Pre-Analysis Plan:
Why Do People Change Their Work Plans?

Zachary Breig, Matthew Gibson, and Jeffrey Shrader*

April 14, 2023

Contents

*Affiliations. Breig: School of Economics, The University of Queensland, Level 6 Colin Clark Building (39), St Lucia, Brisbane, Australia, 4072 (e-mail: z.breig@uq.edu.au); Gibson: Department of Economics, Williams College, 24 Hopkins Hall Drive, Williamstown MA 01267, and IZA (e-mail: mg17@williams.edu); Shrader: School of International and Public Affairs, Columbia University, 420 West 118th Street, New York NY 10027 (e-mail: jgs2103@columbia.edu).

1 Introduction

1.1 Abstract

Research has shown that procrastination has significant adverse effects on individuals, including lower savings and poorer health. Procrastination is typically modeled as resulting from present bias. We study an alternative model of procrastination: excessively optimistic beliefs about future demands on an individual's time. The two models can be distinguished by how individuals respond to information on their past choices. We propose two complementary treatments to test the predictions of the models. If the experimental results refute the hypothesis that present bias is the sole source of dynamic inconsistency, this will have important implications for the large literature on present-biased discounting behavior. Moreover, it will have important practical implications. The findings will offer an explanation, for example, for low takeup of commitment and suggest that personalized information on past choices could instead be an important tool for mitigating procrastination.

1.2 Motivation

Procrastination is an important feature of everyday life that has consequences across a wide range of areas including retirement saving, exercise, and education (Thaler and Benartzi, 2004; DellaVigna and Malmendier, 2006; Ariely and Wertenbroch, 2002). The most widely applied model of procrastination in economics posits that it originates from discounting that is not exponential (Strotz, 1955; Laibson, 1997). In the most common of these discounting-based models, dynamically inconsistent choices come from a utility function that places lower weight on the more distant future relative to the present or near future. The agent exhibits present-biased dynamic inconsistency (procrastination) because choices made far enough in advance will be governed by exponential discounting while choices made about the immediate future will not. If an agent is naẅe about her own present-biased discounting, she believes that she will behave more consistently than she actually does.

Discounting is not the only source of dynamic inconsistency, however. An alternative model based on excessive optimism about future demands on one's time can also lead to procrastination behavior. Importantly, this behavior is observationally equivalent to that of the discounting model in terms of predictions tested by previous experimental work. For intuition on the second model, consider an agent who does not accurately anticipate the arrival of a time-consuming task. Once the task arrives, the agent will need to defer planned time use to accommodate the unanticipated shock, leading to procrastination. For example, a student might be optimistic about how long a problem set will take. When the actual duration of the problem set becomes apparent, the student would need to delay other, planned activities in order to finish their work. One specific model that captures this type of biased belief is the "planning fallacy" model of Kahneman and Tversky (1982). Beliefs are increasingly being seen as an important determining factor for time preferences (Acland and Levy, 2015; Augenblick and Rabin, 2018; Börsch-Supan et al., 2018; Carrera et al., 2021).

Although the two models predict similar procrastination behavior, they can be distinguished by how individuals respond to information. Both models rely on biases in beliefs held by agents. In the discounting model, the biased belief is about the discount rate. For the time-shock model, the biased belief is about future demands on one's time. Providing individualized information relevant to these beliefs can lead to changes in behavior. In response to information about past procrastination, a naïve, discounting-based, dynamically inconsistent agent can learn about her own present bias—learning, for instance that her discounting is more present-biased than she previously thought. This can increase commitment demand for time-use choices made far enough in advance. In contrast, individuals optimistic about their time shocks have erroneous expectations about the external state of the world, so information on previous procrastination can help correct their beliefs and lead them to make more realistic plans.

In previous work (Breig et al., 2021), we experimentally tested the predictions of the two models and rejected the hypothesis that discounting is the sole source of dynamic inconsistency. Instead, both discount rates and beliefs about time shocks matter. This is practically important because the typical prescription for procrastination—offering people the chance to tie themselves to the mast, committing to decisions in advance—does not fully remedy problems from biased beliefs. Personalized historical information can be an important additional tool for helping belief-biased people make consistent decisions over time.

1.3 Research questions

- How does personalized information on a past intertemporal choice over real-effort tasks affect subsequent intertemporal choice?
- How does personalized information on the distribution of future time shocks affect intertemporal choice over real-effort tasks?
- How does personalized information on a past intertemporal choice over real-effort tasks affect subsequent commitment demand?
- How does personalized information on the distribution of future time shocks affect commitment demand?

2 Research design

2.1 Sampling

2.1.1 Sampling frame

Our experiment will employ large representative samples of the U.S. population recruited using the Prolific platform. Similar to Amazon's MTurk, Prolific provides access to on-demand research participants from across the U.S. (Palan and Schitter, 2018). Comparisons between the two platforms show that Prolific subjects are of higher quality (they exhibit greater attention to the experimental task, provide more self-consistent answers, suffer from lower attrition, etc.) than MTurk workers. The data quality from Prolific users is typically also higher than that from samples of undergraduates while having the benefit of greater subject diversity Eyal et al. (2021). The expected sample size is at least 1,000, with the exact count depending on payments to subjects.

2.1.2 Statistical power

We previously executed a similar study at a smaller scale (Breig et al., 2021). A total of 274 subjects completed the study, and we found evidence that both present bias and optimism were significant contributors to procrastination. For a sample of 1,000, we will be able to detect effects that are 52% the size of our previous estimates with 80% power in a test with 5% size.

2.1.3 Assignment to treatment

Subjects will be assigned to one of three equal-size experimental groups (conditions) based on a random number drawn within Qualtrics, without blocking or stratification. Because of this simple randomization procedure, chance imbalances are possible and there will likely be scope to increase statistical power using control variables. Our approach to control variables is described in Section 3.3. The three groups are described in Section 2.3.

2.1.4 Attrition from the sample

Based on pilot experiments in Prolific, attrition is expected. In the most recent pilot, 33 of 150 (22 percent) subjects attrited. We will attempt to reduce attrition in the full experiment through larger lump-sum payments and email reminders. Attrition from the experiment will be assessed by regressing an attrition dummy on all baseline observables and reporting the results of a joint F test.

2.2 Survey details

2.2.1 Instruments

As in Breig et al. (2021), we will use Qualtrics to field the experiments and gather data. Subjects will complete five survey instruments, as illustrated below. Part 1 covers the experimental calendar, consent, demographics, payment explanation, example slider tasks, and comprehension questions. Part 2 includes an IQ quiz and a set of task choices, under different information conditions and piece rates, to be completed in Part 3. A probabilistic commitment device is offered. Part 3 allows subjects to make a new set of choices before performing slider tasks.

Part 4 begins with another IQ quiz. For subjects randomized into information treatments (described in Section 2.3), messages are then presented. Task choices are then made by all subjects. Part 5 allows subjects to make a new set of choices before performing slider tasks.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
This Week					Part 1: Introduction		
Next Week		Part 2: 1.Complete IQ quiz 2.Choose tasks for Part 3			Part 3: 1. Choose tasks for Part 3 2. Complete Part 3 tasks		
Week After Next		Part 4: 1.Complete IQ quiz 2.Choose tasks for Part 5			Part 5: 1. Choose tasks for Part 5 2. Complete Part 5 tasks		Submissions approved and payments sent

Current Part
Future Parts

All survey instruments were developed by the authors. Some elements were taken from the Qualtrics surveys used in Breig et al. (2021). All surveys went through multiple rounds of piloting and revision on Prolific, with the goals of eliminating coding errors and making questions intelligible to subjects.

2.2.2 Data collection and processing

Data collection will occur over approximately four and a half weeks, as detailed in Section 4. Data sets for analysis will be downloaded from Qualtrics in anonymized form, without any sensitive fields such as subject names.

Because processing code has already been written for the pilot data, processing time for the full experimental data is expected to be brief (perhaps 1-2 days at most). The code is quite simple, serving primarily to merge results from the five survey instruments using a randomly generated subject ID.

2.3 Experimental design

We will conduct an experiment evaluating responses to both experimentally induced time shocks and naturally occurring time shocks. Each subject will complete the experiment over five sessions, of which the first is only an introduction. We will inform subjects that they will be offered the opportunity to complete real-effort tasks at a piece rate in sessions three and five of the experiment.

In the second session of the experiment, subjects start by competing in a contest. The contest will be between pairs of participants and will involve completing real-effort tasks (an intelligence quiz). Subjects will be told that the winner of the contest will receive easier tasks in a future session, while the loser will receive more difficult tasks. After the contest, we will elicit each subject's beliefs about the likelihood that she has won the contest. The subjects will then make a series of choices: (1) how many unpaid tasks they would be willing to complete in order to increase the likelihood that their choices in the first session are implemented, (2) how many tasks they would commit to doing in session two at various piece rates *without* knowing the outcome of the contest, (3) how many tasks they would commit to doing in session two at various piece rates assuming that they win the contest, and (4) how many tasks they would commit to doing in session two at various piece rates assuming that they lose the contest.

In the third session of the experiment, which will be completed at least 24 hours after the first, subjects will again choose (1) how many tasks they would commit to doing in this session at various piece rates *without* knowing the outcome of the contest. They will then learn the outcome of the contest and choose (2) how many tasks they would commit to doing in this session at various piece rates given the contest outcome. The subject will then learn which piece rate has been randomly selected to be implemented, and which of their choices (session two or session three, and conditional on contest outcome or not) will be implemented. Then they will be required to complete their chosen number of tasks.

The fourth and fifth sessions of the experiment will be completed in the week following sessions two and three, and will involve the same series of choices with one added step. Between completing the contest and choosing whether they want to commit, subjects will be randomly divided into thirds, with each third receiving one information treatment. In the *no information* (control) condition, subjects will proceed to the belief elicitation and the commitment decision. In the *contest information* treatment, subjects will receive information of the form "We also matched you with two other randomly drawn participants from the previous study, and you (lost against both/won against one/won against both) of them." They will then complete the belief elicitation and the commitment decision. In the *task information* treatment, subjects will complete the belief elicitation then receive information of the form "In Session 2, for a payment rate of Z per set and not knowing whether the sets would be easy or hard, you agreed to complete X1 sets. In Session 3, in the same setting, you agreed to complete X2 sets." The reported comparison of X2 and X1 will be randomly selected from the set of repeated allocations, with uniform probabilities. Subjects in the task information treatment will then be asked, "Why might someone's choices change over time?"

The median time for each session was computed based on pilots, and completion payments for each session were chosen to achieve an hourly payment rate of $12. Accounting for both completion payments and task payments, the average payment per participant was approximately $15 in pilots. Therefore, we expect to pay $15,000 for 1000 participants.

3 Empirical analysis

3.1 Outcomes

Task reallocation in response to contest information The week-two difference between the committed unconditional task allocation and the uncommitted realized allocation. Intuitively, this is the difference between the subject's work plan under probabilistic commitment (elicited in part four) and the work she chooses in part five. Note that her part-five uncommitted plan may not correspond to work performed if her probabilistic commitment binds. Task reallocation is defined for three piece rates: $.06, $.12, and $.18 per set of sliders.

Task reallocation in response to task information The week-two difference between the committed task allocation and the uncommitted allocation. Unlike the above outcome, this difference does not depend on the *realized* allocation (which depends on the outcome of the contest). Intuitively, this is the difference between the subject's work plan under probabilistic commitment (elicited in part four) and the work she chooses (but does not necessarily perform) in part five. Task reallocation is defined for three piece rates ($.06, $.12, and $.18 per set of sliders) and three information conditions (easy, hard, or unconditional). Not all subjects will be observed in all information conditions, as in Part 5 we elicit only the uncommitted allocation corresponding to the subject's contest outcome. Because the contest outcome is conditionally random, however, this selection will not bias estimates in expectation.

Change in commitment demand The task-denominated difference between commitment demand in week two and commitment demand in week one. Subjects will be offered the chance to increase the probability that their ex ante work plan, made in either part two or part four, is binding. More specifically, demand will be elicited using a multiple price list. Subjects will make five binary choices between 1) doing three easy tasks and being committed with probability .2; and 2) doing one to five easy tasks and being committed with probability .8. Commitment demand will equal the number of choices in which the subject selected a higher commitment probability, so it will range from zero to five. Our regression analysis will involve the difference between a subject's week-two commitment demand and her week-one commitment demand.

3.2 Balance checks

A standard covariate balance table will be reported, including all baseline demographic variables. The table will present means for the three experimental conditions and t-tests of the null hypothesis of zero difference between treatment and control conditions. Attritors will also be evaluated for balance. First, attrition rates will be reported for all three experimental conditions. The difference between these rates will not be formally tested, as rejecting the null hypothesis of zero difference does not necessarily imply bias in our estimators. Second, attritors and non-attritors will be compared on observables. Let A be an attrition dummy, Z a dummy for the contest information treatment, W a dummy for the task information treatment, and X a vector of baseline observables including a wave indicator (see Section 4 for an explanation of waves). Let i index subjects. The following regression will be estimated.

$$A_i = \kappa_1 Z_i + Z_i \boldsymbol{X_i'} \boldsymbol{\kappa_2} + \kappa_3 W_i + W_i \boldsymbol{X_i'} \boldsymbol{\kappa_4} + \boldsymbol{X_i'} \boldsymbol{\kappa_5} + \omega_i \tag{1}$$

The F statistic for this regression will be reported. This will test whether attritors and non-attritors differ on observables, including treatment. A version of this regression without the treatment variables will also be reported, including subjects who attrited prior to randomization. Lee bounds for estimated effects on primary outcomes will appear in the appendix.

3.3 Treatment effects

Let treatment variables, demographic variables, and the subject index be denoted as in Section 3.2 above. In addition, let r index piece rate (as described in Section 2.3). For outcomes, let T_1 denote task reallocation in response to contest information (commited unconditional minus uncommited realized), T_2 task reallocation in response to task information (committed minus uncommitted, within piece rate and information condition), and C the change in commitment demand. Let $\boldsymbol{N}$ (mnemonic: number) be a vector of indicators for zero, one and two wins out of a possible two against randomly selected competitors, which will be reported to subjects in the contest information treatment ($Z_i = 1$). Let R be the randomly selected week-one reallocation, which will be reported to subjects in the task information treatment ($W_i = 1$).

3.3.1 Intent to treat

The estimating equation for task reallocation in response to contest information will be as follows.

$$T_{1ir} = Z_i \boldsymbol{N_i'} \boldsymbol{\beta} + \boldsymbol{N_i'} \boldsymbol{\gamma} + \boldsymbol{V_i'} \boldsymbol{\delta} + \zeta W_i R_i + \eta R_i + \boldsymbol{X_i'} \boldsymbol{\theta} + \boldsymbol{\kappa_r} + \varepsilon_{ir} \tag{2}$$

Each subject will contribute a maximum of three observations to this regression, one per piece rate r. Observations for piece rates under which the subject's corresponding week-one choice was censored (0 easy tasks or 20 hard tasks) will be excluded from the sample to increase statistical power. Such selection on a predetermined characteristic does not introduce bias.

The coefficients of interest are β_0, β_1, and β_2 (elements of $\boldsymbol{\beta}$ corresponding to zero, one, and two wins). Task information treatment W and reported reallocation R are included as controls, but their estimated coefficients will not be presented or interpreted, as task-information treatment effects will be estimated using equation (3). The vector $\boldsymbol{V}$ (mnemonic: victory) contains a set of indicators for week-one and week-two contest score, and indicators for whether the subject won the week-one and week-two contests against her randomly drawn opponent. The controls for contest score are endogenous, but one may think of them as reflecting type or ability. The coefficient on winning the week-one contest may reflect some learning that occurs as a result of completing easy tasks (as compared to hard). Demographic variables and a wave indicator will be included in $\boldsymbol{X}$.[1] The vector $\boldsymbol{\kappa_r}$ represents piece-rate fixed effects.

The estimating equation for task reallocation in response to task information will be as follows.

$$T_{2irc} = Z_i \boldsymbol{N_i'} \boldsymbol{\beta} + \boldsymbol{N_i'} \boldsymbol{\gamma} + \boldsymbol{V_i'} \boldsymbol{\delta} + \zeta W_i R_i + \eta R_i + \boldsymbol{X_i'} \boldsymbol{\theta} + \boldsymbol{\kappa_{rc}} + \varepsilon_{irc} \tag{3}$$

Each subject will contribute a maximum of six observations to this regression, one per piece rate r and one per information condition (unconditional and hard/easy)[2]. Again observations

[1] Algorithmic control selection will be implemented as a robustness check (see Section 3.5).

[2] As mentioned previously, subjects will make uncommitted allocations for either the hard or the easy information condition

for piece rates under which the subject's corresponding week-one choice was censored (0 easy tasks or 20 hard tasks) will be excluded from the sample to increase statistical power.

The coefficient of interest is ζ. The contest information treatment Z and wins indicators in $\boldsymbol{N}$ are included as controls, but their estimated coefficients will not be presented or interpreted, as task-information treatment effects will be estimated using equation (2). The vector $\boldsymbol{\kappa_{ir}}$ represents piece-rate-information-condition fixed effects. Other controls are identical to those in equation (2).

The estimating equation for commitment demand will be strongly similar. As commitment did not differ by piece rate, however, each subject will contribute only a single observation to the sample.

$$C_i = Z_i \boldsymbol{N_i'}\boldsymbol{\beta} + \boldsymbol{N_i'}\boldsymbol{\gamma} + \boldsymbol{V_i'}\boldsymbol{\delta} + \zeta W_i R_i + \eta R_i + \boldsymbol{X_i'}\boldsymbol{\theta} + \varepsilon_i \quad (4)$$

Again, the coefficient of interest is ζ.

These regressions impose potentially consequential functional form assumptions. First, the estimating equation imposes linearity of the outcomes with respect to the randomly selected week-one reallocation R. If this assumption were violated, then using subjects in the task information treatment as an input to the counterfactual outcomes for subjects in the contest information treatment could cause bias. Second, functional form choices in controls for subject type (elements of $\boldsymbol{V}$) could influence whether using subjects in the contest information treatment as an input to the counterfactual for subjects in the task information treatment causes bias. In ongoing work, we are using simulations to evaluate the impact of functional form choices on bias and statistical power. It may be necessary to allow the nuisance treatment to enter each ITT regression non-parametrically (e.g. a full set of dummies for possible values of R), or estimate contest-information effects in a sample without task-information-treated subjects (and vice versa).

3.3.2 Treatment on the treated

The treatments in this experiment are informative messages that appear during a survey (see Section 2.3). For non-attritors takeup is either deterministically complete (if takeup is seeing the message) or unobservable (if takeup is internalizing the message). Therefore separate TOTs will not be estimated, and ITTs will be discussed as equal to TOTs in our setting.

3.4 Heterogeneous effects

We plan to estimate heterogeneous treatment effects by piece rate, in part because the piece rate influences the degree of censoring in responses. In addition, we will estimate heterogeneous ITTs on the dimensions of risk and time preferences. Standard economic theory predicts that these preferences mediate the behavioral response to a given change in beliefs over time shocks. Regressions will follow equations (2) and (3), but $\boldsymbol{\beta}$ and ζ will interact with measures of heterogeneity.

We intend to classify subject behavior as consistent with present-biased preferences, biased beliefs, or both using an adapted version of the procedure from Breig et al. (2021). In brief, this procedure matched treatment and control subjects, then compared individual task-reallocation and commitment-demand responses.

3.5 ML control selection

As a robustness check, we will employ machine-learning techniques to choose a precision-maximizing control set. This is consistent with the recommendation of Ludwig et al. (2019). According to Wager et al. (2016), ridge regression, LASSO, elastic net, and random forest procedures can all be used to improve efficiency without introducing bias into estimated treatment effects.

3.6 Statistical inference

Standard errors will be clustered at the subject level for task reallocation regressions. Heteroskedasticity-robust standard errors will be used for commitment demand regressions. For the contest information treatment, we expect that positive (negative) news about the likelihood of winning the first contest should make subjects more (less) optimistic about winning the second contest, and thus should increase (decrease) their willingness to complete tasks in the unconditional information condition. Thus, our one-sided test will be against the null hypothesis that $\beta_2 - \beta_0 \leq 0$. For the task information treatment, we expect that reminders about past task reallocation tempers optimism (lowering task reallocation) and makes subjects believe that they are more present biased (increasing commitment demand). Thus, our one sided tests will be against the null hypotheses that $\zeta \geq 0$ in equation (3) and that $\zeta \leq 0$ in equation (4).

4 Calendar

The experiment will take place in two waves. Wave 1 will run from April 6, 2023 through April 21, 2023.[3] Wave 2 will run from April 20, 2023 through May 5, 2023. Analysis of the data will take place beginning May 6, 2023. Paper submission is expected by mid June, 2023.

[3]Randomization and treatment will not begin until April 17.

References

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

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

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

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

Breig, Z., M. Gibson, and J. Shrader (2021). Why do we procrastinate? present bias and optimism. Technical report, IZA Working Paper.

Carrera, M., H. Royer, M. Stehr, J. Sydnor, and D. Taubinsky (2021). Who chooses commitment? evidence and welfare implications. *The Review of Economic Studies*.

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

Eyal, P., R. David, G. Andrew, E. Zak, and D. Ekaterina (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1–20.

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

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

Ludwig, J., S. Mullainathan, and J. Spiess (2019). Augmenting pre-analysis plans with machine learning. In *AEA Papers and Proceedings*, Volume 109, pp. 71–76.

Palan, S. and C. Schitter (2018). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance 17*, 22–27.

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

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

Wager, S., W. Du, J. Taylor, and R. J. Tibshirani (2016). High-dimensional regression adjustments in randomized experiments. *Proceedings of the National Academy of Sciences 113*(45), 12673–12678.