, Keith M., John Myles White, David Laibson, and Jonathan D. Cohen. "Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting does." Psychological science, 2015, 26(6): 826-833.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. "The Nature and Predictive Power of Preferences: Global Evidence," Working Paper 2017

Farmer, J. Doyne and John Geanakoplos "Hyperbolic Discounting is Rational: Valuing the Far Future with Uncertain Discount Rates," 2009 Cowles Foundation Discussion Paper 1719

Frederick, Shane. "Cognitive Reflection and Decision Making," Journal of Economic Perspectives, 2005 19 (4): 25–42.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. "Time discounting and time preference: A critical review." Journal of economic literature 40.2 (2002): 351-401.

Gabaix, Xavier and David Laibson, "Myopia and Discounting," Working Paper 2017

Halevy, Yoram. "Diminishing Impatience: Disentangling Time Preference from Uncertain Lifetime," 2004, Working paper, University of British Columbia

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

Halevy, Yoram, “Some comments on the use of Monetary and Primary rewards in the measurement of Time Preferences,” Discussion Paper, 2015

Harvey, Charles, “Value Functions for Infinite-Period Planning,” Management Science, 1986, 32(9): 1123-1139

Hayden, Benjamin Y. “Time discounting and time preference in animals: a critical review.” Psychonomic bulletin & Review, 2016, 23(1): 39-53.

Herrnstein, Richard, “Self-Control as Response Strength,” in Quantification of SteadyState Operant Behavior. Christopher M. Bradshaw, Elmer Szabadi and C. F. Lowe, eds. 1981

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.

Loewenstein, George, and Drazen Prelec. "Anomalies in intertemporal choice: Evidence and an interpretation.” The Quarterly Journal of Economics 107.2 (1992): 573-597.

Mazur, James, "An Adjustment Procedure for Studying Delayed Reinforcement,” Chapter 2 in M. L. Commons, J. E. Mazur, J. A. Nevins, and H. Rachlin, eds., Quantitative Analysis of Behavior: The Effect of Delay and of Intervening Events on Reinforcement Value (Hillsdale, NJ: Ballinger, 1987).

McClure, Samuel, Keith M 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. “Doing It Now or Later,” American Economic Review, 1999a, 89 (1): 103-124.

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

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

Phelps, Edmund S., and Robert A. Pollak. "On second-best national saving and game-equilibrium growth.” The Review of Economic Studiesm, 1968, 35.2: 185-199.

Read, Daniel. “Is time-discounting hyperbolic or subadditive?.” Journal of risk and uncertainty, 2001, 23(1) : 5-32.

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.

Read, Daniel, George Loewenstein, and Shobana Kalyanaraman. “Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic.” Journal of Behavioral Decision Making, 1999, 12(4): 257.

Read, Daniel, Shane Frederick, and Marc Scholten. “DRIFT: An analysis of outcome framing in intertemporal choice.” Journal of Experimental Psychology: Learning, Memory, and Cognition 2013, 39(2): 573.

Rubinstein, Ariel, “Economics and Psychology? The Case of Hyperbolic Discounting,” International Economic Review, 2003, 44 (4) 1207-16.

Sozou, Peter, “On Hyperbolic Discounting and Uncertain Hazard Rates.” Proceedings of the Royal Society B: Biological Sciences 1998, 265 (1409): 2Less015-20.

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).

Vandekerckhove, J., Matzke, D., & Wagenmakers, E. J. “Model comparison and the principle of parsimony,” In J. R. Busemeyer, Z. Townsend, J. Wang, & A. Eidels (Eds.), Oxford handbook of computational and mathematical psychology (pp. 125–144). Oxford: Oxford University Press, 2014.

Weitzman, Martin, “Gamma Discounting.” American Economic Review, 2001, 91(1): 260-71

FOR ONLINE PUBLICATION ONLY

A Online Appendix

A.1 Overview of Other Results and Robustness

This section summarizes the large number of robustness checks and additional analyses found in this Appendix. Appendix A.3 provides more detail on the individual estimates, and Appendix A.5 replicates the analysis when adding additional data from AR. Appendix A.4 estimates a variety of linear splines and piecewise exponential-discounting functions to estimate the discount curve, confirming the basic shape shown in Figure 3. For example, the linear spline with knots at 1 hour, 1 day, and 1 week, yields an estimated discount factor of .953, .925, and .875 at those time periods.

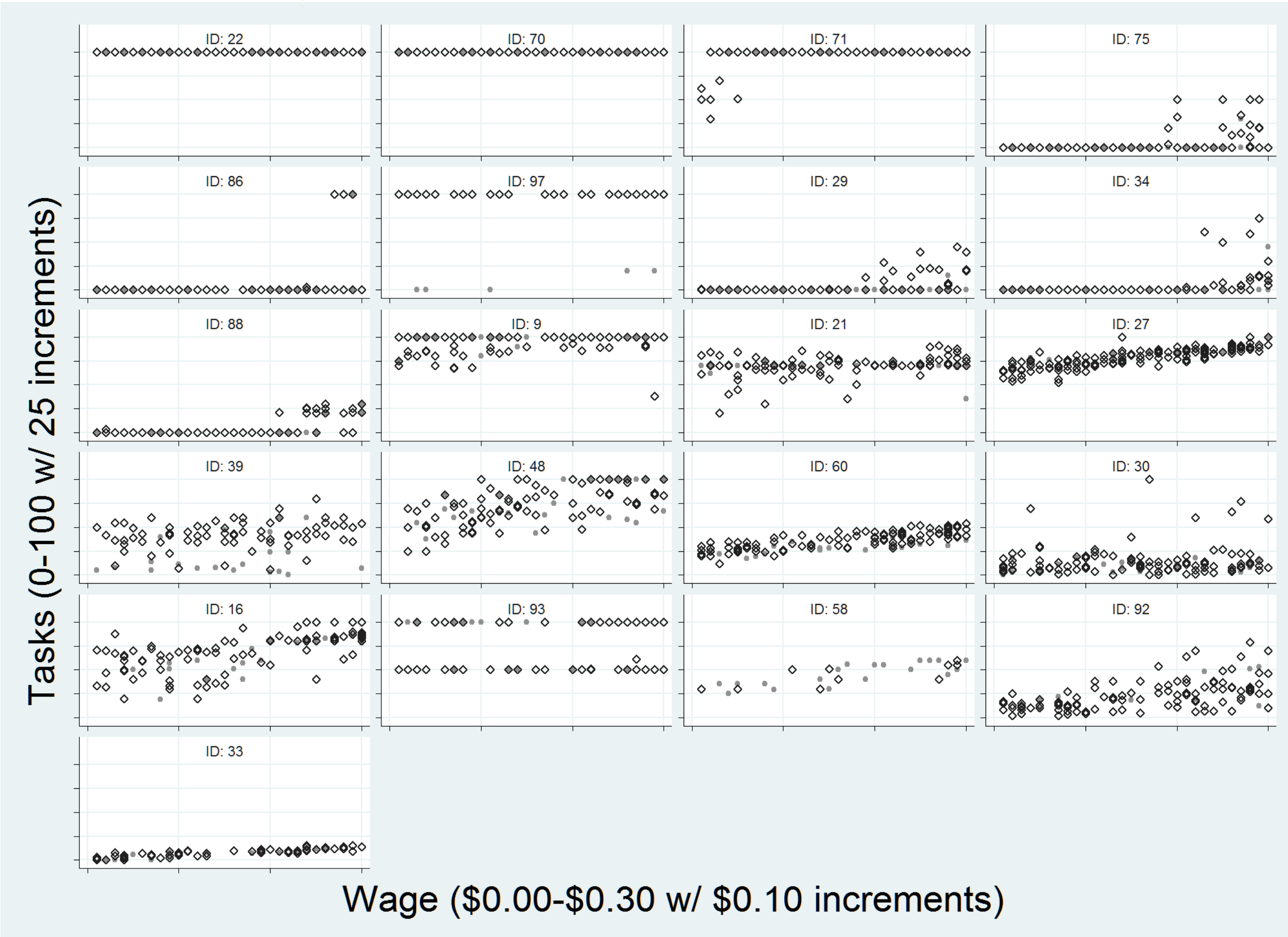
Appendix sections A.6, A.7, A.8, A.9, and A.10 contain many versions of the two curves in Figure 3 and of the estimates in Table 2 given different data samples, subject samples, use of fixed effects, measures of work timing, and smoothing methods. These analyses suggest that (1) the raw task curve appears shrunken by about 25% when focusing only on the final two work dates, but this appears due to a changing cost function as the estimated discount curve and parametric estimates are nearly unchanged; (2) the subject sample – such as the inclusion of all subjects or removing attritors – has little effect on either curve or the estimates; (3) the removal of subject fixed effects has a meaningful effect on all of the results given the heterogeneity in subject choices; (4) alternative measures of timing based on ex-ante elicited expected work time rather than actual work time has almost no impact on the curves or estimates; (5) changing the smoothing technique or smoothing parameters leads to either noisier or smoother curves that broadly follow the same pattern as the curves shown in the paper.

In addition, Appendix sections A.11, A.12, A.13, and A.14 contain additional versions of the estimated time discounting curves in Figure 3 and the estimates in Table 2 using a variety of different structural assumptions. These analyses suggest that (1) assuming even extreme curvature in the monetary utility function has almost no effect on the curve or estimates, (2) parameterizing the cost curve using a more flexible form has little effect on the curve or estimates, (3) different specifications of the error term can shrink or magnify the smoothed curves, with the largest effect (a 50% magnification) occurring when placing a normally-distributed error on the *log* of task decisions, (4) removing fewer outliers or using less discount factor estimates to construct the curve exaggerates the non-monotonic hump in the discount curve when work is about two days away.[29]

Appendix section A.15 presents correlations between individual behavior and eleven survey questions covering demographics, the cognitive-reflection task, and self-reported measures of time discounting and risk willingness. There is some evidence that older students and those with higher math SAT scores exhibit less severe discounting and that students with higher scores on the cognitive-reflection task exhibit more severe discounting. However, this evidence is fairly weak and the statistical confidence is not corrected for the large number of hypotheses tested.

[29]One timing issue is not well addressed in the Appendix. In the paper, subjects are implicitly assumed to receive all of the disutility from effort at the time that work starts. In reality, subjects receive disutility continuously over the course of 5-60 minutes, so that the time distance to the marginal task is endogenous to the number of tasks chosen. This modification significantly complicates the analysis with little apparent benefit given the always-small temporal distance between the beginning and end of work in this experiment.

Figure A1: Scatterplots of Decisions of Removed Subjects



Note: The main analysis focuses on a primary sample of 78 subjects. This graph shows scatter plots of task decisions (y-axis) given a wage (x-axis) of all of the removed subjects. The subjects are ordered by the general reason of their removal, which is discussed in the text. Immediate and future task decisions are represented by filled circles and hollow diamonds, respectively.

A.2 Subjects Removed From Main Sample

The main sample in the paper does not include 21 subjects. Figure A1 contains the scatter plot of all of the task decisions given different wages for each of these subjects, separately marking decisions made within a day of work.

These subjects can be broadly segmented based on the issue that causes their removal. Subjects 22, 70, 71, 75, 86, 97, 29, 34, and 88 have little or no variation in decisions across wages and time. For these subjects, the experiment simply provides no information about their time discounting preferences. For example, if a subject has a very high cost curve and does not want to perform any tasks for the offered wages, there is no way to know if preferences change given the timing of work without additional questions employing higher wages. Subjects 9, 21, 27, 39, 48, 60, and 30 give medium-level task decisions for very low wages (such as 50 tasks for $0.01/task) and don't raise the number of chosen tasks much as the wage rises (stating, for example, 60 tasks for $0.15/task). This pattern is difficult to rationalize with a smooth cost function as, in the example, it suggests that the subject's marginal cost is extremely low (under \$0.01/task) up until 50 tasks but then suddenly rises dramatically (to, in this example, \$0.15/task at 60 tasks). While this is difficult to rationalize with the theoretical model in the paper, it might be that these subjects experience little displeasure from completing the task, but continually only have a medium fixed amount of time to complete tasks. Subjects 16 and 93 have very erratic answers, with strange amounts of variation across wages and decisions sets. Subjects 58 and 92 lead to parameter estimations that are transparent outliers: A Grubbs' test identifies these observations as outliers with a p-value less than 10^{-9}. Finally, subject 33 states positive but extremely low task decisions for every wage, averaging only 8 tasks across wages, which causes the aggregate estimation to fail to converge.

A.3 Individual Estimations

Table A1 provides summary statistics of the results when each subjects' parameters are estimated individually. The table contains the median, mean and the standard deviation of the distribution of each parameter across individuals (the standard deviations should not be confused with the (unreported) average standard error of the estimates). Largely, the estimates are close to the aggregate estimates in Table 2, although large outliers in the generalized hyperbolic parameters and the "slope" parameter ϕ lead to large deviations. For the generalized hyperbolic specification in Column (6), many estimations do not properly converge and the parameters are highly variable when convergence is achieved, reinforcing the point that precise parameter identification is difficult for this model in smaller samples.

A.4 Linear Splines and Changing Discount Factors

The estimate of the discount factors for various linear splines are shown in the left panel of Table A.4, with the third column including data from AR. For example, the second column suggests that the best fit for the discount factors of a four-knot linear spline with knots at (1) the point of work, (2) one hour from work, (3) one day from work, and (4) one week from work are 1,.953, .925, and .875, respectively.[30]The estimate of the discount rates for various piecewise exponential functions are shown on the right panel of Table A.4. For example, the second column suggests that the best fit for the exponential daily discount factor between (1) the point of work and one hour from work, (2) one hour from work and one day from work, (3) one day from work and one week from work are 0.316, 0.969, and 0.990, respectively. Note that these discount rate estimates imply a discount factor of $0.316^{\frac{1}{24}} = .953$ an hour from work, $(0.316^{\frac{1}{24}})(0.969^{\frac{23}{24}}) = .925$ a day from work, and $(0.316^{\frac{1}{24}})(0.969^{\frac{23}{24}})(0.990^{6}) = .871$ a week from work, which closely mirrors the results from the second column in the left panel.

The results from the splines map very closely to the estimates in the main paper from smoothing discount factors. However, when adding data from AR, the smooth curve shows almost no drop from one to three weeks (from around 0.87 to 0.86), while the linear spline drops from 0.90 to 0.84. Still, the implied exponential discount factor is still apparently rising across the entire three weeks, to 0.997 between one and three weeks from work.

[30]The table does not report significance levels, which commonly border statistical significance at standard confidence levels. For example, the p-values of the tests that each set of bordering discount factors in the second column are the same are 0.17, 0.20, and 0.09, respectively.

Table A1: Individual Estimates

	(1) Delta Only	(2) β-δ δ:=1	(3) β-δ -	(4) β-δ-η -	(5) Hyper-bolic	(6) General Hyperb.
mean($\hat{\delta}_i$)	0.972	1	0.978	0.978		
median($\hat{\delta}_i$)	0.978	1	0.977	0.983		
sd($\hat{\delta}_i$)	(0.044)	(.)	(0.041)	(0.091)		
mean($\hat{\beta}_i$)		0.900	0.941	0.990		
median($\hat{\beta}_i$)		0.925	0.952	0.865		
sd($\hat{\beta}_i$)		(0.185)	(0.186)	(0.428)		
mean($\hat{\eta}_i$)				27.483		
median($\hat{\eta}_i$)				22.954		
sd($\hat{\eta}_i$)				(23.604)		
mean($\hat{\kappa}_i$)					0.043	68903499
median($\hat{\kappa}_i$)					0.024	8
sd($\hat{\kappa}_i$)					(0.071)	(4.1e+08)
mean($\hat{\alpha}_i$)						241423
median($\hat{\alpha}_i$)						0
sd($\hat{\alpha}_i$)						(1019555)
mean($\hat{\gamma}_i$)	2.073	2.076	2.074	2.055	2.084	1.938
median($\hat{\gamma}_i$)	1.974	1.998	1.981	1.978	1.987	1.963
sd($\hat{\gamma}_i$)	(0.757)	(0.748)	(0.755)	(0.761)	(0.759)	(1.459)
mean($\hat{\phi}_i$)	142722	130065	133027	174009	137219	21609
median($\hat{\phi}_i$)	298	311	301	276	317	251
sd($\hat{\phi}_i$)	(1063111)	(1006654)	(1041373)	(1423810)	(1031502)	(88646)
Observations	77	78	78	78	78	61

This table presents the statistics of the parameters for the individual structural estimations of different discount function specifications. sd($\hat{x}_i$) is the standard deviation of the distribution of individual estimates (not the average standard error). Additional subjects are removed from column (7) as the estimations did not converge for the generalized hyperbolic specification.

Table A2: Estimated Linear Splines and Piecewise Discount Rates

Discount Factor (Linear Spline)				Discount Rate (Piecewise Exponential Function)			
At work	1	1	1				
	(.)	(.)	(.)				
One Hour		0.953	0.962	>One Hour		0.316	0.395
		(0.034)	(0.030)			(0.271)	(0.295)
One Day	0.938	0.925	0.929	>One Day	0.938	0.969	0.965
	(0.022)	(0.029)	(0.027)		(0.023)	(0.024)	(0.023)
One Week	0.894	0.875	0.899	>One Week	0.991	0.990	0.993
	(0.027)	(0.034)	(0.027)		(0.006)	(0.006)	(0.005)
~Three Weeks			0.839	>~Three Weeks			0.997
			(0.049)				(0.003)

Note: This table presents the estimates of linear splines and piecewise exponential discounting functions given the basic structural parameterization used throughout the main paper.

A.5 Adding Data From Previous Experiment

As discussed in the main paper, AR run a similar experiment with the same task decision for the same wage using the same subject population. AR largely focus on decisions at the time of work or around 4-27 days from work. Figure A2 and Table A3 replicate those in the text when including the task-choice data from the main sample of 72 subjects in AR. So that the short-term data is still consistent with that in the main paper, the 9% of choices in AR that occur between 1-6 days of work are dropped (this has virtually no impact on the results). Finally, as in the main figures in the paper, the figures do not show the most-distant 0.5% of observations (between 22 and 27 days) as the smoothed estimates become volatile near the borders. If anything, during the omitted period, there is a steep drop in the task residuals (to around 2) and a steep rise in the estimated discount factor (to around 0.89), although there are many reasons to be very skeptical of the validity of these estimates.

Figure A2 suggests that the task decisions and estimated discount function are largely flat between weeks, with the discount factor at three weeks estimated at around 0.86. The estimates in Table A3 reflect the effect on the parametric estimates. Most notably, the estimated exponential discount factor δ rises (from 0.983 to 0.991) and the estimated standard hyperbolic parameter falls (from 0.018 to 0.010) to account for the little discounting across later weeks in the added AR data. These models are heavily disfavored by both information criteria, with AIC and BIC weights of essentially zero. The AIC weight favors the quasi-hyperbolic discounting model with the additional now-vs-later division parameter η (estimated at around 18 hours), while the BIC weight (heavily) favors the generalized hyperbolic model.

A.6 Focusing on Data From Last Two Weeks

Figure A3 and Table A4 replicate those in the text when only using data from the final two weeks. There does appear to be some effect on the raw task decisions, with the changes in task decisions across the week shrunken by about 25%. However, the effect is absent in the estimated discount curve and the parametric discounting-function estimates, which are nearly unchanged from those derived from all of the data. The likely explanation for this discrepancy is a change in the cost curve across the weeks, a trend also seen in AR. The later cost curves are estimated to be steeper (the parameter φ_D is estimated at around 310 for the entire sample and 135 for the last two weeks) but with less curvature (the parameter γ is estimated at around 1.95 for the entire sample and 1.82 for the last two weeks). While a change in the cost curve directly affects raw task decisions, the parametric discounting estimates remain unchanged because they are constructed to account for the change.

A.7 Different Subject Samples

The left panel of Figure A4 presents the smoothed task decisions for four additional subject samples: (1) the main sample minus six subject who left the experiment; (2) the main sample minus six subjects with "little" curvature (individual estimates of $\gamma < 1.25$); (3) the main sample minus five subjects with "extreme" curvature (individual estimates of $\gamma > 4$); and (4) all subjects in the experiment. The use of different samples has virtually no impact on the smoothed curves.

The right panel displays the smoothed discount curve for the first three of these subject samples (as discussed in the main text, the structural estimation does not converge with all subjects included). Similarly, the estimates of the quasi-hyperbolic and hyperbolic specifications for these three samples are shown in Table A5 (given the number of comparison subject samples, the full tables for each sample are not shown). The smoothed discount curves and estimated parameters are largely the same, although there is a slight drop in discounting at later time periods when removing those with little curvature.

A.8 Fixed Effect Choices

The main smoothed task decision graph is constructed by smoothing the residuals of a regression of task decisions on wage, subject and week fixed effects. To understand the impact of these fixed effects, Figure A5 presents the smoothed task decisions using different fixed effects in the residual construction: (1) no fixed effects; (2) subject fixed effects; (3) subject and wage fixed effects. Similarly, the right panel presents the smooth discounting graph with the first two of these specifications (the structural estimation always accounts for the impact of wages through the estimated cost curve). Finally, the estimates of the quasi-hyperbolic and hyperbolic specifications for these two different fixed-effect specifications are shown in Table A6.

Broadly, the results are much noisier – to the point of being qualitatively different from those in the main paper – when subject fixed effects are not included. This is perhaps not surprising since around 27% of the heterogeneity in all task decisions is explained by variation in the mean task decisions of different subjects. This result suggests the benefit of a within-subject design, in which the discount curve is identified from decisions across time within the same subject.

A.9 Exact Timing of Consumption

Recall that subjects completed tasks for each chosen-work date during a self-chosen three-hour time window. The window interval was chosen to balance the desire to pinpoint the exact timing of future work with the need to allow subjects some flexibility in login times.

In the main analysis, the time from the decision to work ($t - k$ in the model) is calculated as the interval between the decision time k and the subject's *actual* login time t during the three-hour window. This method presupposes that the subject correctly predicts the future login time t when making decisions, which is problematic if the subject's perception of the login time is biased. A natural concern is that a naive time-inconsistent subject will consistently predict an earlier login time $\widehat{t}$ than the true time t, because she doesn't appreciate her future desire to delay work. As a result, this naive subject at time k will mistakenly express work preferences using $D(\widehat{t} - k)$ rather than $D(t - k)$.The magnitude of this effect is determined by the difference between $\widehat{t}$ and t, which is bounded by the time-window interval of three hours and therefore likely small. During the initial laboratory session, subjects were asked to identify the *expected* login time for each future date.[31] The average absolute distance between the two times was 44 minutes, and – somewhat reassuringly – the predicted login time was non-statistically-significantly 2.4 minutes earlier than the actual login time ($Z = 0.60, p = 0.28$).

To address the broad endogeneity of the login time within the work window, I use two modified calculations of the distance from a decision to work using the expected login time rather than the actual login time. The difference between the two definitions lies in the treatment of decisions that occur after the expected login time: the first definition treats all of these decisions as occurring when work is immediate, while the second entirely removes these decisions from the data.

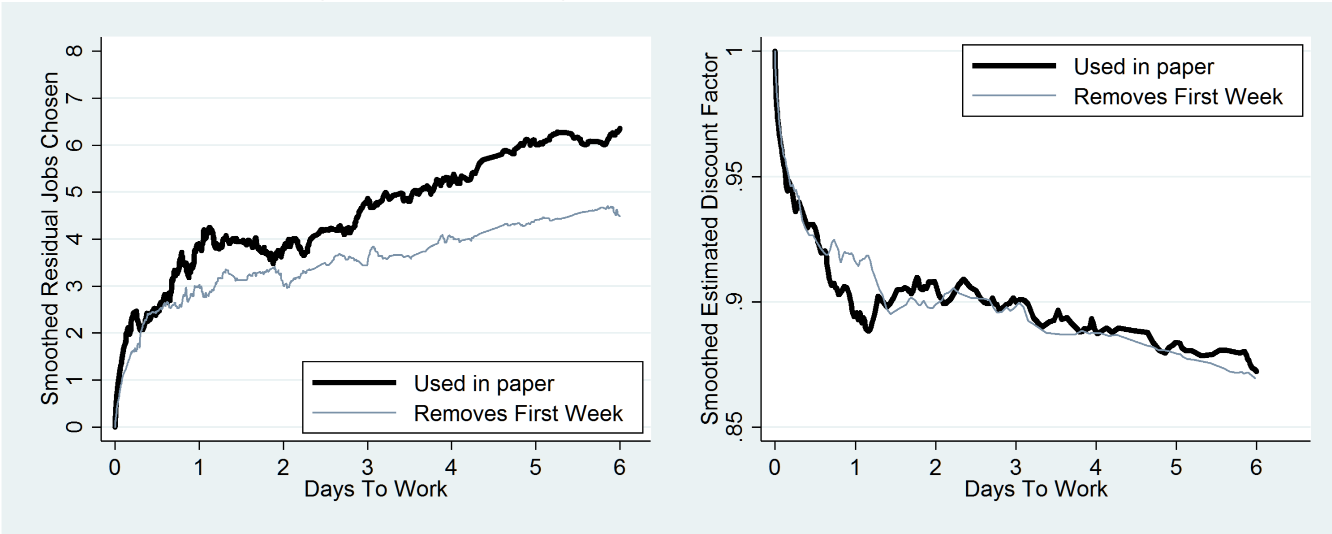
Figure A.9 presents the smoothed discount curves and Table A.9 presents the quasi-hyperbolic and hyperbolic specifications using the two time definitions. In both cases, there is virtually no impact of the use of these different definitions on the results.

Figure A2: Main Figure Including Data from Augenblick & Rabin (2017)

Note: This figure replicates Figure 3 with the inclusion of data from Augenblick & Rabin (2017). The additional portion of the curve derived from the additional data uses a dashed line. In the histogram in the right panel, the frequency bar for decisions made at the time of work is excluded for readability (there are 3,475 observations at this point).

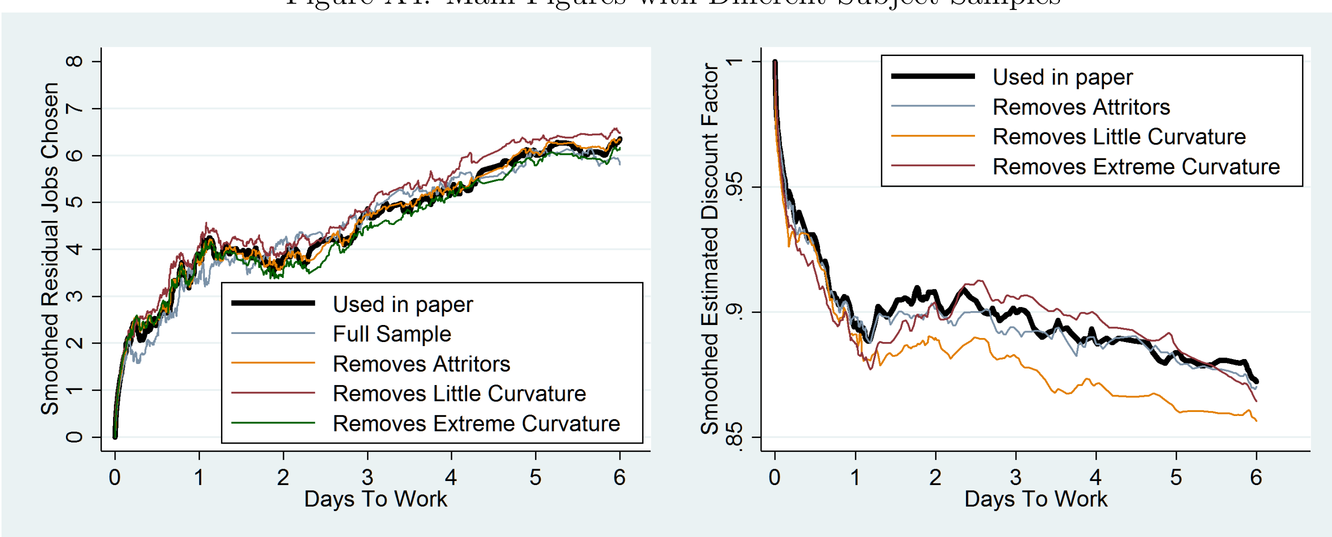
[31] 4 subjects produced 6 expected login times that did not fall within the three-hour window. These login times are recoded as the midpoint of the interval.

Figure A3: Main Figure with Different Data Subsamples



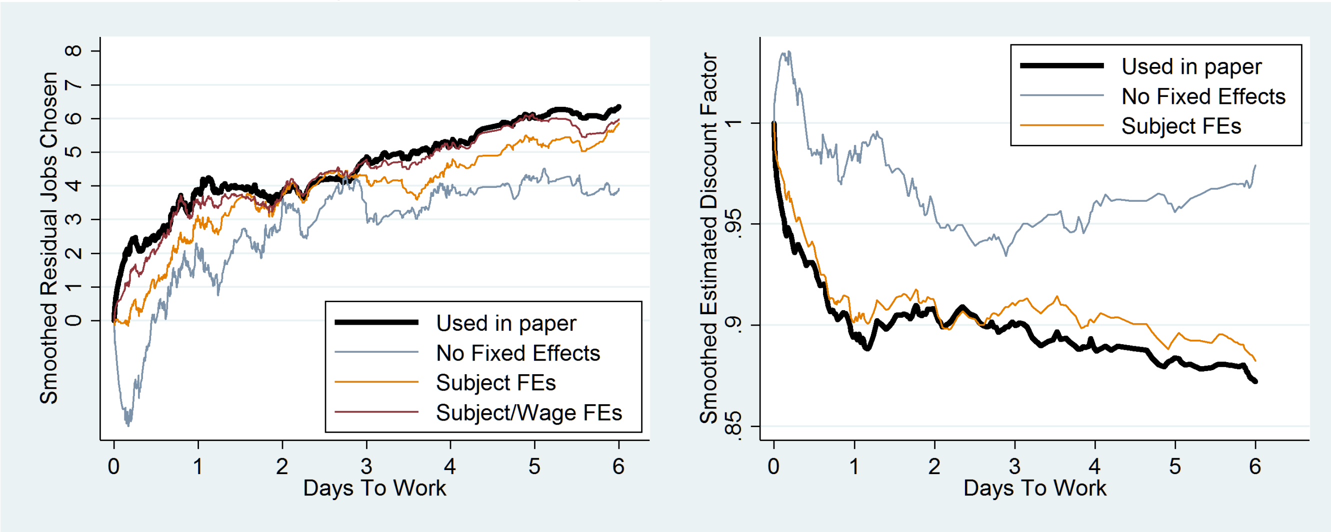
Note: This figure compares the estimated curves from Figure 3 (in dark bold) with those created only using data from the final two work weeks.

Figure A4: Main Figures with Different Subject Samples



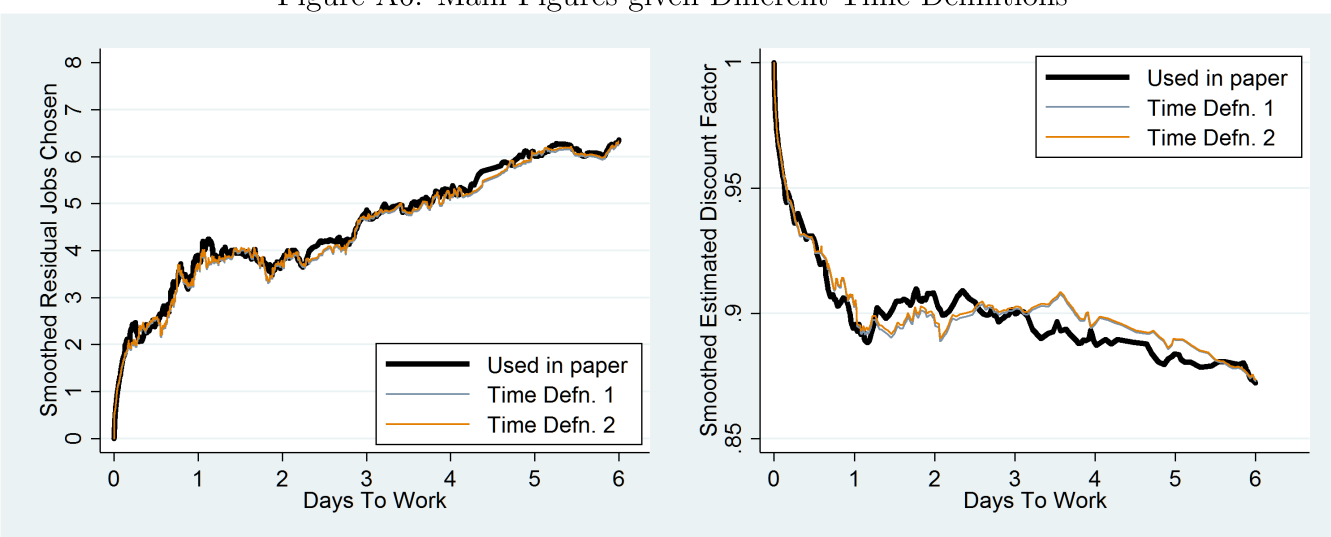
Note: This figure compares the estimated curves from Figure 3 (in dark bold) with those created using different subject samples. The left panel includes a curve in which the full sample of 99 subjects are included (inclusion of these subjects leads the estimation to create the right panel to fail). Both panels present results given the removal of subjects that (1) prematurely left the experiment (2) almost always choose either 0 or 100 tasks, and (3) are estimated to have extreme cost curvature parameters ($\hat{\gamma}_i > 4$ or $\hat{\gamma}_i < 1$).

Figure A5: Main Figures given Different Fixed Effects



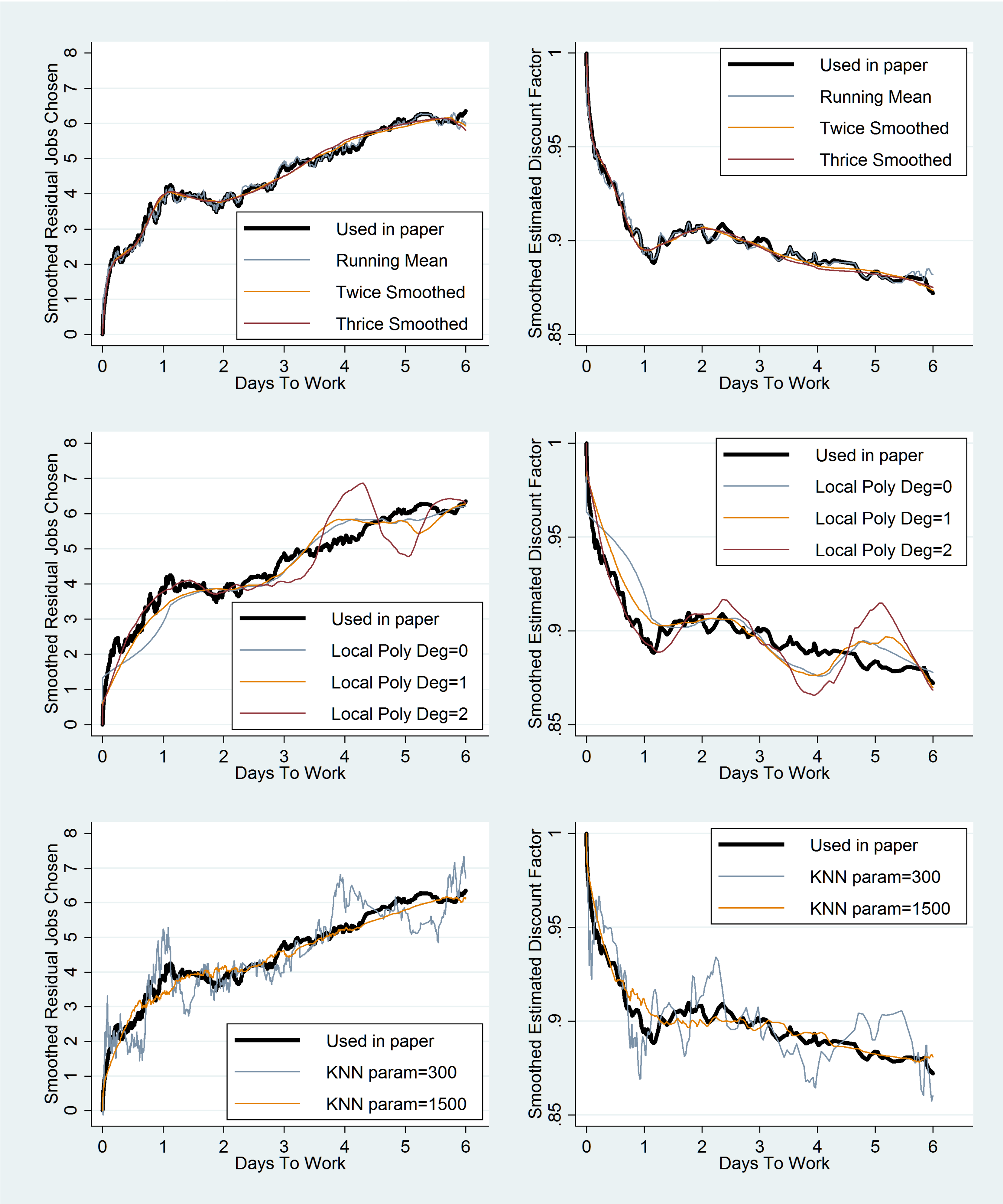
Note: This figure compares the estimated curves from Figure 3 (in dark bold) with those created given the inclusion of different sets of fixed effects. Both panels include specifications with no fixed effects and only using subject fixed effects. The left panel includes a specification with subject and wage fixed effects (the effect of wage is structurally parameterized in the estimation for the right panel).

Figure A6: Main Figures given Different Time Definitions



Note: This figure compares the estimated curves from Figure 3 (in dark bold) with those created given different definitions of work time by using the predicted work time elicited from the subjects at the start of the experiment.

Figure A7: Main Figures with Different Smoothing Options



Note: This figure compares the estimated curves from Figure 3 (in dark bold) with those created given different smoothing techniques and smoothing parameters. In the third row, "KNN" represents the number of nearest neighbors used on each side of the smoothed point.

Table A3: Main Table Including Data from Augenblick & Rabin (2017)

	(1) Delta Only	(2) β-δ δ:=1	(3) β-δ -	(4) β-δ-η -	(5) Hyper-bolic	(6) General Hyperb.
Discount Factor δ	0.991	1.000	0.994	0.995		
	(0.002)	(.)	(0.002)	(0.002)		
Present Bias β		0.905	0.939	0.936		
		(0.023)	(0.027)	(0.016)		
Now (Hours) η		0.0	0.0	17.8		
		(.)	(.)	(.)		
Hyperbolic κ					0.010	61.415
					(0.003)	(291.758)
Gener. Hyper. α						1.16
						(4.72)
Cost Curvature γ	2.09	2.10	2.10	2.10	2.10	2.10
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Cost Slope φ_D	367	357	358	367	367	358
	(133)	(129)	(130)	(137)	(133)	(131)
Observations	13515	13515	13515	13515	13515	13515
Subjects	150	150	150	150	150	150
Log Likelihood	-41367.9	-41366.9	-41355.6	-41350.9	-41366.4	-41352.5
Akaike IC	83051.7	83049.7	83029.2	83021.9	83048.7	83023.0
Bayesian IC	84238.5	84236.5	84223.6	84223.7	84235.6	84217.3
AIC (weight)	0.000000	0.000001	0.015829	0.622490	0.000001	0.361679
BIC (weight)	0.000023	0.000062	0.040509	0.037319	0.000102	0.921986
$H_0(\hat{\delta}=1)$	p<0.001		p= 0.023	p= 0.050		
$H_0(\widehat{\beta}=1)$		p<0.001	p= 0.025	p<0.001		
$H_0(\widehat{\kappa}=0)$					p<0.001	p= 0.833
$H_0(\widehat{\alpha}=1)$						p= 0.974

Note: This table replicates Table 2 with the inclusion of data from Augenblick & Rabin (2017).

Table A4: Main Table with First Week Removed

	(1) Delta Only	(2) β-δ δ:=1	(3) β-δ -	(4) β-δ-η -	(5) Hyper-bolic	(6) General Hyperb.
Discount Factor δ	0.985	1.000	0.989	0.997		
	(0.005)	(.)	(0.005)	(0.006)		
Present Bias β		0.927	0.951	0.930		
		(0.037)	(0.037)	(0.020)		
Now (Hours) η		0.0	0.0	18.8		
		(.)	(.)	(.)		
Hyperbolic κ					0.016	65.502
					(0.006)	(393.352)
Gener. Hyper. α						1.09
						(5.73)
Cost Curvature γ	1.82	1.82	1.82	1.82	1.82	1.82
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Cost Slope φ_D	137	134	134	136	137	135
	(59)	(58)	(58)	(59)	(59)	(59)
Observations	6015	6015	6015	6015	6015	6015
Subjects	77	77	77	77	77	77
Log Likelihood	-17549.3	-17552.6	-17544.3	-17537.9	-17549.2	-17545.2
Akaike IC	35262.7	35269.2	35254.6	35243.8	35262.4	35256.3
Bayesian IC	35812.2	35818.7	35810.9	35806.8	35812.0	35812.6
AIC (weight)	0.000080	0.000003	0.004399	0.993522	0.000090	0.001907
BIC (weight)	0.049495	0.001927	0.095405	0.756305	0.055432	0.041436
$H_0(\hat{\delta}=1)$	p= 0.004		p= 0.021	p= 0.559		
$H_0(\widehat{\beta}=1)$		p= 0.046	p= 0.183	p<0.001		
$H_0(\widehat{\kappa}=0)$					p= 0.007	p= 0.868
$H_0(\widehat{\alpha}=1)$						p= 0.987

Note: This table replicates Table 2 but only includes data from the final two work weeks.

Table A5: Comparison Table: Different Subject Samples

	(1) β-δ Main	(2) β-δ -Attr.	(3) β-δ -$\downarrow$ curv.	(4) β-δ -$\uparrow$ curv.	(5) Hyp. Main	(6) Hyp. -Attr.	(7) Hyp. -$\downarrow$ curv.	(8) Hyp. -$\uparrow$ curv.
Discount Factor δ	0.987	0.986	0.986	0.987				
	(0.004)	(0.004)	(0.005)	(0.004)				
Present Bias β	0.946	0.946	0.930	0.945				
	(0.031)	(0.032)	(0.031)	(0.032)				
Now (Hours) η	0.0	0.0	0.0	0.0				
	(.)	(.)	(.)	(.)				
Hyperbolic κ					0.018	0.019	0.021	0.018
					(0.005)	(0.005)	(0.005)	(0.005)
Cost Curvature γ	1.95	1.94	2.09	1.97	1.95	1.94	2.09	1.97
	(0.11)	(0.11)	(0.10)	(0.10)	(0.11)	(0.11)	(0.10)	(0.10)
Cost Slope φ_D	309.48	283.12	560.85	356.03	318.15	290.80	581.25	366.07
	(160.61)	(153.65)	(286.69)	(183.62)	(164.87)	(157.55)	(296.70)	(188.50)
Observations	8875	8545	8145	8420	8875	8545	8145	8420
Subjects	78	73	72	73	78	73	72	73
Log Likelihood	-26877.2	-25775.3	-25519.9	-25655.1	-26884.4	-25782.2	-25531.0	-25662.7
Akaike's IC	53926.5	51712.7	51199.8	51472.2	53938.9	51724.3	51220.0	51485.4
Schwarz's IC	54536.3	52284.0	51760.2	52042.3	54541.6	52288.6	51773.4	52048.5
$H_0(\hat{\delta}=1)$	p= 0.003	p= 0.001	p= 0.002	p= 0.004				
$H_0(\widehat{\beta}=1)$	p= 0.082	p= 0.089	p= 0.024	p= 0.079				
$H_0(\widehat{\kappa}=0)$					p<0.001	p<0.001	p<0.001	p<0.001

Note: This table compares the results for different subject samples for the quasi-hyperbolic and hyperbolic specifications. Columns (1) and (5) exactly replicate the quasi-hyperbolic and hyperbolic results from Table 2 for comparison. Columns (2) and (6) remove subjects that prematurely left the experiment. Columns (3) and (7) remove subjects that almost always choose either 0 or 100 tasks. Columns (4) and (8) remove subjects that are estimated to have extreme cost curvature parameters.

Table A6: Comparison Table: Fixed Effects

	(1) β-δ Main	(2) β-δ No FE	(3) β-δ Subject FE	(4) Hyp. Main	(5) Hyp. No FE	(6) Hyp. Subject FE
Discount Factor δ	0.987	0.988	0.986			
	(0.004)	(0.004)	(0.004)			
Present Bias β	0.946	0.984	0.959			
	(0.031)	(0.028)	(0.026)			
Now (Hours) η	0.0	0.0	0.0			
	(.)	(.)	(.)			
Hyperbolic κ				0.018	0.014	0.018
				(0.005)	(0.004)	(0.005)
Cost Curvature γ	1.95	1.81	1.93	1.95	1.81	1.93
	(0.11)	(0.08)	(0.11)	(0.11)	(0.08)	(0.11)
Cost Slope φ_D	309.48	138.62	190.47	318.15	140.24	195.59
	(160.61)	(49.22)	(79.66)	(164.87)	(50.11)	(82.31)
Observations	8875	8875	8875	8875	8875	8875
Subjects	78	78	78	78	78	78
Log Likelihood	-26877.2	-30196.1	-27089.1	-26884.4	-30196.3	-27092.9
Akaike's IC	53926.5	60396.2	54338.3	53938.9	60394.6	54343.8
Schwarz's IC	54536.3	60410.4	54905.6	54541.6	60401.7	54904.0
$H_0(\hat{\delta}=1)$	p= 0.003	p= 0.004	p= 0.002			
$H_0(\widehat{\beta}=1)$	p= 0.082	p= 0.554	p= 0.110			
$H_0(\widehat{\kappa}=0)$				p<0.001	p<0.001	p<0.001

Note: This table compares the results given the inclusion of different fixed effects for the quasi-hyperbolic and hyperbolic specifications. Columns (1) and (4) exactly replicate the quasi-hyperbolic and hyperbolic results from Table 2 for comparison. Columns (2) and (5) do not include any fixed effects. Columns (3) and (6) include subject fixed effects but no week-level fixed effects.

Table A7: Comparison Table: Definitions of Work Timing

	(1)	(2)	(3)	(4)	(5)	(6)
	β-δ	β-δ	β-δ	Hyp.	Hyp.	Hyp.
	Main	Expect	Expect	Main	Expect	Expect
	-	Time1	Time2	-	Time1	Time2
Discount Factor δ	0.987	0.987	0.987			
	(0.004)	(0.004)	(0.004)			
Present Bias β	0.946	0.946	0.946			
	(0.031)	(0.031)	(0.031)			
Now (Hours) η	0.0	0.0	0.0			
	(.)	(.)	(.)			
Hyperbolic κ				0.018	0.018	0.018
				(0.005)	(0.005)	(0.005)
Cost Curvature γ	1.95	1.95	1.95	1.95	1.95	1.95
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Cost Slope φ_D	309.48	309.48	309.48	318.15	318.15	318.15
	(160.61)	(160.61)	(160.61)	(164.87)	(164.87)	(164.87)
Observations	8875	8875	8875	8875	8875	8875
Subjects	78	78	78	78	78	78
Log Likelihood	-26877.2	-26877.2	-26877.2	-26884.4	-26884.4	-26884.4
Akaike's IC	53926.5	53926.5	53926.5	53938.9	53938.9	53938.9
Schwarz's IC	54536.3	54536.3	54536.3	54541.6	54541.6	54541.6
$H_0(\hat{\delta}=1)$	p= 0.003	p= 0.003	p= 0.003			
$H_0(\widehat{\beta}=1)$	p= 0.082	p= 0.082	p= 0.082			
$H_0(\widehat{\kappa}=0)$				p<0.001	p<0.001	p<0.001

Note: This table compares the results given two different definitions of work time for the quasi-hyperbolic and hyperbolic specifications. Columns (1) and (4) exactly replicate the quasi-hyperbolic and hyperbolic results from Table 2 for comparison. Columns (2) and (5) use the first alternative definition of timing discussed in the text. Columns (3) and (6) use the second definition of timing.

A.10 Different Smoothing Techniques

Figure A7 displays the effect of using different smoothing techniques and parameters on the smoothed task decision and discount curve graphs.

The first row of the figure demonstrates the effect of using a (1) running mean smoother, (2) running line smoother, run twice, and (3) running line smoother, run three times. There is nearly no difference between the use of a running line or running mean smoother. The effect of multiple smoothing runs is, predictably, a smoother curve.

The second row demonstrates the effect of using a kernel-weighted local polynomial smoother of three different degrees. The results are broadly locally smoother than the results in the main paper, but globally much more erratic. Notably, this pattern was also apparent in pre-experiment data simulations, where a known discount function was used to generate the data. In many cases, these smoothers appeared to "oversmooth" sharp changes in the discount function, but "undersmooth" longer stable areas.

Finally, the third row shows the effect of using different "KNN" parameters, which is the number of nearest neighbours used on each side of the smoothed point. Not surprisingly, as this number falls, the smoothed curve becomes more erratic. In the main paper, I use a relatively low parameter to capture the steep movement close to the work time.

A.11 Monetary Utility Function Curvature

In the main text, the utility function is assumed to be quasi-linear in monetary payments. In this section, the robustness of this parameterization is checked by adding curvature to the monetary utility function. Specifically, adding initial wealth y to the subject's monetary payment, suppose that $U(y+e\cdot w)$ is equal to $-\exp(-a\cdot(y+w\cdot e))$, the standard CARA utility function (with a used instead of the traditional α to differentiate from the use of α in the generalized hyperbolic discount function). Given this parameterization, equation (2) can be rewritten as:

$$e^* = \arg\max_e -D_m(T)\cdot exp(-a\cdot(y+w\cdot e)) - D(t-k)\cdot\frac{1}{\varphi\cdot\gamma}(e+10)^\gamma. \tag{6}$$

Taking the first-order condition of equation (6) with respect to e and solving for e yields the predicted choice e^*:

$$e^* = \frac{(\gamma-1)}{a\cdot w}\cdot\mathbf{W}\left(\frac{a\cdot w\cdot\left(\frac{\varphi_{Dy}\cdot w}{D(t-k)}\cdot\frac{exp(10\cdot a\cdot w)}{D(t-k)}\right)^{\frac{1}{\gamma-1}}}{\gamma-1}\right)-10, \tag{7}$$

where (1) $\mathbf{W}(z)$ represents the principle value of the Lambert W function (the principal solution for $\mathbf{w}$ in the implicit equation $z=\mathbf{w}\cdot exp(\mathbf{w})$), and (2)$\varphi_{Dy}\equiv D_m(T)\cdot\varphi\cdot exp(-a\cdot y)$ given that $D_m(T)$ is constant throughout the experiment and assuming that $exp(-a\cdot y)$ is constant as well.[32] Appropriately adjusting the likelihood function in equation (6), it is possible to perform a similar analysis to that in the text.

[32] In fact, this is strictly not true as the subjects' wealth can be seen as rising as the experiment progresses and the subject earns wages. However, as this change in wealth is both small and (relatively) predictable, wealth is assumed to be constant throughout the experiment.

Figure A.13 reproduces the right panel of Figure 3 given a monetary curvature parameter a equal to .001, .005, .01, and .02. These parameters imply that the ratio of the marginal utility of an additional dollar at the start of the experiment in comparison to the end of the experiment (after receiving ~\$100) is 1.1, 1.6, 2.7, and 7.4, respectively. Even the extreme levels of curvature have almost no impact on the estimated discounted curve.

Table A.13 contains the estimates of the quasi-hyperbolic and hyperbolic specifications for the different risk parameters. The discount factor parameters are very stable. The effect is largely absorbed by the cost parameters: for example, $\hat{\gamma}$ changes from 1.95, 1.81, 1.70, and 1.51 in Columns (1)-(4), respectively. Given that either curvature in the monetary utility function or the cost function can explain the observed shape of the effort-wage relationship, forcing a larger amount of monetary curvature leads to a lower estimate of effort curvature to fit the relationship, but does not broadly change the estimates of how the relationship changes over time.

A.12 A More Flexible Parameterization of Cost Curve

In the main paper, the cost function component is parameterized as a two-parameter power function $\frac{1}{\varphi\gamma}(e+10)^{\gamma}$. In this section, the cost function is instead parameterized as a second- or third-degree polynomial, leading equation (2) to be rewritten as either:

$$e^*_{2^{nd}} = \arg\max_{e} \; e \cdot w - D(t-k) \cdot (\psi_1(e+10) + \frac{1}{2}\psi_2(e+10)^2) \text{ or} \tag{8}$$

$$e^*_{3^{rd}} = \arg\max_{e} \; e \cdot w - D(t-k) \cdot (\psi_1(e+10) + \frac{1}{2}\psi_2(e+10)^2 + \frac{1}{3}\psi_3(e+10)^3). \tag{9}$$

Taking the first-order condition of these equations with respect to e and solving for e yields the predicted choice e^* of either:

$$e^*_{2^{nd}} = \frac{w - D(t-k)\psi_2}{D(t-k)\psi_2} - 10 \text{ or} \tag{10}$$

$$e^*_{3^{rd}} = \frac{\sqrt{(D(t-k)^2(\psi_2^2 - 4\psi_1\psi_3) + D(t-k)(4w\psi_3)} - D(t-k)\psi_2}{2D(t-k)\psi_3} - 10. \tag{11}$$

The smoothed discount curves for these parameterizations of the cost curve are shown in Figure A.13, with the results from the quasi-hyperbolic and hyperbolic estimations in Table A9.[33] There is very little effect of this alternative parametric specification, suggesting that the misspecification of the cost function is unlikely to be driving the results.

A.13 Different Error Specifications

Throughout the paper, observed effort choices e are assumed to be composed of predicted effort choice and an additive reduced-form normally-distributed error term ε with a Tobit correction. This section demonstrates the effect of using different assumptions about the location and form of the error term.

[33] The generalized hyperbolic specification does not converge when the cost curve is parameterized as a third-degree polynomial.

The first specification is the direct estimation of 5 without the Tobit correction:

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

The second specification uses a normally-distributed error term on the *log* of effort decisions:

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

The third specification places the normally-distributed error on the cost parameter γ. In this case, observed effort would be equal to:

$$e = (\frac{\varphi \cdot w}{D(t-k)})^{\frac{1}{\gamma + \varepsilon_\gamma - 1}} - 10,$$

with ε_γ distributed normally with mean zero and standard deviation σ_γ. Given this,

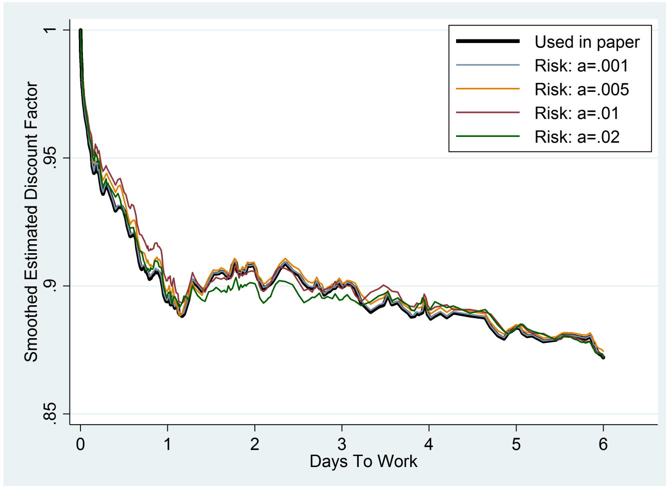
$$\gamma^* \equiv 1 + \frac{\ln(\frac{\varphi \cdot w}{D(t-k)})}{\ln(e+10)} \sim N(\gamma, \sigma_\gamma^2),$$

which leads the likelihood of observation i to become:

$$L^{LnError}(e_j) = \mathbf{1}(e_j < 100)\phi\left(\frac{ln(\gamma^*) - ln(\gamma_j)}{\sigma}\right) + \mathbf{1}(e_j = 100)\Phi\left(\frac{ln(\gamma_j^*) - ln(\gamma)}{\sigma}\right).$$

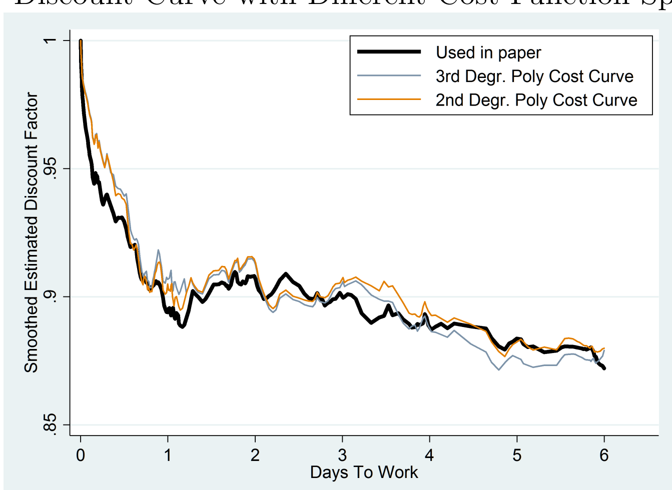
The smoothed discount curves for these parameterizations of the error are shown in Figure A.13, with the results from the quasi-hyperbolic and hyperbolic estimations in Table A10. Although the shape is similar across specifications, the discount curve appears shrunken by around 30% when the Tobit correction is not applied, and expanded by around 50% when the error is assumed to be log-normal.

Figure A8: Discount Curve with Different Levels of Utility Curvature



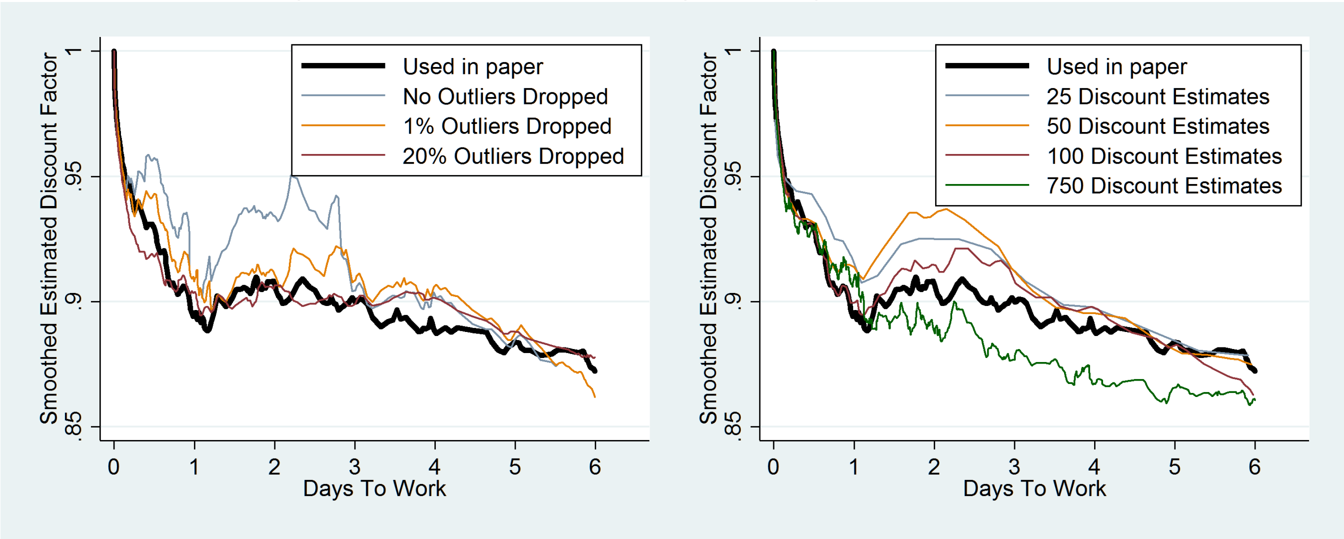
Note: This figure compares the estimated discounting curves from the right panel of Figure 3 (in dark bold) with those created when assuming that the monetary utility function takes a CARA form with varying levels of curvature.

Figure A9: Discount Curve with Different Cost Function Specifications



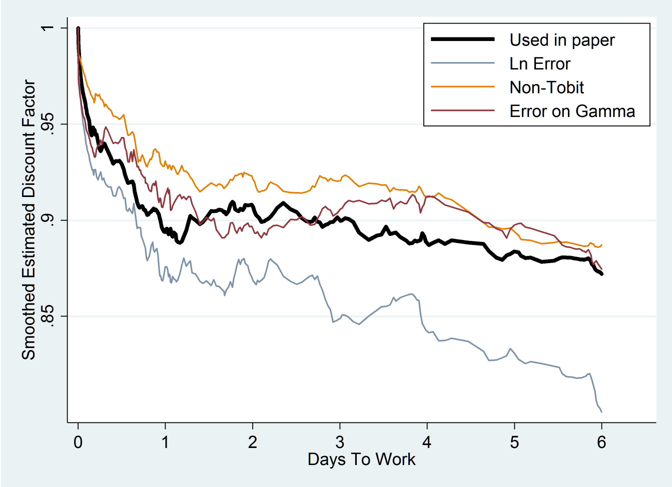
Note: This figure compares the estimated discounting curves from the right panel of Figure 3 (in dark bold) with those created when parameterizing the cost curve as a second- or third- degree polynomial.

Figure A11: Main Discount Figure Using Different Methods



Note: This figure compares the estimated discounting curves from the right panel of Figure 3 (in dark bold) with those created when using different cutoffs for outlier removal (left panel) or with a different number of points at which the discount factor is estimated (right panel).

Figure A10: Discount Curve with Different Error Specifications



Note: This figure compares the estimated discounting curves from the right panel of Figure 3 (in dark bold) with those created when parameterizing the error term such that the error is log-normal, there is no Tobit correction, and with the location of the error moved to the cost curvature term.

Table A8: Comparison Table: Monetary Curvature

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	β-δ	β-δ	β-δ	β-δ	Hyp.	Hyp.	Hyp.	Hyp.
	a=0	a=0.005	a=0.01	a=0.02	a=0	a=0.005	a=0.01	a=0.02
Discount Factor δ	0.987	0.986	0.986	0.986				
	(0.004)	(0.004)	(0.003)	(0.003)				
Present Bias β	0.946	0.950	0.953	0.948				
	(0.031)	(0.026)	(0.022)	(0.018)				
Now (Hours) η	0.0	0.0	0.0	0.0				
	(.)	(.)	(.)	(.)				
Hyperbolic κ					0.018	0.019	0.019	0.020
					(0.005)	(0.004)	(0.004)	(0.003)
Cost Curvature γ	1.95	1.81	1.70	1.51	1.95	1.81	1.69	1.51
	(0.11)	(0.10)	(0.10)	(0.09)	(0.11)	(0.10)	(0.10)	(0.09)
Cost Slope φ_D	309.48	179.34	116.58	57.61	318.15	183.99	119.38	59.02
	(160.61)	(90.59)	(55.58)	(23.58)	(164.87)	(93.24)	(56.92)	(24.31)
Observations	8875	8875	8875	8875	8875	8875	8875	8875
Subjects	78	78	78	78	78	78	78	78
Log Likelihood	-26877.2	-26789.7	-26734.9	-26727.3	-26884.4	-26796.4	-26741.4	-26736.5
Akaike's IC	53926.5	53751.4	53641.8	53626.6	53938.9	53762.8	53652.7	53643.1
Schwarz's IC	54536.3	54361.2	54251.6	54236.4	54541.6	54365.5	54255.4	54245.8
$H_0(\hat{\delta}=1)$	p= 0.003	p<0.001	p<0.001	p<0.001				
$H_0(\widehat{\beta}=1)$	p= 0.082	p= 0.054	p= 0.033	p= 0.003				
$H_0(\widehat{\kappa}=0)$					p<0.001	p<0.001	p<0.001	p<0.001

Note: This table compares the results when assuming different levels of monetary utility curvature for the quasi-hyperbolic and hyperbolic specifications. Columns (1) and (5) set the CARA curvature parameter at 0, and therefore exactly replicate the quasi-hyperbolic and hyperbolic results from Table 2 for comparison. Columns (2) and (6) set a=0.005, Columns (3) and (7) set a=0.01, and Columns (4) and (8) set a=0.02.

Table A9: Comparison Table: Different Cost Parameterization

	(1) β-δ Main	(2) β-δ 2D Poly	(3) β-δ 3D Poly	(4) Hyp. Main	(5) Hyp. 2D Poly	(6) Hyp. 3D Poly
Discount Factor δ	0.987	0.985	0.985			
	(0.004)	(0.004)	(0.004)			
Present Bias β	0.946	0.954	0.946			
	(0.031)	(0.023)	(0.022)			
Now (Hours) η	0.0	0.0	0.0			
	(.)	(.)	(.)			
Hyperbolic κ				0.018	0.020	0.021
				(0.005)	(0.004)	(0.004)
Cost Curvature γ	1.95			1.95		
	(0.11)			(0.11)		
Cost Slope φ_D	309			318		
	(161)			(165)		
Poly Param 1 ψ_1		0.0224	0.0248		0.0217	0.0239
		(0.0103)	(0.0088)		(0.0100)	(0.0085)
Poly Param 2 ψ_2		0.0039	0.0037		0.0038	0.0036
		(0.0002)	(0.0002)		(0.0002)	(0.0001)
Poly Param 3 ψ_3			0.000002			0.000002
			(0.000000)			(0.000000)
Observations	8875	8875	8875	8875	8875	8875
Subjects	78	78	78	78	78	78
Log Likelihood	-26877.2	-26964.1	-26902.7	-26884.4	-26968.7	-26908.8
Akaike's IC	53926.5	54088.2	53967.5	53938.9	54095.4	53977.5
Schwarz's IC	54536.3	54655.5	54541.8	54541.6	54655.6	54544.8
$H_0(\hat{\delta}=1)$	p= 0.003	p<0.001	p<0.001			
$H_0(\widehat{\beta}=1)$	p= 0.082	p= 0.042	p= 0.014			
$H_0(\widehat{\kappa}=0)$				p<0.001	p<0.001	p<0.001

Note: This table compares the results when using different parameterizations of the cost curve for the quasi-hyperbolic and hyperbolic specifications. Columns (1) and (5) replicate the quasi-hyperbolic and hyperbolic results from Table 2 for comparison. Columns (2) and (5) parameterize the cost curve as a second-degree polynomial while Columns (3) and (6) parameterize the cost curve as a third-degree polynomial.

Table A10: Comparison Table: Error Forms

	(1) β-δ Main	(2) β-δ NoTobit	(3) β-δ LnError	(4) β-δ Error γ	(5) Hyp. Main	(6) Hyp. NoTobit	(7) Hyp. LnError	(8) Hyp. Error γ
Discount Factor δ	0.987	0.987	0.972	0.987				
	(0.004)	(0.003)	(0.005)	(0.005)				
Present Bias β	0.946	0.949	0.924	0.926				
	(0.031)	(0.018)	(0.030)	(0.027)				
Now (Hours) η	0.0	0.0	0.0	0.0				
	(.)	(.)	(.)	(.)				
Hyperbolic κ					0.018	0.018	0.039	0.020
					(0.005)	(0.004)	(0.007)	(0.005)
Cost Curvature γ	1.95	2.46	2.04	1.15	1.95	2.46	2.04	1.15
	(0.11)	(0.13)	(0.09)	(0.09)	(0.11)	(0.13)	(0.09)	(0.09)
Cost Slope φ_D	309.48	2231.40	297.39	13.31	318.15	2305.59	313.28	13.95
	(160.61)	(1252.02)	(99.57)	(4.38)	(164.87)	(1297.77)	(105.59)	(4.61)
Observations	8875	8875	8875	8875	8875	8875	8875	8875
Subjects	78	78	78	78	78	78	78	78
Log Likelihood	-26877.2	-38950.8	-8554.6	-334.0	-26884.4	-38957.0	-8558.5	-338.3
Akaike's IC	53926.5	78073.7	17281.2	839.9	53938.9	78084.1	17287.0	846.5
Schwarz's IC	54536.3	78683.5	17891.0	1449.7	54541.6	78686.8	17889.8	1449.3
$H_0(\hat{\delta}=1)$	p= 0.003	p<0.001	p<0.001	p= 0.006				
$H_0(\widehat{\beta}=1)$	p= 0.082	p= 0.005	p= 0.011	p= 0.007				
$H_0(\widehat{\kappa}=0)$					p<0.001	p<0.001	p<0.001	p<0.001

Note: This table compares the results when using different parameterizations of the error term for the quasi-hyperbolic and hyperbolic specifications. Columns (1) and (5) replicate the quasi-hyperbolic and hyperbolic results from Table 2 for comparison. Columns (1) and (5) assume that the error is log-normal, Columns (2) and (6) remove the Tobit correction, and Columns (4) and (8) place the location of the error on the cost curvature term.

A.14 Alternative Discount Function Creation Methods

To construct the smoothed discount curve in the main text, the structural model is estimated with 250 discount factors across the week. Then 5% of these outliers are dropped and smoothed across the week. Figure A11 separately displays the effect changing the outlier-exclusion criteria (to 0%,1% and 20%) and the number of estimated discount factors used for smoothing (to 25, 50, 100, and 750).

As mentioned in the main text, the exclusion of outliers has a meaningful effect on the shape of the curve. By estimating 250 discount factors, each non-immediate discount factor is estimated using only around 15 observations. It is therefore not surprising that some of these estimates are extremely high or low. When these outliers are included, the non-monotonic "hump" around 1-2 days from work is exaggerated. However, removing the top and bottom 0.5% of discount factor estimates largely dampens this effect, suggesting that it is driven by a small number of extreme estimations. Reassuringly, there is a small effect when an additional 4% of outliers are removed and very little effect when a further 15% are removed. The estimation of a smaller number of discount curves also inflates the non-monotonic hump. When 750 discount factors are estimated (nearly one for every non-immediate decision set), the curve is more locally erratic but globally more monotonically decreasing. The main paper uses 250 estimated discount factors to attempt to balance these effects.

A.15 Demographic Correlates

In the experiment, subjects were asked a variety of survey questions. As noted in the PD, "I do not have strong ex-ante hypotheses concerning the interaction of demographic variables (such as gender, intelligence, or major) with the results. However, I have collected these data mainly for use by future researchers who might have an interest in this data. Therefore, I will not use demographic information in the main analysis. I will present all heterogeneity results in the Appendix regardless of the outcome and make a small reference to the results in the paper."

Specifically, during the initial laboratory session, subjects were asked four survey questions. The exact questions were: (1) What is your gender? (2) What year are you at Berkeley? (3) What was your math SAT score? (4) On a scale of 1-10, how much do you procrastinate? (10=more procrastination).

At the end of the experiment, subjects were asked two questions taken from the Global Preference Survey (Falk *et al* (2017), three questions from the cognitive reflection test (Frederick (2005)), and two generic monetary time-discounting questions (Read, Frederick & Scholten (2013)). The exact questions were: (1) On a scale of 0 to 10, how willing or unwilling are you are to take risks? (where 0 means "completely unwilling to take risks" and a 10 means you are "very willing to take risks") (2) On a scale of 0 to 10, how willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? (where 0 means "completely unwilling to give something up" and a 10 means you are "very willing to give something up".), (3) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?, (4) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?, (5) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?, (6) Imagine that you were going to receive $20 in one month with certainty. How

Table A11: Demographic Summary Statistics

	(1) FullMean	(2) FullSD	(3) MainMean	(4) MainSD	(5) DropMean	(6) DropSD
Gender (1=male)	0.28	0.45	0.31	0.47	0.38	0.52
School year	2.65	1.22	2.56	1.26	2.38	1.06
Math SAT	723.11	72.94	721.53	76.76	734.29	80.59
Procrastination	6.55	1.82	6.64	1.84	7.25	1.58
Risk Willingness	5.93	2.00	5.96	2.04	.	.
-Delay Willingness	-7.10	2.04	-7.16	2.05	.	.
CRT (correct of 3)	1.93	1.16	2.05	1.09	.	.
Pay for Today	0.48	0.97	0.56	1.06	.	.
Pay for Month	0.96	1.66	0.97	1.68	.	.
Pay Today>Month	0.11	0.31	0.14	0.35	.	.
Pay Difference	-0.48	1.47	-0.40	1.41	.	.

Notes: This table presents the means and standard deviations of the various survey answers for the full sample, the main sample of 78 subjects, and the sample of 9 subjects who exited the experiment. As this last group did not answer the end survey, there is no data for the final questions. One subjects who reported a gender of "other" is not included for the gender variable. Only 72 and 90 of the subjects in the full sample and main sample reported their math SAT.

much money would you pay in order to receive the money today instead?, (7) Imagine that you were going to receive $20 in two months with certainty. How much money would you pay in order to receive the money in one month instead?

Table A11 contains the mean and standard deviation of the answers, and Table A12 contains the correlation of the answers with the non-parametric and parametric measures of time discounting.

A.16 Log Likelihood Curves for Time Discontinuity Location in Quasi-Hyperbolic Model

The paper provides a variety of estimates for the parameter η: the location of the discontinuity between "now" and "later" in the quasi-hyperbolic discounting model. In Column (4) of the main table (Table 2), this parameter is estimated at around one hour. However, the individual median estimate, the aggregate estimate including data from AR, and the aggregate estimate when dropping the first week (Tables A.3, A3 and A4, respectively) are around 23, 19, and 18 hours, respectively.

To understand the difference in the estimates, Figure A12 plots the log likelihood from the maximum likelihood estimates of the parameter η given four specifications: the main estimation, the main estimation forcing the exponential discounting parameter δ to be 1, the estimation including data from AR, and the estimation when dropping the first week. In each case, the peak of likelihood clearly occurs within one day, but there are two local maxima at around 1 hour and around 20 hours. In the main estimate, the first point has higher likelihood, while in the other estimations, the second is higher. Therefore, while the best estimate of η is consistently within one day of work, the precise

Table A12: Demographic Coorelations

	(1) e[t>1]-e[t=0]	(2) α_{effort}	(3) $-\delta$	(4) $-\beta$	(5) κ
Gender (1=male)	0.20*	0.18	0.07	-0.05	0.05
School year	-0.16	-0.10	-0.24**	-0.29**	-0.21*
Math SAT	-0.14	-0.21*	-0.20*	-0.13	-0.24**
Procrastination	-0.06	-0.08	-0.13	-0.08	-0.07
Risk Willingness	0.17	0.03	0.10	0.04	0.15
-Delay Willingness	-0.15	-0.11	-0.09	0.02	-0.12
CRT (correct of 3)	0.29**	0.29**	0.14	0.12	0.14
Pay for Today	0.03	0.07	0.23**	0.03	0.16
Pay for Month	0.02	-0.02	0.15	0.02	0.09
Pay Today>Month	-0.04	-0.02	-0.01	-0.06	-0.02
Pay Difference	-0.00	0.08	-0.01	0.00	0.02

Notes: This table presents the correlation and significance of five measures of discounting (columns) with multiple survey answers (rows). The signs of measures (3) and (4) are changed such that a higher number for all measure implies more severe discounting. The stars represent statistical significance (* for $p<0.10$, ** for $p<0.05$, and *** for $p<0.01$), with no correction for multiple hypothesis testing.

estimate within in this interval is variable.

A.17 Alternative Wage Figures

The PD states: "Based on simulations and data from other papers, I will aggregate the data into 10 wages bins for visual ease. I will provide a similar graph which shows decisions for each wage in the Appendix." However, after looking at the data, I decided that the non-aggregated graphs were clear enough to include them in the main text. For completeness, Figure A13 presents a variety of graphs of the tasks chosen given wages, which are – not surprisingly – less noisy than the disaggregated graphs presented in the paper.

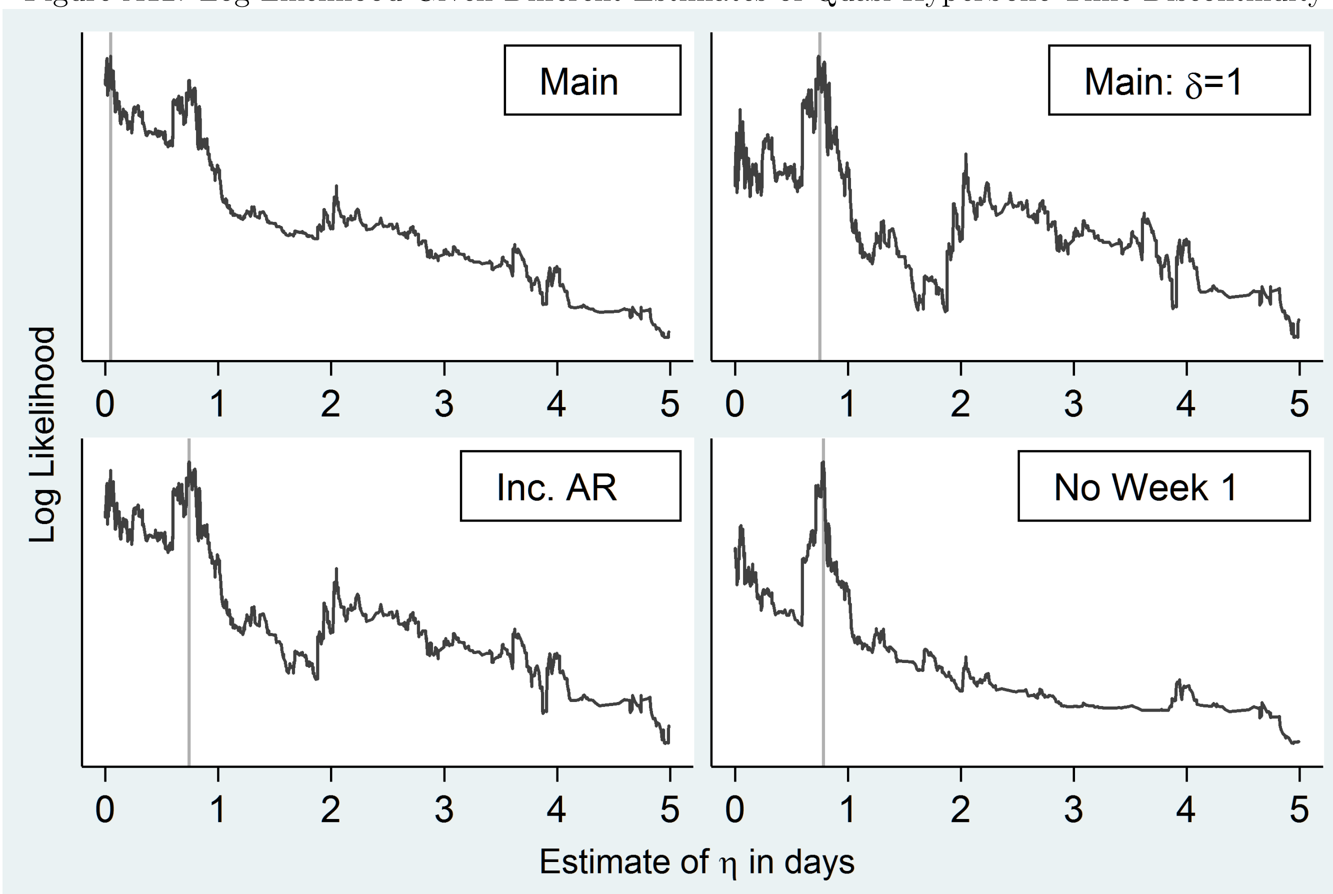
A.18 Preregistration Document

Please see https://www.socialscienceregistry.org/trials/1061/ for the full documented history of the preregistration documents. The document below was posted on February 16, 2016 (before the beginning of the experiment). One sentence was added at the end of the document on February 19, 2016 (after recruitment but prior to any data collection).

Study Outline:

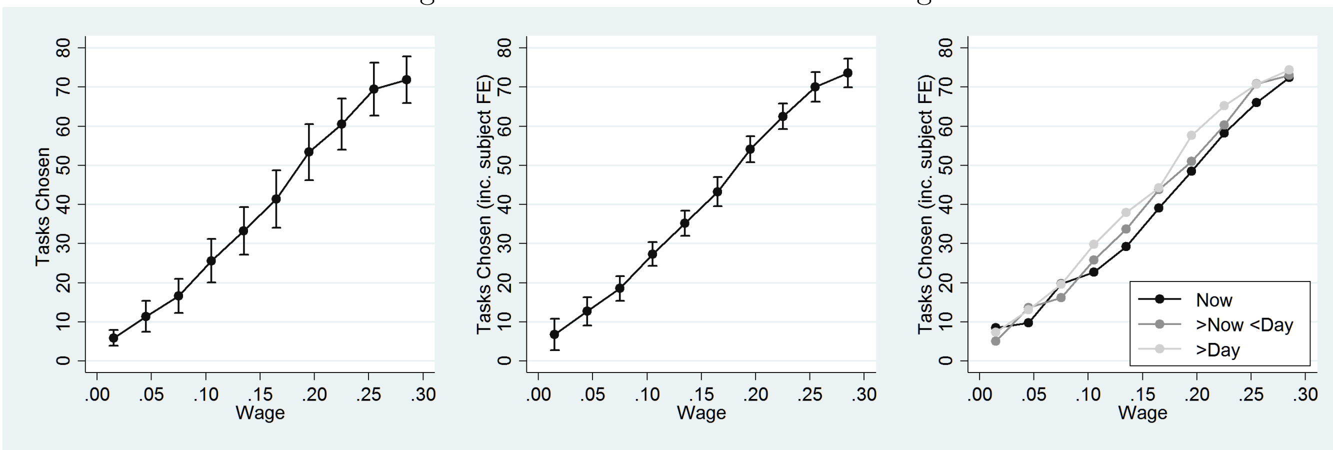
The goal of this study is to estimate the shape of short-term time preferences using incentive-compatible decisions over a real-effort task. The basic idea of the study is fairly simple: after practicing the task in the first week, study subjects are asked the number of tasks they wish to complete for different wages on a given date as that date approaches over the next three weeks. For

Figure A12: Log Likelihood Given Different Estimates of Quasi-Hyperbolic Time Discontinuity



Note: These graphs present the log likelihood associated with different estimates of the parameter η for four situations: the main estimation in the paper, the same estimation where δ is forced to be 1, the estimation including data from AR, and the estimation when dropping the first week. In each case, there are two local maxima at around 1 hour and 20 hours from work, with the global maximum (highlighted by a vertical line) depending on the specification.

Figure A13: Task Decisions versus Wages



Note: This graph is a reproduction of Figure 2 in the paper, but using the aggregation of 10 wage bins rather than individual wages.

example, a subject might be asked questions (via text message) on Monday morning, Wednesday afternoon, and Thursday afternoon about work preferences for that Friday afternoon. Broadly, the study design and analysis plan largely follows that of Augenblick and Rabin (2016), using the same task and potential wages. The main differences are that the choices are all within one week in order to determine a more precise shape for short-term time preferences and there is no focus on sophistication or projection bias.

Outcome Data:

The main data of the study is fairly simple - each decision represents the number of tasks chosen by a specific subject at a given time for a given wage for work that occurs for a given time window.

Sample:

As this is an experiment that takes place over multiple weeks, significant attrition is expected (my expectation is 15-20%). Some subjects will likely quit prior to making a decision. Consequently, in an attempt to gather 100 subjects to reach the point of making decisions, 110 subjects will be initially recruited into the study. In addition to this pre-decision attrition, I expect that other subjects will quit after making some decisions. My expectation is not to include any analysis of the drop-out decision in the main text as I don't believe there will be enough power to make meaningful statements.

Non-Parametric Analysis:

[All of the analysis will cluster the standard errors at the subject-level]

The non parametric analysis will summarize the raw data in a variety of ways (there are some important degrees of freedom in the analysis, which I discuss below):

- Show the aggregate level of tasks chosen for different wages.
- Raw comparisons and difference tests of aggregate wage choices as the work time approaches and/or smoothed comparison of wage choices as the work approaches.
- Raw comparisons and difference tests that loosely test the predictions of different models on the aggregate data (exponential, quasi-hyperbolic, hyperbolic, fixed-cost, etc.).

- Discussion of heterogeneity in individual responses.

Parametric Analysis:

[All of the analysis will cluster the standard errors at the subject-level]:

The functional form of the non-parametric analysis will largely follow Augenblick-Rabin 2016 (there are some important degrees of freedom in the analysis, which I discuss below):

- The aggregate analysis will structurally estimate parameters from the main time discounting models: exponential, hyperbolic, quasi-hyperbolic, and (secondarily) fixed-cost (Benhabib, Bisin, and Schotter 2010) and potentially others.

- In the structural estimation, the disutility of work will be modeled as a power function (with exponent parameter gamma and scaling parameter 1/psi) with the utility of wages modeled as a linear function of money earned.

- For some models (such as fixed-cost and quasi-hyperbolic), it is possible to estimate multiple models with different parameters. For example, a one-parameter quasi-hyperbolic model specifies that the discount function is 1 for t=0 and β for all t>0. One two-parameter model would allow the discount function to be to allow $\beta\delta^t$ for all t>0. However, both of these models do not precisely specify the time-period length and therefore it is likely necessary to define an additional parameter which determines when the discounting discontinuity occurs.

- The fit of the models will be compared. Here, there is an issue which I am unfortunately not sure of the appropriate solution. Usually, measures of goodness-of-fit penalize additional parameters. However, in this case, although the fully specified quasi-hyperbolic model includes three parameters and the hyperbolic model includes one parameter, my intuition is that the quasi-hyperbolic model is still effectively as restrictive as the hyperbolic model. I require more work to determine the effective way to compare the fit of these models.

- One issue is that averaging over quasi-hyperbolic discounters might create patterns that appear hyperbolic in the aggregate. It will should be possible to estimate individual parameters and potentially categorize individual choices to deal with this issue.

Main degrees of freedom:

There are a few degrees of freedom in the analysis. Broadly, my goal is to present the most natural or representative analysis in the main paper and then include other choices in the Appendix.

- Bins/Smoothing over the week: The goal of the non-parametric analysis is to show how decisions change over the week. This presentation requires either a choice of bins or a smoothing method (and smoothing parameter). Unfortunately, based on simulated data, I do not believe it is appropriate to pre-specify the exact bins or smoothing parameters. That is, in the simulated data, different assumed discount functions and parameters create different patterns that only appear visually given different smoothing parameters. My plan is to use multiple smoothing techniques/bins and present the ones that I feel best represent the data in the paper. I will then include a representative set of other possibilities in the Appendix.

- Bins over wages: Either in the paper or the Appendix, I expect to show a graph that demonstrates raw changes in work decisions given wage changes. Based on simulations and data from other papers, I will aggregate the data into 10 wages bins for visual ease. I will provide a similar graph which shows decisions for each wage in the Appendix.

- Participant Sample: It will likely be difficult to properly identify the structural parameters for some subjects, either due to lack of decisions due to attrition, lack of variation in decisions, or decisions that consistently violate monotonicity. The analysis will be run on the entire sample as well as the subsample for which it is possible to identify parameters. I expect to focus on the subsample in the paper and only include the results for the entire sample in the Appendix.

- Decision Sample: There are two ways in which I foresee restricting or breaking up the sample of decisions. First, it might be argued that subjects will be potentially confused in the first week of decisions - therefore, I will provide analysis which either removes the first week or separates the results week-by-week. Second, there is a concern that boredom or a desire for consistency leads people to choose similarly after making many decisions. Therefore, I will potentially restrict the analysis to earlier decisions, exploiting the fact that I assign some subjects in some weeks to only make decisions closer to the work time.

- Controls: When presenting the non-parametric results in the text, I will using controls for wage/individual fixed effects, although I presume that these should not affect the qualitative conclusions. If the results do change substantially given the controls, I will discuss the differences in the text. When possible, I will present results both with and without controls in the text. Given that the graphs will likely only have one set of controls, I will include graphs with different sets of controls in the Appendix if the results appear to differ.

- Demographic information: I do not have strong ex-ante hypotheses concerning the interaction of demographic variables (such as gender, intelligence, or major) with the results. However, I have collected these data mainly for use by future researchers who might have an interest in this data. Therefore, I will not use demographic information in the main analysis. I will present all heterogeneity results in the Appendix regardless of the outcome and make a small reference to the results in the paper.

- Location of error term: For the structural estimation, I believe there are two reasonable locations for the error term. The first location, which I expect to be in the main text, is an additive zero-mean normal-shaped error term on the work decision. The second, which I expect to include in the Appendix, is an error term on the effort-cost parameter gamma.

- Estimation of individual parameters: hopefully, it will be possible to estimate individual parameters for a large set of the subjects. However, for some of the discounting models with multiple parameters, it might be that there are not enough individual observations to achieve this goal. In this case, a second-best solution is to estimate certain individual parameters, but constrain other parameters to be constant across subjects.

- Definition of when work occurs: In the experiment, subjects choose a three-hour window to complete their work (Friday, February 26 10am-1pm) rather than specifying an exact time (Friday, February 26 10:30am). This is done to reduce the possibility that workers have small events arise that are unrelated to time discounting but make it difficult for them to complete work at an exact time. However, this creates endogeneity in the exact timing of work and complicates the task of tightly identifying very short-term (hourly) effects. An alternative analysis is not completely obvious, but it might be possible to use the average work time within the time window or the pre-stated expected work time to test for these effects.

[[[Added 2/19/2016]]] - It is unclear how people think of work that starts now but ends in one hour: do they discount the next task differently than the tasks completed in an hour? In the main analysis, I plan to assume that people treat the entire set of tasks as occurring exactly at the start time. However, I believe it is possible to complete the analysis under the assumption that people discount the set of tasks differently depending on their distance from the start time. This analysis exploits the fact that higher wages lead to larger task choices and therefore place the marginal task

farther away from the start time. I expect that this analysis will not change the conclusions of the main analysis and therefore I plan to include it in the Appendix.

A.19 Experimental Instructions

Welcome:

Thank you for participating in the study. We will begin shortly.

Summary:

This is a study about how many transcription tasks people want to complete for different wages. Today, you will learn about the study. There are **four** more participation dates that occur online. For the next online participation date, you will practice the transcription task. Then, for the final three dates, you can choose to complete more tasks for money. You will decide how many tasks you want to complete for different wages by answering text messages in the week proceeding each date.

Informed Consent

Placed in front of you is an informed consent form to protect your rights as a subject. Please read it. If you would like to choose not to participate in the study you are free to leave at this point. If you have any questions, we can address those now. We will pick up the forms after the main points of the study are discussed.

Anonymity

Your anonymity in this study is assured. Your name will never be recorded or connected to any decision you make here today. Your email and phone number will be collected solely in order to send reminder emails and text messages. The emails and phone numbers will be encrypted when we collect them. Furthermore, immediately following your last payment for the study, your email and phone information will be destroyed and will not be connected to your responses in the study.

Rules

Please turn your cell phones off. Please put away any books, papers, computers, etc. If you have a question at any point, *just raise your hand.* There will be a quiz once we have finished with the instructions. If it is clear that you do not understand the instructions when we review your answers, you will be emailed and removed from the study.

Eligibility for this study:

To continue in this study, you need to be able to complete the full study.

You must be willing to participate on the same day of the week (eg each Monday or each Tuesday...) for four (4) consecutive weeks (excluding today). The future four participation dates require the use of a computer to access the experiment through the Internet.

You will be able to choose your preferred day of the week to participate. For example, if you choose to participate on Tuesdays, you will be required to participate next Tuesday, the following Tuesday, and so on. For each day, you must choose a three-hour time window to start the experiment, which can vary across weeks. For example, you could participate next Tuesday between 9am-12pm, the following Tuesday between 1pm-4pm, and so on. The only requirement for these windows is that one of the windows must start on 11am or before, one must start between 12pm-3pm, and one must start on 4pm or after.

The next participation day is a practice day and will require one hour of participation. After that, each participation day will require at least 20 minutes of your time. On these days, you can choose participate for additional time each day to receive supplementary payments.

In addition, you must be willing to receive no more than 10 text messages each week which will point you to a mobile website to answer a small set of questions. Therefore, you must have a smart phone to participate in the study. You must respond to at least 80% of the text questions within one hour in order to complete the study. You will be able to select a set of times for which you cannot receive text messages (i.e. during times you are in class).

You must be willing to receive your payment from this study as three payments by electronic payment at the end of the study. Payments will be made around 6,7, and 8 weeks from today.

If you do not meet these criteria, please inform us of this now.

Your Earnings

You will be paid a one-time completion payment of $60 for completing the minimum requirements of the study. Furthermore, you will have the chance to earn a supplementary payment of between $1-$25/hour for further participation in weeks 2-4. Based on payments in previous studies, we expect the average payment to be around $100.

It is very important for the study that you participate on your chosen participation days at your chosen time windows. You cannot modify participation dates or times. Unfortunately, if you miss one of your participation dates or do not answer 80% of the text messages, *you will forgo the $60 completion payment and will be immediately removed from the study* (you will receive any supplementary payments you have already earned). **There will be absolutely no exceptions to this rule, regardless of the reason.**

All payments (completion, supplementary, bonus) will be made as payment by three emails at the end of the study around 6,7, and 8 weeks from today, regardless of if you are removed from the study. The completion payment will be equally distributed across these three payments. The supplemental payments for weeks 2-4 will be distributed around 6,7, and 8 weeks from today, respectively.

Choosing Future Participation Days

As stated above, you will participate in the study for the same day of the week for 4 consecutive weeks. Today, you will first select your chosen day of the week (e.g. Monday, Tuesday, etc.) You will then choose a three-hour time window to participate for each of these days between 5am and 12pm. The only constraint is that, for weeks 2-4, one of the windows must start on or before 11am, one must start between 12pm-3pm, and one must start after 4pm.

On future participation dates, you will be texted a reminder of your chosen time the day before your participation date. You must log into the experiment website within your chosen three hour window. If you fail to do this, you will be immediately removed from the study and will forgo the $60 completion payment.

You will now register and pick the times and dates for the experiment. Please open the computer link and listen for instructions.

Task: Greek Transcription Job

For the study, we have designed a task involving transcribing a line of blurry letters from a greek text for this study. In the study, you can choose to do this task for different wages. The task has no value to us beyond understanding these decisions.

We will now spend a few minutes practicing this job on the computer. In the task, Greek text will appear in a Transcription Box on your screen. For each letter you will need to find and select the corresponding letter and enter it into the Completion Box on your screen. *One task is one row of greek text.* For the task to be complete your accuracy must be 80% or better.

As part of the task, an auditory "beep" will sound randomly throughout the transcription process. Please put on your headphones so that you can hear the beeping noise. After you hear this beeping noise, you must press the "noise" button at the bottom left of the screen. If you do not press the "noise" button within five seconds of hearing the beeping noise, your transcription will be reset. If you press the noise button erroneously (when there was no beeping noise), your transcription will be reset.

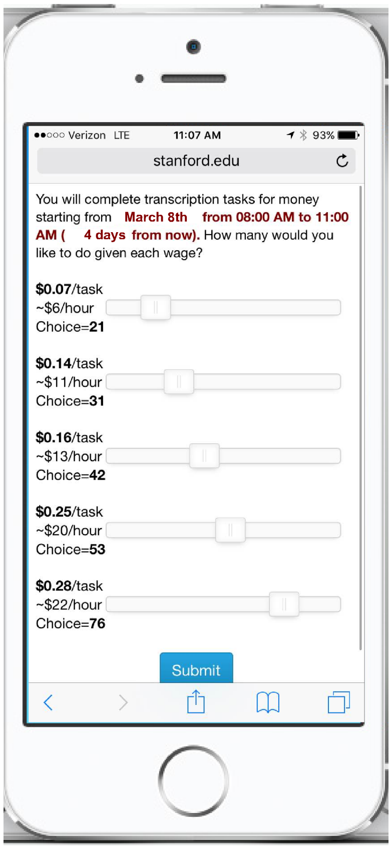
On average, people with some experience complete a task in about 45 seconds (about 80/hour).

Each day of participation, you will have to complete 10 *mandatory tasks* (10 lines of greek text). Furthermore, you will complete additional *supplementary tasks* for supplementary payments. The number of supplementary tasks you must complete on each participation day and the supplementary payment will depend on your choices in the study.

You will now complete 5 of these tasks and then answer some questions about yourself.

Text Messages

This study examines people's decisions about doing supplementary work for monetary payments. We are interested in these work decisions at different points in time. To gather your answers at different times, we will send you text messages with links to a mobile website with work questions. For example, suppose that one of your participation windows is Tuesday, March 8th from 8am to 11am. You will receive text messages in the week prior to March 8th. For example, you might receive a text message on Friday, March 4th at 1pm. That text message will link to a mobile website that looks something like this:



On this screen, you are asked to choose the number of tasks that you want to complete on March 8th for five different wages. For example, on the first line, you are asked the number of tasks you would like to complete for $.07/task to be paid 4 weeks from the decision date. You will use the slider bar to choose a number between 0 and 100. You will also do this for the other 4 wages. The hourly wage estimates have been calculated using a task time of 45 seconds. The wages in each question have been chosen at random from a set of wages from $.01/task to $.31/task. Your decisions cannot in any way affect the choice of future wages—**the wages for this study have already been randomly chosen by a random number generator.**

Each one of these decisions could be randomly chosen as the **decision-that-counts** (the process for choosing that decision-that-counts is discussed below). If a decision is chosen as the decision-that-counts, you must complete that amount of supplementary tasks for the supplementary wage. Therefore, **it is in your own interest to answer honestly about your work preferences**, because you might actually have to complete the work you specify for the given wage. For example, if given a wage of $.12/task, you would like to do 50 tasks (and make an extra $6.00 four weeks from the decision date), you should answer "50."

First Online Participation Date: Practice

For the first participation date, you will complete a total of 50-80 practice tasks.

Later Online Participation Dates: Timeline

For the later participation dates, you will go through a variety of steps:

Answer Questions about Tasks

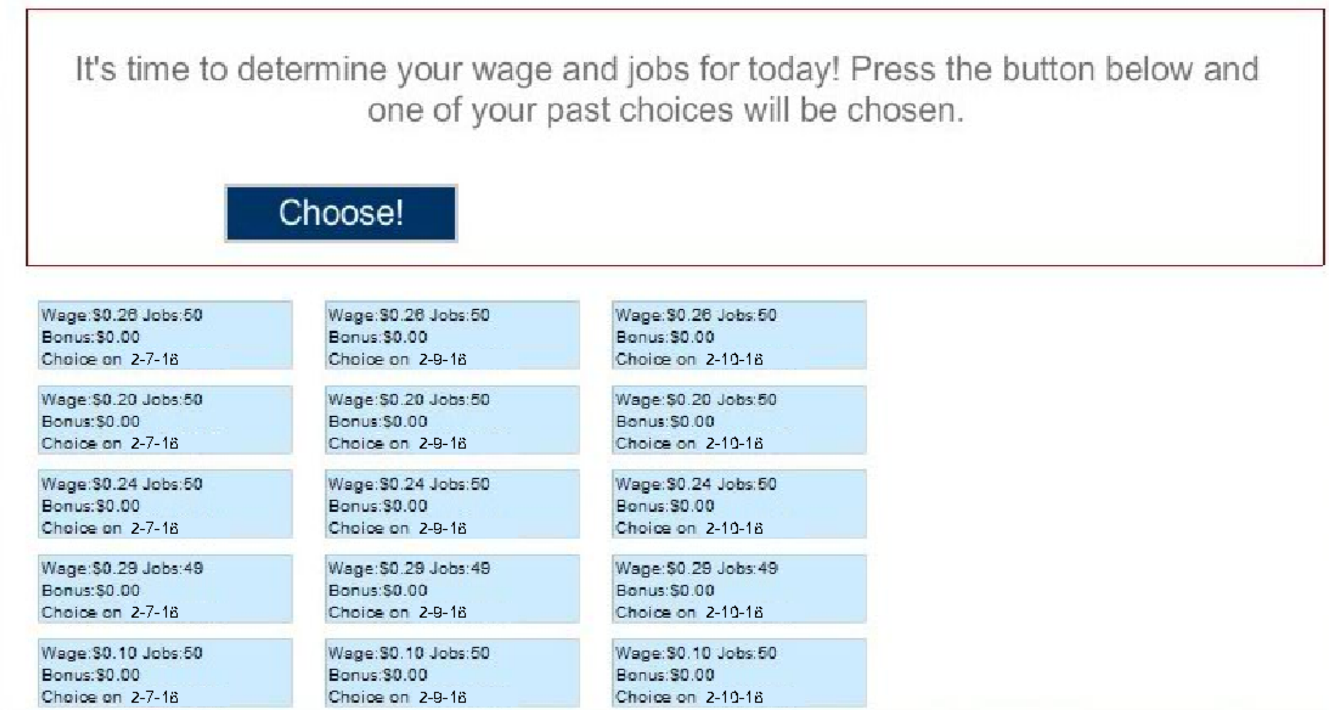
When you first log into the website, you will be sent a final text for that week asking your preferences over the supplemental tasks you want to complete.

Completion of Mandatory Tasks

Recall that you are required to complete 10 mandatory tasks on each participation date. These mandatory tasks require that you set aside at least 15 minutes for each participation date.

Random Selection of the "Decision-That-Counts"

You will be asked questions about how many tasks you want to do in the future and how many tasks you want to do on that day. Therefore, when a given participation date arrives, you will have answered many questions about work on that date. We will collect all of those decisions and randomly choose **one** as the decision-that-counts. This is the screen that collects all of the past decisions:



When you press the "choose" button, one of these decisions will be randomly chosen as the decision-that-counts. The decision-that-counts **has already been chosen by a random number generator in the computer**. You cannot affect how the decision-that-counts is chosen with your choices.

Completion of Supplementary Work

Once the decision-that-counts is chosen, you must complete the amount of supplementary tasks you chose for the wage in the decision-that-counts. For example, if you answered "40" to the question: "*For $.18/task, how many tasks do you want to complete TODAY?",* and this decision is chosen as the decision-that-counts, you would complete 40 supplementary tasks and make a supplementary payment of 40·$.18=$7.20 four weeks from the date of the decision-that-counts.

Recap:

- The study requires participation for four more days over the next 4-5 weeks.
- This is a study about work decisions. We are interested in *decisions* about work for different wages at different points in time.
- You will make decisions about completing tasks involving transcription of greek letters.
- Each participation date, you will be asked to complete a minimum requirement of tasks.
- On the next participation date, you will practice the task.
- For the other participation dates, you will receive text messages asking how many supplementary tasks you would like to complete on the next participation date for different per-task wages. Given the expected pace, these wages are between $1-$25/hour.
- When these participation dates arrive, one of the decisions you made about supplementary tasks will be randomly chosen as the decision-that-counts. You must complete your chosen supplementary number of tasks for the chosen supplementary wage (paid four weeks from the day of the decision-that-counts). Therefore, you should just answer honestly about the number of tasks you would like to complete for each wage decision.
- You will be paid a one-time-completion payment of $60 for completing the minimum requirements of the study on each day. Furthermore, you will be paid for your supplemental tasks at the supplemental rate in the decision-that-counts. You will be paid via email around 6,7, or 8 weeks from today.
- If you choose to no longer participate, or do not complete the jobs you chose, you will forgo the completion payment of $60 and be removed from the study. You will still receive any payments around 6,7, or 8 weeks from today.

Consent

Now that we have explained the study, you are free to leave if you would like to choose not to participate in the study.

Otherwise, please sign the consent form and we will pick these up now.

Quiz

Please now complete the quiz in order to make sure that you understand the study.