

# Inattention, task completion, and retention among undergrads

by

**Kevin Laughren**

M.A., University of Calgary, 2013

B.Comm., Queen's University, 2008

Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Doctor of Philosophy

in the  
Department of Economics  
Faculty of Arts and Social Sciences

©Kevin Laughren 2021  
**SIMON FRASER UNIVERSITY**  
**Fall 2021**

Copyright in this work is held by the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

# Declaration of Committee

**Name:** Kevin Laughren

**Degree:** Doctor of Philosophy

**Thesis title:** Inattention, task completion, and retention among undergrads

**Committee:** **Chair:** Simon Woodcock  
Associate Professor, Economics

**David Freeman**  
Supervisor  
Associate Professor, Economics

**Jasmina Arifovic**  
Committee Member  
Professor, Economics

**Arthur Robson**  
Examiner  
Professor, Economics

**Victor Aguiar**  
External Examiner  
Associate Professor, Economics  
Western University

## Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

- c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library  
Burnaby, British Columbia, Canada

Update Spring 2016

# Abstract

Three economic experiments are used to answer questions about human decision making: when to pay attention, when to complete a task, and when to enroll in second year university. Chapter 1 explores inattention as an explanation for economic mistakes like overpaying. A novel experiment design separates intention to collect information from its actual collection and use. The results show that many mistakes are the result of optimal inattention, when someone is inattentive because the costs of information exceed the benefits of paying attention. But the experiment also demonstrates addressable mistakes which are driven by dynamic inconsistency and ineffective attention, even in a very simple information setting. Chapter 2 explores procrastination through a novel real-effort task experiment with no commitment. Unlike many previous economic experiments on intertemporal choice which use monetary rewards or record choices at a single point in time, this experiment finds participants have a preference to complete the real-effort task early, in line with psychological experiments on working memory. Structural estimation of the classic model of hyperbolic discounting finds the present-bias parameter  $\beta > 1$ , indicating a future-bias and desire to finish costly tasks immediately. Chapter 3 uses a novel control group in a field experiment to study the causal effect of a first-year seminar on undergraduate student outcomes such as enrollment, academic standing, GPA, and psychological well-being. Students enrolled in seminars have higher GPA and retention than the overall pool of students, but these effects disappear when comparing seminar students to the novel control group of students who were willing but not enrolled in a seminar. Self-selection into a seminar course may be indicative of student learning style and strengths in ways that are not measured by registrar data; this experiment demonstrates that studies that fail to control for willingness to enroll are at risk of identifying selection effects rather than genuine effects of a seminar. The participants sampled in all chapters are drawn from the same undergraduate population, though they are distinct.

**Keywords:** economic experiments; limited attention; ineffective attention; time preferences; future bias; student retention; selection effects

# Dedication

For my father Les and my grandfather Wyman.

# Acknowledgements

David Freeman provided substantial supervision of my development as a researcher throughout my PhD. Jasmina Arifovic provided guidance and supported my development as a researcher and teacher. I could not have done this work without them. This research was conducted during my doctoral studies, which have been partially funded by the Bombardier Fellowship from Social Sciences and Humanities Resource Council and the Peter Kennedy Memorial Graduate Fellowship from Simon Fraser University.

# Table of Contents

Declaration of Committee	ii
Ethics Statement	iii
Abstract	iv
Dedication	v
Acknowledgements	vi
Table of Contents	vii
List of Tables	ix
List of Figures	x
<b>1 Decomposing Inattention</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Model and Experiment Design . . . . .	5
1.3 Results . . . . .	12
1.4 Related Literature . . . . .	21
1.5 Discussion . . . . .	23
<b>2 Early Completion of a Real Effort Task</b>	<b>25</b>
2.1 Introduction . . . . .	25
2.2 Theoretical Framework . . . . .	28
2.3 Experiment Design . . . . .	30
2.4 Results . . . . .	34
2.5 Discussion . . . . .	41
<b>3 Selection Effects in a First Year Seminar</b>	<b>42</b>
3.1 Introduction . . . . .	42
3.2 Methods . . . . .	46
3.3 Results . . . . .	49

3.4 Discussion . . . . .	52
<b>Bibliography</b>	<b>55</b>
<b>Appendix A Supplements to Chapter 1</b>	<b>62</b>
A.1 Applying Model Axioms . . . . .	62
A.2 Appendix: Additional Screenshots . . . . .	66
A.3 Appendix: Experiment Instructions . . . . .	67
<b>Appendix B Supplements to Chapter 2</b>	<b>77</b>
B.1 Additional Tables of Results . . . . .	77
B.2 Structural Model Parameter Estimates and Standard Errors: . . . . .	78
B.3 Data Generation and Classification . . . . .	80
<b>Appendix C Supplement to Chapter 3</b>	<b>82</b>
C.1 Correlation of Well-being Measures and Covariates . . . . .	82



# List of Tables

Table 2.1	Experiment Effort Schedules . . . . .	33
Table 2.2	Classifying Observed Choice Quads by Effort Schedule (All data) . .	35
Table 2.3	Classifying Non-endogenous subsample . . . . .	35
Table 2.4	Proportion choosing to work Today when a choice is available . . . .	38
Table 2.5	Time Invariance Violations by Choice Set . . . . .	39
Table 2.6	Structural Estimates of Time Preference Parameters $\beta, \delta$ . . . . .	40
Table 3.1	Baseline Measures . . . . .	50
Table 3.2	Seminar Effect on GPA . . . . .	51
Table 3.3	Seminar Effect on Survey Measures of Well-being . . . . .	52
Table 3.4	Seminar Effect on Academic Standing after two semesters . . . . .	53
Table 3.5	Seminar Effect on Second-year Enrollment . . . . .	53

# List of Figures

Figure 1.1	Timeline of Experiment . . . . .	7
Figure 1.2	All Observed Choices by Stage . . . . .	17
Figure 1.3	Classifying Mistakes: All Choices . . . . .	18
Figure 1.4	Classifying Mistakes: Manual Choice Environment. . . . .	19
Figure 1.5	Implied Average Time Costs and Cognitive Costs . . . . .	22
Figure 2.1	Experiment data generating process . . . . .	31
Figure 2.2	Participant Houtman Maks Index - Monotonicity . . . . .	37

# Chapter 1

## Decomposing Inattention

Kevin Laughren

### 1.1 Introduction

Gathering information is costly. Yet it can inform a decision-maker about the available options. When the cost of gathering information exceeds the benefit, it is optimal for a decision-maker to make their decision under incomplete information. For example, overpaying is optimal when, in expectation, the cost of finding a lower price is greater than the savings.

Limited attention theories describe decision-makers who first gather information subject to some constraint, then make an action choice. This literature begins with the model of rational inattention by [Sims, 2003], and is generalized by the attention cost model of [Caplin and Dean, 2015].<sup>1</sup> A decision-maker in this literature will be inattentive to some information as an ex ante optimal response to their attention cost relative to the benefits of information. The amount of information gathered by a decision-maker is ex ante optimal, regardless of whether this leads to an ex post mistake in the final action choice. Following the literature, I use the term *mistake* to refer to any action that does not maximize a decision-maker's expected payoff under complete information.

Using an experiment, I am able to identify optimal mistakes separately from nonoptimal mistakes by considering ways in which a participant's information gathering may be nonoptimal.<sup>2</sup> Practically, this involves structuring the experiment to observe a participant's

<sup>1</sup>The optimization of information has been a topic of economics since as early as [Hayek, 1945] which discussed prices as a means of decentralizing knowledge and [Stigler, 1961] which discussed diminishing returns to search when sellers have random prices. Alternative models of limited attention which inspired this paper include [Masatlioglu et al., 2012] which shows how action choices reveal a decision-maker's preceding information choices, and [Manzini and Mariotti, 2014] which shows conditions under which a decision-maker considers only subsets of the available actions before making an action choice.

<sup>2</sup>I use *decision-maker* to refer to a theoretical agent who makes choices, and *participant* to refer to a person observed in an experiment.

intent-to-collect information separately from their actual information collection. Additionally, I can categorize the various types of nonoptimal mistakes by observing the stage at which a participant deviates from the optimal choice process. The majority of participant mistakes are cases of optimal inattention (OI) – when information costs are large and information benefits are small – but there are two substantial causes of mistakes which were not previously described in the context of limited attention.

One cause of nonoptimal mistakes is *ineffective attention*, which occurs when a participant has the information to make the optimal choice but fails to do so. Another cause is *dynamic inconsistency*, which occurs when a participant makes internally inconsistent choices across time – for example, when a participant initially states their intent-to-collect information but later makes a choice without having gathered that information. These two causes represent most of the nonoptimal mistakes observed in this experiment, but I document other causes of nonoptimal mistakes as well.

In the limited attention literature, decision-makers have an internal attention cost function that determines the expected cost of gathering information. While this internal attention cost function is unobservable to experimental researchers, we can back out some of its structure if we know the expected benefits of information to decision-makers and observe their intent-to-collect information. Toward this end, the experiment makes explicit for participants the exact benefits of information, and captures their subsequent choices to gather information. The resulting data reveal enough about the attention cost function to identify when mistakes are due to OI, nonoptimal information gathering, or nonoptimal use of information at the action choice stage.

To learn more about the ways in which participants differ from the decision-makers in limited attention models I split the information gathering stage into intent-to-collect versus information-collection, and relax the assumption that mistakes must be consistent with optimal decision-making in limited attention models. This design allows for nonoptimal mistakes in each of three stages as follows:

1. Intent-to-collect stage: internal inconsistencies can occur, e.g., when a participant states an intent-to-collect information to obtain a specific benefit, but also states an intent-to-not-collect the same information for a larger benefit.
2. Information-collection stage: dynamic inconsistency can occur, e.g., when a participant actually collects less information than they stated in the intent-to-collect stage.
3. Action stage: ineffective attention can occur, e.g., when a participant collects sufficient information to identify the optimal action but fails to choose it.

While a single participant mistake can always be explained as OI by assuming high attention cost, a set of choices made by a participant implies some bounds on their attention

cost, and these bounds can be used to rule out OI in some cases.<sup>3</sup> Practically this is done by testing if a participant’s choices are consistent with the axioms of [Caplin and Dean, 2015], which are necessary and sufficient conditions for a dataset to be generated by optimal decision-making under general attention costs and information benefits.

The experiment has 10 paid rounds and in each round there are two equally probable states. A participant wins a monetary prize in a round if they correctly predict the state. The experiment presents the problem to participants as the opportunity to predict the binary element (red or blue) that follows a specific prompt, and the prompt is presented as a sequence of binary elements.<sup>4</sup> In each round there is a finite set of possible sequences that act as prompts (e.g., all binary sequences of length four), and each sequence has one randomly-determined state that follows. The number of possible sequences in the round determines the expected time it will take a participant to learn the true state that follows their specific prompt.

[Zhong, 2019] shows that the theoretically optimal information gathering process for a decision-maker in a context like the experiment – with two states and a prior belief they are equally likely – involves observing a Poisson process and stopping immediately after the arrival of a signal that fully reveals the state. The experiment imposes this Poisson process information flow on participants but allows them to stop at a time of their choosing. A participant who stops before, or significantly after, the fully revealing signal arrives will incur unnecessary time costs without any marginal information benefits. A decision-maker in a limited attention model with constant time costs would not exhibit this dynamically inconsistent pattern of behaviour.

The experiment informs a participant of the prize values and expected time at the beginning of a round, then the participant decides whether they want to gather information or make an uninformed guess. Gathering information takes time, but an uninformed guess of the state provides a 50% chance of winning and allows a participant to proceed immediately to the next round. Participants who choose to gather information face a Poisson process information flow. This information flow is framed as the opportunity to randomly sample one sequence every 10 seconds (with replacement) and observe the state that follows. The participant can record this observation in a table with a button click. Only one of the sequences in the finite set is relevant for the participant, and when that sequence is observed (the arrival of the Poisson signal) the state is fully revealed, improving the (theoretical) probability of winning a prize from 50% to 100%. A participant’s ex ante choice to gather

<sup>3</sup>Attention costs presumably vary across participants due to differences in opportunity costs and preferences to exert cognitive effort, but the attention cost function is assumed to be unchanged across a participant’s decisions.

<sup>4</sup>Giving participants an opportunity to gather information before entering a prediction using red and blue shapes to represent bits of information is inspired by [Dean and Neligh, 2019].

information for a given prize is thus made under known information benefits and is affected by their personal (and unobserved) attention costs.

In the first stage of a round a participant faces a list of prize values and must decide *ex ante* for each prize whether to gather information or guess the state immediately and proceed to the next round. In each round, the experiment randomly selects one prize and implements the participant’s choice for that prize. This first stage data allows for classification of attention choices into “optimal” – those that are optimal given the known information benefits and implied attention costs – and “suboptimal” – those choices which imply a participant could have achieved greater expected monetary benefits for the same expected attention costs. Seventy-five percent of the mistakes I observe in the experiment are identified as OI at this stage.<sup>5</sup>

The second stage of a round is the actual gathering of information. Every 10 seconds, one of the sequences is randomly drawn, and the state that follows is displayed. The participant may collect this information in a table with a button click but if they fail to click within the 10-second window the information is not added to the table for future reference. Participants can stop gathering information at any time and use the table to make a third stage action choice. I identify cases of *dynamic inconsistency* when a participant stops before drawing the relevant sequence, because they previously indicated a preference to gather information to improve their probability of winning a prize, but by stopping before drawing the relevant sequence they indicate that preference has changed. Dynamic inconsistency accounts for 8% of all mistakes I observe in the experiment. The information gathering process can be described as optimal dynamic information acquisition with two states, a flat prior, and information that arrives in a Poisson point process.

In the third stage of a round, participants must choose an action. Participants may either use the information they have collected in the table to make a specific prediction of the state (red or blue), or they may randomize. Randomization is programmed to win the prize with 51% probability to eliminate indifference. I identify cases of *ineffective attention* when a participant chooses an action other than that which maximizes expected payoffs given the information collected in the second stage. Ineffective attention accounts for 12% of all mistakes I observe in the experiment.

There are two versions of the third stage, and all participants complete five rounds of each version. In the *automatic environment* a participant is prevented from making any prediction of their own, but the computer will choose optimally on their behalf using the information the participant collected in the second stage. In contrast, in the *manual environment* participants are required to choose an action. By measuring a participant’s choices in both environments I can estimate their attention cost as a sum of time cost (affecting

<sup>5</sup>Two other types of mistakes can be made in the first round: suboptimal inattention (5% of mistakes) and suboptimal attention (which does not always lead to a mistake, but preceded 1% of mistakes).

both environments) and cognitive cost (affecting only the manual environment). Participants tend to pay equal amounts of attention in the automatic and manual environments, suggesting cognitive cost is small relative to time cost.

To summarize the experimental outcomes, participant choice data is highly consistent with the axioms of [Caplin and Dean, 2015], implying that participants make attention choices as if they have a known attention cost function. 75% of all observed mistakes are the result of optimal inattention in the intent-to-collect stage, but two types of nonoptimal mistakes that are novel in this context are observed in the information gathering and processing stages.

Section 1.2 outlines a model of costly attention, the experiment design, and the testable conditions the model imposes on the experiment data. Section 1.3 details the results of the experiment. Section 1.4 reviews related literature and the contributions of this experiment, and Section 1.5 concludes. Appendix A includes a detailed discussion of the axioms of [Caplin and Dean, 2015], experiment screenshots, and a full copy of the instructions.

## 1.2 Model and Experiment Design

Consider a decision-maker (DM) who has the opportunity to gather information before making a choice, and for whom gathering information is costly. This problem has often been described as a two-part process: first the DM chooses an information structure at a cost, then chooses the action with the greatest expected benefit given the new beliefs (posteriors) formed by the information structure. For a very general set of attention costs and information structures, [Caplin and Dean, 2015] showed that choice data generated by such a process must satisfy two conditions. First, each final action choice must be optimal given the posterior.<sup>6</sup> Second, the information structure choices must be collectively optimal.<sup>7</sup>

Choosing the best action given the information at hand is a familiar context, but the first stage choice of information structure and the process of forming beliefs given that structure is an abstract context and there have been many novel modeling approaches to reflect this [Sims, 2003, Manzini and Mariotti, 2007, Manzini and Mariotti, 2014, Gabaix, 2014, Enke, 2017, Caplin et al., 2018]. This paper presents a three-stage choice process experiment in which the information structure is decomposed and measured in two distinct stages: intent to collect information and collection of information. The third stage is typical, a DM chooses an action given the available information.

<sup>6</sup>NIAS - No Improving Action Switches: a DM must always choose optimally given the observed posterior.

<sup>7</sup>NIAC - No Improving Attention Cycles: Over multiple decision problems, a DM's choices to gather information must not be able to be re-arranged across problems to yield greater information benefits at no additional attention cost.

The underlying decision problem is to predict the binary state (red circle or blue triangle) that follows a specific sequence of binary elements, with correct predictions awarded a monetary prize. Participants can immediately randomize their prediction and be correct with 51% probability, or incur time costs to improve their probability of a correct prediction by gathering information. A participant who gathers information is shown a randomly drawn sequence and the state that follows every 10 seconds.<sup>8</sup> If the randomly drawn sequence matches the specific sequence from the decision problem, the state is fully revealed, but all other randomly drawn sequences have no information about the state. In this very simple environment it is possible to distinguish three sources of mistakes: optimal inattention, dynamic inconsistency, and ineffective attention. In the first stage participants can choose not to gather information, and this inattention is often optimal for small prizes and large expected time costs. In the second stage, participants can be impatient and stop gathering information before possibly learning anything, which I attribute to dynamic inconsistency. In the third stage, participants can fail to choose optimally despite being previously presented with information fully revealing the state, which I call ineffective attention. Figure 1.1 outlines the three stages as seen by an experiment participant and the interpretation of mistakes at each stage.

## Model Primitives

The DM is assumed to have a utility function over prize values ( $u(P) \in \mathbb{R}$ ) and an expected utility function that is linear in probabilities. There are two possible states of the world  $\omega \in \{R, B\}$  (red or blue), and the DM has correct prior belief  $\mu(R) = \mu(B) = \frac{1}{2}$ . There are three possible actions  $a \in \{r, b, g\}$  (predict red, predict blue, or guess), and a gross payoff for each action-state combination. Participants who choose action  $a = r$  when the true state is  $\omega = R$  win prize  $P$  with certainty. Participants who choose  $a = r$  when the true state is  $\omega = B$  win nothing with certainty.<sup>9</sup> Participants who choose  $a = g$  win prize  $P$  with probability=51% regardless of the true state. The 51% prize probability implies choosing  $a = g$  is the unique payoff-maximizing action for a participant whose belief has not changed from the prior. Had I designed the experiment so that choosing  $a = g$  resulted in 50% prize probability, the DM would be indifferent between all actions when they believe both states are equally likely.

In each decision problem, the DM makes a first stage binary choice of information structure  $\pi \in \{\pi_0, \pi_L\}$  (where  $L$  is the length of preceding sequence for the round) and has some unobserved expected attention cost  $K(\pi) \in \mathbb{R}$ . An information structure is a mapping

<sup>8</sup>These sequences are drawn with replacement from the set of binary sequences of the same length as the sequence presented in the decision problem. The length of sequence varies from 3 to 7 elements across rounds, doubling the expected time with each additional element.

<sup>9</sup>The action-state payoffs for  $a = b$  (predict blue) are analogous to  $a = r$ .



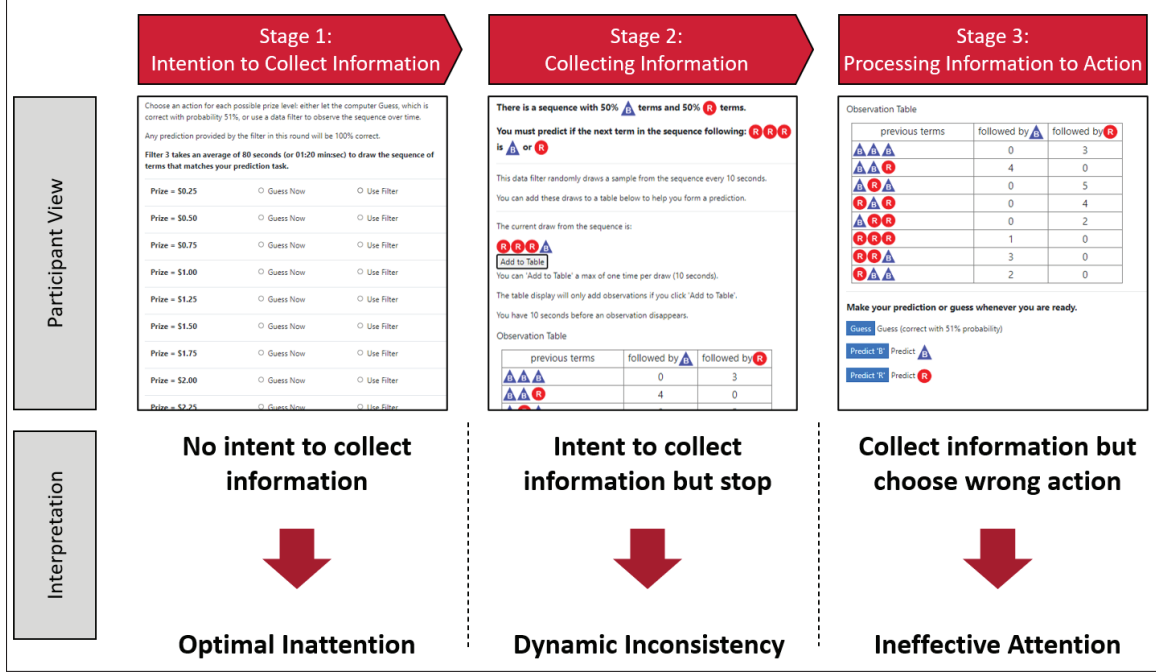


Figure 1.1: Timeline of Experiment

A three stage choice process experiment separating the intention to collect information from the actual collection and use of information.

from the state to a set of possible posteriors about the state. The inattentive information structure  $\pi_0$  generates the posterior  $\gamma(R) = \frac{1}{2}$  because it collects no information to update the prior. The attentive information structure  $\pi_L$  generates a distribution of posteriors  $\Gamma(\pi_L)$  which depends on the number of IID observations, but it always begins as  $\gamma(R) = \frac{1}{2}$  and eventually results in  $\gamma(R) \in \{0, 1\}$  as the IID sampling process eventually reveals the state. I normalize the attention cost of being uninformed to zero ( $K(\pi_0) = 0$ ).

### First Stage: Intent to Collect Information

The choice to collect information is restricted to be binary. A DM who chooses the inattentive information structure  $\pi_0$  must guess immediately with a 51% chance of winning the prize. Alternatively the DM may choose the attentive information structure  $\pi_L$  to randomly sample sequences of length L and record the states that follow each draw before making a prediction. The gross benefit of choosing to gather information is therefore an improvement in the prize probability from 51% to 100%. Participants choose this for 20 different prize levels ( $P = \{\$0.25, \$0.50, \dots, \$4.75, \$5.00\}$ ) in each of ten paid rounds with no feedback. The gross benefit is denoted  $G(\pi, A)$  where  $A$  denotes a unique decision problem.<sup>10</sup> The net

<sup>10</sup>A decision problem is a choice for a single prize in a single round, formally it is an element of  $\{\$0.25, \$0.50, \dots, \$4.75, \$5.00\} \times \{3, 4, 5, 6, 7\} \times \{\text{'auto'}, \text{'manual'}\}$ .

benefit is  $G(\pi, A) - K(\pi)$ , where  $K(\pi)$  is not directly observed but assumed to be known by the DM.

I assume that  $K(\pi)$  is increasing in the expected time to observe useful information and the DM’s beliefs about their personal probability of ineffective attention. The expected time in seconds is  $10 * 2^L$  with  $L \in \mathbb{L} = \{3, 4, 5, 6, 7\}$ . A DM faces each sequence length  $L$  twice, once in the manual environment and once in the automatic environment. I assume  $K(\pi)$  can be split into additive terms for time cost  $T(\pi)$  and cognitive cost  $C(\pi)$ ;  $T(\pi)$  is estimated using choices in the automatic environment where there is plausibly no cognitive cost (just a need to click a button every 10 seconds), and  $C(\pi)$  is estimated by comparing choices in the automatic and manual environments.

Ten paid rounds vary the expected attention cost. Each of the five sequence length  $L$  conditions are completed twice, once in the automatic environment and once in the manual environment. In the manual environment only, participants report what they believe is their personal probability of ineffective attention ( $q(\pi) \in \mathbb{R}$ ) using an incentive compatible binarized scoring rule [Karni, 2009, Holt and Smith, 2009].

## Second Stage: Information Collection

Every 10 seconds, the experiment draws a random sequence and the subsequent state is revealed. The DM has 10 seconds to add this observation to a table with a button click; when the next observation is drawn, the previous observation can no longer be recorded. The observation table is organized to exactly mirror the prediction task: the table presents every possible sequence, and counts the frequency of observed red and blue states that follow. There is no limit on the length of time spent gathering information (in the manual environment), but the state is fully revealed upon observing the sequence of interest one time.

When a DM fails to add an observation to the table with a click, a difference is created between the information that has been presented and the information that has been collected. When a missed observation would have revealed the state, the “true” posterior differs from the “observed” posterior that would be formed using only the recorded observations. When this occurs I classify participants who could have collected information that revealed the state but failed to do so as demonstrating ineffective attention, even though they could be classified as dynamically inconsistent if I were to use the “observed” posterior.

## Third Stage: Translating Information to Action

The IID nature of the prediction task restricts the posterior space to three possible beliefs:  $\gamma(R) \in \{0, \frac{1}{2}, 1\}$ , and until the relevant observation has been drawn,  $\gamma(R) = \frac{1}{2}$  is the

only plausible belief.<sup>11</sup> Because there are only three plausible posteriors and three actions, it is easy for me to check if a participant chooses the optimal action given the available information.

## Data

Participants completed ten paid rounds with no feedback, five in each of the automatic and manual environments. In the automatic environment participants make attention decisions that incur time costs but no cognitive costs, as the computer interprets the data a participant clicks and makes the optimal prediction of red or blue as soon as the state is revealed. In the manual environment, participants face the same set of decision problems but must interpret the data and choose an action themselves.

Each participant provides their choice of information structure  $\pi \in \{\pi_0, \pi_L\}$  in each decision problem  $A$ . There are 10 paid rounds each with 20 prize levels requiring an attention choice; one of the 20 choices is randomly realized and paid in each round, with no feedback. The ten paid rounds followed a paid comprehension quiz and two practice rounds and with feedback. There is no exogenous error in these environments; participants are informed that with enough time there is always sufficient information to make a correct prediction and therefore there is always a positive gross marginal benefit when choosing  $\pi_L$  over  $\pi_0 - G(\pi_L) = \frac{1}{2}u(P)$ .

In the manual environment only, a participant provides their belief they will make an error in the prediction stage if they are collecting information using  $\pi_L$ , defined as  $q_L \in [0, \frac{1}{2}]$ . The elicitation mechanism uses a binarized scoring rule, where reporting truthfully is a dominant strategy and invariant to risk attitudes (see [Schotter and Trevino, 2014] for a review of lab implemented belief mechanisms, including the binary scoring rule used here and previously by [Karni, 2009] and [Holt and Smith, 2009]).

Let  $I_{PLE}^i \in \{0, 1\}$  be the indicator variable equal to 1 if and only if participant  $i$  chose to pay attention with information structure  $\pi_L$  when the prize was  $P$  in environment  $E \in \{\text{'automatic'}, \text{'manual'}\}$ . I suppress the variables for  $E$  and  $i$  whenever the context is a single decision environment or participant. I refer to a participant's full set of binary information gathering choices as their *attention allocation*.

The data from the first stage choice to gather information is my primary tool for analysis, but participants also provide a prediction  $a \in \{r, b, g\}$  and (implicitly) a decision time  $t$  in each round. The experiment records several other timestamps, including when it was first possible to form an informed prediction, when the player actually clicked to collect prediction-relevant information, and the posterior at the player's decision time  $\gamma_t(R) \in \{0, \frac{1}{2}, 1\}$ .

<sup>11</sup>Formally an information structure maps from a state  $(\omega, t)$  - where  $\omega \in \{R, B\}$ , and  $t$  is the number of i.i.d. draws observed - to a distribution of posteriors, but  $t$  is suppressed when it is not relevant.

## Costly Information Representation

A choice dataset has a *costly information representation* if the DM:

- (i) makes ex ante optimal choices of attention structure given expected costs and benefits:  $(\pi_A \in \arg \max_{\pi \in \Pi} \{G(A, \pi) - K(\pi)\})$ , and
- (ii) their chosen action is optimal given their observed information:  $(\sum_{\omega \in \{R, B\}} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \{R, B\}} \gamma(\omega) u(a'(\omega)) \text{ for all } a' \in \{r, b, g\})$ .

If a DM's choices have a costly information representation, then the decision maker is behaving as if they have a known attention cost function and optimally choose information and actions given this function. Those with costly information representations have their choices modeled precisely with the two-stage limited attention models.

The primary theoretical contribution of [Caplin and Dean, 2015] is their derivation of two testable conditions that are necessary and sufficient for a state dependent stochastic choice dataset to have a costly information representation. I am interested in whether experimental data has a costly information representation to estimate the extent to which we can use limited attention models to classify mistakes. If a choice dataset is indistinguishable from randomly generated data, classifying a choice as OI is logically tenuous. But if a dataset has a costly information representation, classifying a choice as OI is reasonable, and further we can interpret mistakes that are not OI using the language of the model.

Experimental data differs from the theorists' state dependent stochastic choice dataset in meaningful ways. The utility function  $u()$  is not known or directly observed, and the limited number of observations for a given decision problem makes for coarse observation of the choice probabilities in a given decision problem. However, the experimental data is richer than a state dependent stochastic choice dataset in some dimensions because I can directly observe the choice of information structure  $\pi$  in each decision problem, as well as the exact information received. These differences in observables, require the costly information representation axioms to be modified for this context. I modify the conditions of [Caplin and Dean, 2015] into the Optimal Action ( $Oa$ ) and Optimal Information ( $O\pi$ ) conditions below to account for these differences between theoretical and experimental data.

Optimal Action ( $Oa$ ) is a property of each decision problem, and requires that the observed actions chosen are optimal given the posterior beliefs of the DM. In this decision environment, there is a unique optimal action for any given posterior. It is necessary that:

$$\begin{aligned}
 a = g &\iff \gamma(R) = \frac{1}{2} & (Oa) \\
 a = r &\iff \gamma(R) = 1 \\
 a = b &\iff \gamma(R) = 0
 \end{aligned}$$

If condition  $(Oa)$  is not met, a participant has demonstrated ineffective attention. The condition  $(Oa)$  is this experiment's analogue to [Caplin and Dean, 2015]'s NIAS condition, defined in the appendix.

Optimal information  $(O\pi)$  is a property of the entire choice data set generated by a DM and requires that there could be no gross payoff improvement by reassigning the chosen information structures across decision problems. An example of a violation of  $(O\pi)$  can be generated with a DM who faces an identical decision problem twice with the only difference being the prizes,  $P < P^*$ . If she used an information structure (say  $\pi_L$ ) that improved her probability of receiving  $P$  by  $\frac{1}{2}$ , but chose to be uninformed (using  $\pi_0$ ) when  $\pi_L$  was available and the gross benefit was  $\frac{1}{2}P^*$ , then her dataset fails  $(O\pi)$ . The condition  $(O\pi)$  is this experiment's analogue [Caplin and Dean, 2015]'s NIAC condition, defined in the appendix.

In these binary choice environments, the optimal information condition  $(O\pi)$  can be checked through a series of pairs of decisions. Each pair of decisions that have the same prize  $P$  or sequence length  $L$  offers an opportunity for an information allocation mistake. Formally:

$$I_{PLE} \leq I_{P^*LE} \quad \forall P, P^* \in \mathbb{P} \quad s.t. \quad P < P^*; \quad \forall L \in \mathbb{L} \quad (O\pi 1)$$

$$I_{PLE} \geq I_{PL^*E} \quad \forall L, L^* \in \mathbb{L} \quad s.t. \quad L < L^*; \quad \forall P \in \mathbb{P} \quad (O\pi 2)$$

To satisfy  $(O\pi 1)$ , if a participant gathered information when the sequence length was  $L$  for prize  $P$  in environment  $E \in \{\text{auto}, \text{manual}\}$ , they must also gather information when the sequence length is  $L$  for any higher prizes in that environment. To satisfy  $(O\pi 2)$ , if a participant gathered information when the sequence length was  $L$  for prize  $P$  in environment  $E$ , then they must also gather information when the prize is  $P$  in rounds with shorter sequence lengths  $L < L^*$ , because they have a lower expected time cost.

A dataset has a costly information representation if it satisfies all the conditions jointly -  $(Oa)$ ,  $(O\pi 1)$ , and  $(O\pi 2)$ .

## Time cost

I interpret a participant's reaction to attention costs under an assumption that expected attention cost  $K(\pi)$  can be separated into time cost  $T(\pi)$  and cognitive cost  $C(\pi)$ , so that  $K(\pi) = T(\pi) + C(\pi)$ . The automatic environment suppresses the cognitive requirement as much as possible, a participant needs only to click a button every 10 seconds and the computer will eventually make a correct prediction and win a prize with certainty. I normalize whatever cognitive requirement remains in the automatic environment to  $C(\pi) = 0$ , and interpret the reaction to expected time across rounds.

In the automatic environment, for any prize  $P$  and sequence length  $L$ , a DM with a costly information representation will choose to guess ( $\pi = \pi_0$  and  $a = g$ ) when:

$$\frac{1}{2}u(P) - K(\pi_L) < 0 \quad (1.1)$$

Likewise, a DM with a costly information representation will choose to make an informed prediction ( $\pi = \pi_L$  and  $a \in \{r, b\}$ ) when:

$$\frac{1}{2}u(P) - K(\pi_L) > 0 \quad (1.2)$$

The switching point from inattention to gathering information as  $P$  increases or  $L$  decreases provide bounds on time cost  $T(\pi)$  in the automatic environment, and overall attention cost  $K(\pi)$  in the manual environment. The difference between the attention costs and time cost yields an implied cognitive cost; when this cognitive cost is negative I interpret it as a preference for choice agency despite the added possibility of mistake.<sup>12</sup> For presentation purposes these utility valued bounds are presented as dollar values, under the assumption that the utility function is linear over these small prize values.

One of the primary benefits of choosing action  $g$  is that no time cost is incurred. However, if this experiment were to be conducted over a fixed time in a lab, then the relative benefit of saving time is very small if the participant must wait until all others are finished to receive payment. For this reason, the experiment is available online to participants over a three day period through a unique link. By doing this, it is expected that the participant will choose to complete the experiment at a time when their opportunity cost is low relative to other times, but the marginal time cost of the information gathering task in this remote setup is still greater than what it would be if they were committed to sit in a lab for an hour.

## 1.3 Results

### Implementation

In April-May 2021, participants were recruited through the SFU Experimental Economics Lab portal to participate in an online experiment to be completed online at home over three days, beginning with a 30 minute live video conference introduction. The introduction period included informed consent and a live reading of the instructions with opportunities for questions. The attendance requirement then ended and participants were told they could complete the experiment at any time of their choosing.

<sup>12</sup>Cognitive costs are presented on an adjusted basis, discounting the gross payoffs  $\frac{1}{2}u(P)$  using the elicited mistake probability  $q$ , but because the average observed cognitive costs are near zero the difference in results between the belief-adjusted and naive interpretations is small.

104 participants completed the online introduction and received the show up fee, 101 participants completed the experiment.<sup>13</sup> Of these 101 participants, 37 proceeded through rounds in which time costs were randomly ordered, and the remaining 64 faced time costs that increased by round. There were no significant order effects between the two groups so they are pooled throughout the paper. Participants earned an average variable pay of \$19.91 while spending 34 minutes in the information gathering process. A participant who spends no time gathering information is expected to earn variable pay of \$6.70 through correct guesses. So participants earn around \$23.31 per hour gathering information.<sup>14</sup>

A show up fee of \$7 and comprehension quiz pay averaging \$2.30 brought average payoffs above \$29 with an average time online of over one hour in total.<sup>15</sup>

I coded the experiment in python using oTree [Chen et al., 2016] as an underlying structure to capture data and distribute unique participant links for online participation, with a custom javascript interface for the interactive information gathering stage.

**RESULT 1: Participants behave as if they have a known attention cost function. The median participant dataset requires fewer than 2% of choices dropped to be consistent with the axioms, and 41% of participant datasets satisfy all axiomatic tests.**

If participants in this experiment did not create state-dependent stochastic choice datasets that can be rationalized by a costly information representation, then the data interpretation offered in section 1.2 is tenuous. So the first question to answer is whether the model is suitable for this data, and this is effectively a test of the theory and whether the experiment incentives worked as designed.

To satisfy  $(Oa)$  a participant must only choose  $a = g$  until the moment the state is revealed, and after that time, must only make a correct prediction. To satisfy  $(O\pi1)$  a participant facing sequence length  $L$  in environment  $E$  who chooses to gather information when the prize is  $P$  must also gather information for any greater prize values. To satisfy  $(O\pi2)$  a participant facing prize  $P$  in environment  $E$  who chooses to gather information when the sequence length is  $L$  must also gather information for any shorter sequence lengths when the prize is  $P$  in environment  $E$ .

## Optimal Information

If participants make optimal decisions in a costly information representation, the optimal information conditions  $(O\pi1)$  and  $(O\pi2)$  are satisfied for all possible comparisons in the

<sup>13</sup>Three participants opted out during the informed consent and instructions.

<sup>14</sup>Marginal earnings of  $(\$19.91 - \$6.70) = \$13.21$  per 34 minutes = \$23.31 per hour.

<sup>15</sup>Some players take long breaks between rounds so the most consistent measure of active time is the information gathering time even though it is a lower bound.

dataset. In this unlikely case, I could not reject the hypothesis of a costly information representation.

Under a null hypothesis that participants choose the binary  $I_{PL}$  randomly and independently with probability  $\frac{1}{2}$ , a failure of  $(O\pi 1)$  will be generated with probability  $\frac{1}{4}$ , i.e., when  $I_{PL} = 1$  and  $I_{P*L} = 0$ .

There are  $\frac{|\mathbb{P}|^2 - |\mathbb{P}|}{2}$  opportunities to fail equation  $(O\pi 1)$  for any  $L$ , and  $\frac{|\mathbb{L}|^2 - |\mathbb{L}|}{2}$  opportunities to fail equation  $(O\pi 2)$  for any  $P$ . With  $|\mathbb{P}| = 20$ ,  $|\mathbb{L}| = 5$ , and two decision environments (automatic or manual), This leaves 1900 opportunities to fail equation  $(O\pi 1)$  and 400 opportunities to fail equation  $(O\pi 2)$  per participant.<sup>16</sup>

Participants overall fail 1.3% of tests of  $(O\pi 1)$ , and 71% of participants fail zero of the 1900 conditions  $(O\pi 1)$  places on the data. Condition  $(O\pi 1)$  requires a single switch in first stage choices from not gathering information to gathering information as  $P$  increases *within* a single decision page, a relatively low bar. Participants overall fail 7.3% of tests of  $(O\pi 2)$  which requires a single switch from gathering information to not gathering information as  $L$  increases *across* decision pages, and 58% of participants fail zero of the 400 conditions  $(O\pi 2)$  places on the data.

Only three of 101 participants fail to reject a hypothesis of random binomial data generation in favour of a one-tailed alternative hypothesis of fewer rejections, implying that 98 of 101 participants generate choice datasets which are significantly closer to a costly information representation than randomly generated data.<sup>17</sup>

I construct the predictive success measure inspired by Selten, treating each participant's dataset as an observation. This measure takes the proportion of participants whose choice data perfectly satisfy  $(O\pi 1)$  and  $(O\pi 2)$  and subtracts the probability a randomly generated dataset would satisfy the same conditions [Demuyneck and Hjertstrand, 2019]. Overall, 41% of participants satisfy all 2300 information conditions, versus a near zero probability of a random dataset satisfying those conditions. The Selten score of 41% being positive suggests a costly information model has predictive success for the first stage choice data.

## Optimal Action

The experiment records the exact time when there is sufficient information to make a correct prediction  $\bar{t}_i$ , as well as the participant's decision time  $t^i$  and action  $a_i$ . The random information gathering process is independent across participants, so  $\bar{t}_i$  varies by participant.

A costly information representation requires that any action observed is optimal given the posterior at the time of action. A participant who chooses an action at a time  $t < \bar{t}_i$  must choose  $a_i = g$  because it yields a prize with probability 51%, versus 50% probability

<sup>16</sup>(190 conditions per sequence length  $L * 5$  values of  $L * 2$  environments; 10 conditions per price  $P * 20$  prices \* 2 environments).

<sup>17</sup>The one-tailed test is conducted at a 0.025 significance level.



when choosing  $a_i = r$  or  $a_i = b$ . For any time  $t > \bar{t}_i$ , the posterior is either  $\gamma(R) = 1$  or  $\gamma(R) = 0$ , and for these cases  $(Oa)$  requires  $a_i = r$  and  $a_i = b$ , respectively. An  $(Oa)$  violation occurs before the state is revealed if a participant chooses any to make any specific prediction, and an  $(Oa)$  violation occurs after the state is revealed if a participant guesses or makes an incorrect prediction.

There are two information states that could be used to form a posterior, one that uses all data drawn (the true information state), and one that only uses the observations the player clicked to store in a data table. There is only one marginal violation in the entire sample using the stricter true information state, suggesting participants are unlikely to be using their memory as a replacement for the observation table.

Regardless of whether an action is chosen before or after the state has been revealed, only one of the three available actions  $\{r, b, g\}$  is optimal, so a uniform random dataset would fail  $\frac{2}{3}$  or an average of 6.67 of the 10 observed paid actions. The median participant failed zero of 10  $(Oa)$  conditions and the mean number of conditions failed was 0.77.

61% of participants satisfy all  $(Oa)$  conditions. The probability of one participant's dataset satisfying all  $(Oa)$  conditions if action choice is uniform random is  $(\frac{1}{3})^{10} \approx 0.24\%$  of participants pass all  $(O\pi)$  conditions and  $(Oa)$  conditions. Thus the overall Selten score is 24%.

## Proportion of Data that is Consistent with Axioms

There is substantial literature on measuring the variation in observed choice datasets from an axiomatic standard (see [Demuyne and Hjertstrand, 2019] for a review of measures and recent advances in their computation). One of the common measures of a dataset is the Houtman-Maks Index (HMI) which measures the maximal proportion of choices that can collectively satisfy every condition of  $(O\pi 1)$  and  $(O\pi 2)$  (or analogously drops the fewest observations to have a consistent dataset).<sup>18</sup> I calculate an HMI specific to the information gathering choices, and another using all choices which includes the 10 conditions on paid actions imposed by  $(Oa)$   $(Oa)$ . The median HMI(information) among participants is 0.99 and median HMI(all) is 0.98; this high degree of consistency suggests that I may interpret those few observations that fail one of the costly information axioms as mistakes. In a sample of uniform randomly generated datasets, the median HMI(information) is 0.63 and median HMI(all) is 0.61.

<sup>18</sup>Tests of GARP often prefer the money-metric critical cost efficiency index (CCEI) for goodness of fit because it measures the dollar cost of errors, but that is less appropriate in this context because the budget constraint is known in the GARP context and the attention cost function is not in this context.

After computing the HMI for each participant’s information gathering choices, there is a set of observations that, if omitted, leaves a dataset that is fully consistent with  $(O\pi)$ .<sup>19</sup> If a dropped observation was a choice to gather information –  $I_P L = 1$  – it is recorded as suboptimal attention (SA), and if a dropped observation was a choice to be inattentive –  $I_P L = 0$  – it is recorded as suboptimal inattention (SI). Some participants have multiple subsets of the same size that satisfy the costly information representation axioms, one subset that involves dropping a 0 and one that involves dropping a 1. These ambiguous cases are not included in the counts of IA and SI. The top row of Figure 1.3 shows that of the actions categorized as mistakes in the 10 paid rounds, 22 of 331 can be described as errors at the first stage. These cases saw a participant make a decision deemed “suboptimal” relative to the preponderance of their other information gathering decisions; of these 15 were cases of SI and 7 were cases of SA. Figure 1.4 shows there were 8 cases of SI – 5% of mistakes – and 2 cases of SA – preceding 1% of mistakes – in the five manual rounds.

I prefer to use data from the five manual rounds when discussing the proportion of mistakes because the five automatic rounds restrict the participant from incorrectly choosing  $a = b$  or  $a = r$ , and thus eliminates some types of mistakes. Only one type of ineffective attention is possible in the automatic environment – a participant can fail to click a relevant observation in the 10-second window, then stop gathering information and guess before the observation is drawn again.<sup>20</sup> This is distinct from dynamic inconsistency, where a participant starts the information gathering process but stops and guesses before the relevant observation is ever drawn.

**RESULT 2: The majority of mistakes are caused by optimal inattention, but ineffective attention and dynamic inconsistency are significant sources of mistakes which occur after the choice to gather information.**

This experiment is designed to monitor the information state and the implied posterior at all times, and these data allow me to differentiate between causes of mistakes. A costly information representation can rationalize many mistakes as optimal ex ante, but it can not rationalize dynamic inconsistency and ineffective attention, which occur after the choice to gather information.

If information were costless every participant would be predicted to gather information and make a perfect prediction in every round. So a mistake is any guess or any incorrect prediction, as either fails to maximize the gross payoff. When attention is costly however, a large number of ex ante choices to guess are in fact OI, as they optimally trade off the costs and benefits of attention.

<sup>19</sup>This omitted set is one interpretation of the set of attention allocation violations the participant made ex ante, taking the maximal consistent dataset as the true preference.

<sup>20</sup>This type of fail-to-click ineffective attention occurred 5 times in the 5 automatic rounds.

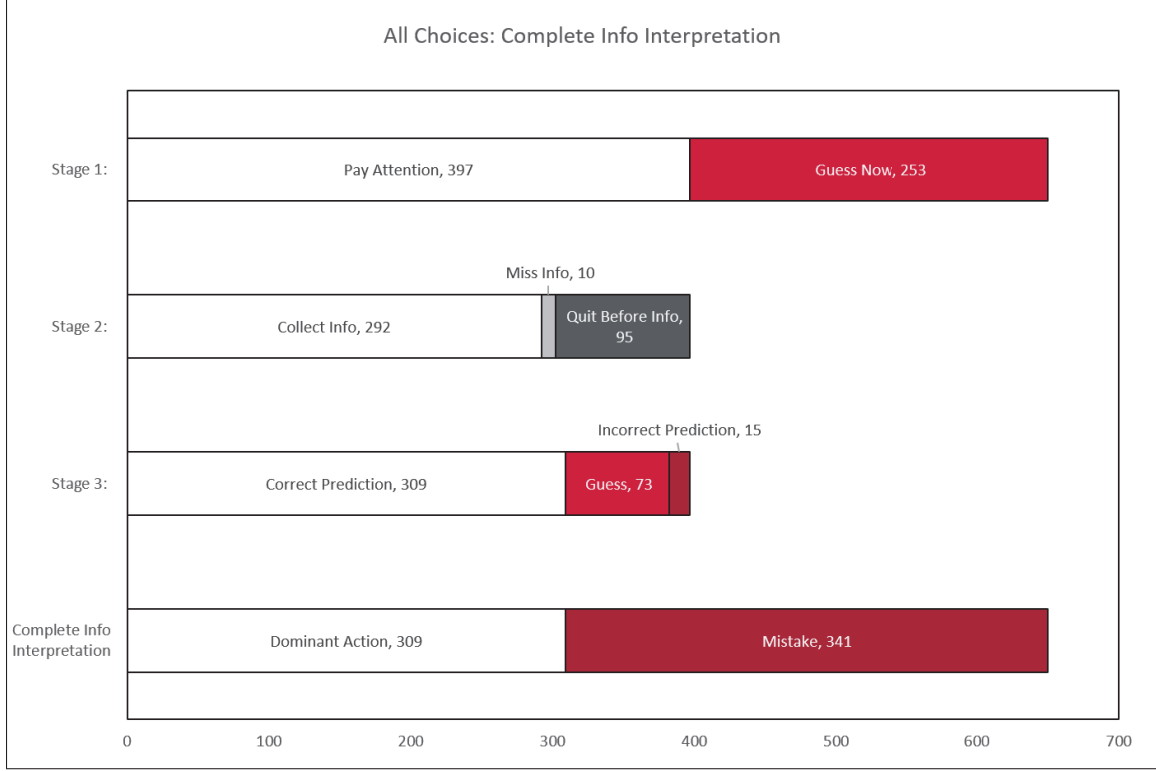


Figure 1.2: All Observed Choices by Stage  
101 participants over 10 paid rounds

If a mistake is not an ( $Oa$ ) violation, then the participant's action was the best choice given the information state, and this implies the participant chose to guess at a time when the posterior was uninformative. If the relevant first stage choice for that round was  $I_{PL} = 0$  and this observation is not among the subset dropped in the process of estimating the HMI, it is classified as being caused by OI. However observations of  $I_{PL} = 0$  that needed to be dropped to form a costly information representation are interpreted as suboptimal inattention (SI), as the majority of similar choices imply the participant's benefits of gathering information exceeded their costs. Likewise, If the relevant first stage choice for that round was  $I_{PL} = 1$ , the participant's choice can be classified as suboptimal attention or optimal attention based on whether it was included in the largest dataset fully consistent with ( $O\pi$ ).

If a mistake is also an ( $Oa$ ) violation, it implies the action was not optimal given the information state. A mistake where the participant failed to interpret the posterior into the optimal action is classified as ineffective attention (NA). Ineffective attention implies the information state was fully revealed and the participant made an incorrect prediction or guessed, or made a specific prediction without information that would change their prior.

The remaining type of mistake has a participant choose to gather information, but then choose to guess in the second stage before any information is revealed. This is a form of dynamic inconsistency but is not an ( $Oa$ ) error. These participants' previous selves wanted

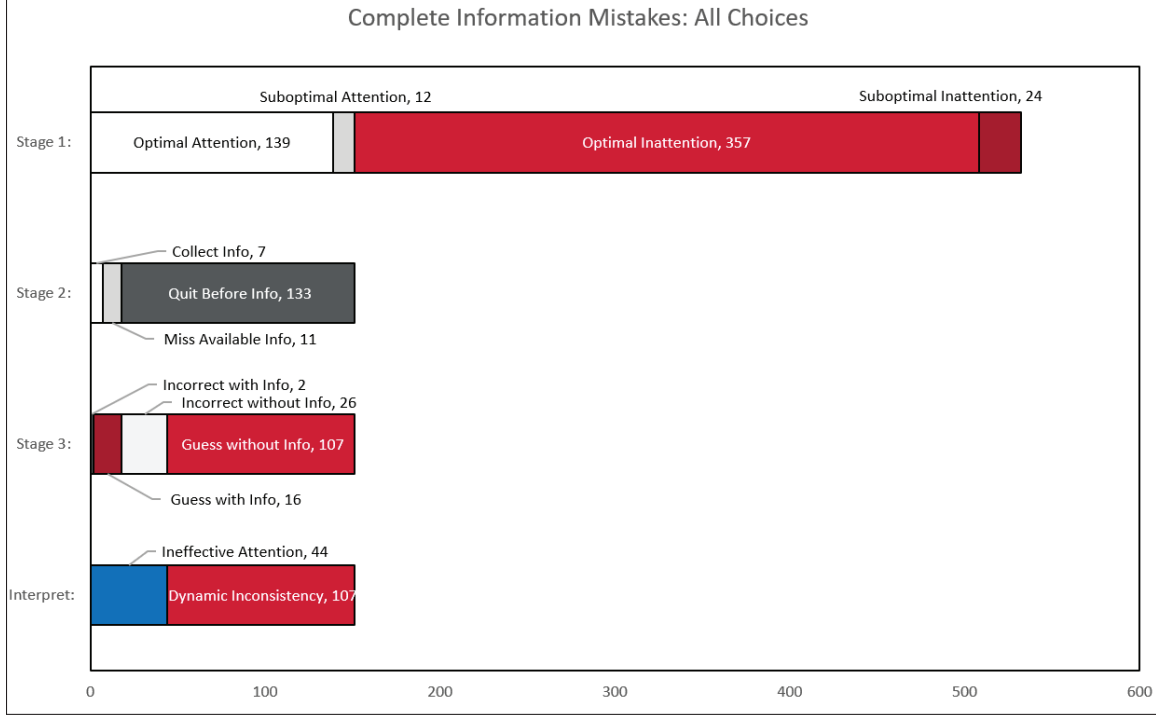


Figure 1.3: Classifying Mistakes: All Choices

101 participants make 331 choices a full information observer would deem a mistake. Note that ineffective attention is impossible in the Automatic Environment.

to gather information, but when they reached the information gathering they quit early. This experiment is not sufficient to test whether the participant faced an unanticipated shock to time costs during the second stage, or if they were previously naive about their future self, or if there was some other cause of their dynamic inconsistency. But it is a distinct cause of mistakes worth separating from ineffective attention.

Figure 1.4 shows that 132 of 175 mistakes (75%) in the manual environment can be explained as OI. 12% of these errors are observed after a participant chooses to gather information but then fails to interpret the given information (NA). Another 8% are driven by participants who choose to gather information in the first stage, but then choose to guess after incurring some time cost and no marginal information in the second stage (dynamic inconsistency). The remaining 5% of mistakes are driven by suboptimal inattention in the first stage, where a participant made several choices to gather information for lower prizes and one difficult-to-explain choice to guess for a larger prize.

**RESULT 3: Participants make low-cost mistakes; observed mistakes are only 30% as costly as mistakes in randomly generated data.** While any non-optimal action is an ( $Oa$ ) violation, their expected costs take three different values, measured as a change in prize probability. If a participant chooses  $a = b$  or  $a = r$  when the posterior  $\gamma(R) = \frac{1}{2}$ , her probability of winning a prize decreases from 51% to 50%, for a loss of

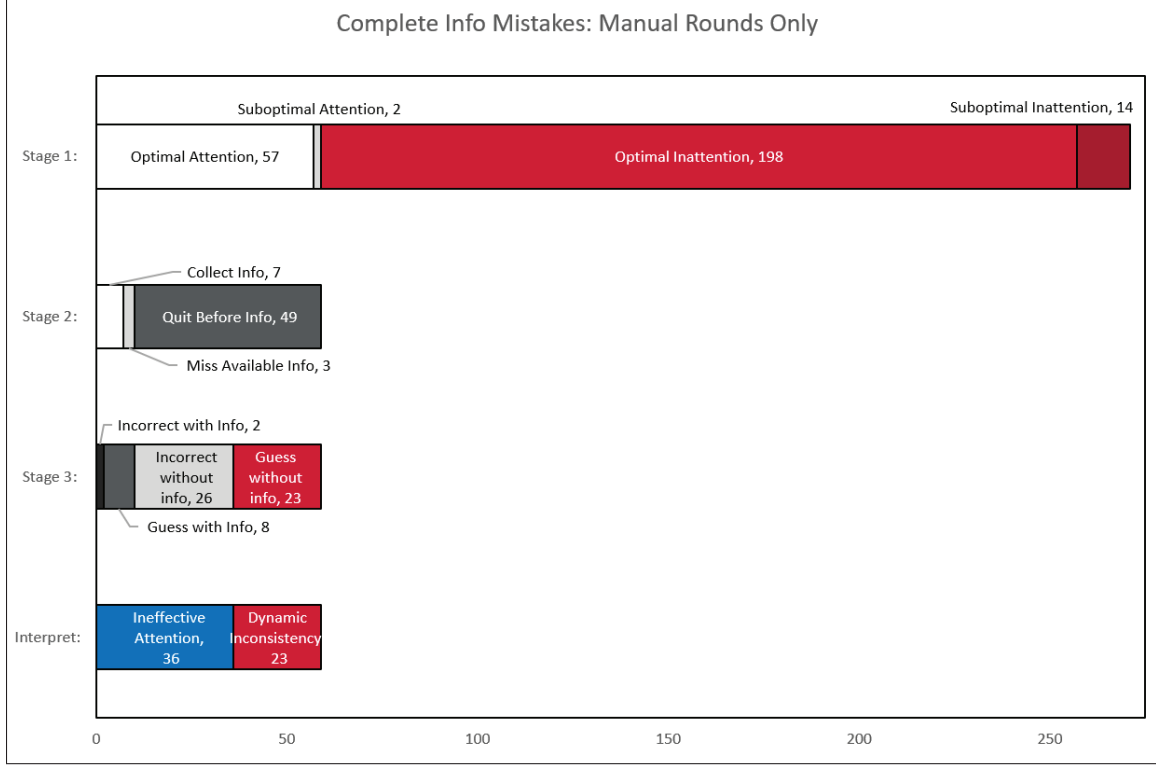


Figure 1.4: Classifying Mistakes: Manual Choice Environment.  
101 participants over 5 paid Manual rounds.

1pp. If a participant chooses  $a = g$  when  $\gamma(R) = 1$ , her probability of winning a prize decreases from 100% to 51%, for a loss of 49pp. Finally, if a participant chooses  $a = r$  when  $\gamma(R) = 0$  or  $a = b$  when  $\gamma(R) = 1$ , her probability decreases from 100% to 0%, for a loss of 100pp. Experiment participants make less costly errors than a set of errors generated by a random dataset. 75% of participant ( $Oa$ ) violations are the 1pp variety, 22% are the 49pp variety, and 3% are 100pp mistakes. The mean participant ( $Oa$ ) violation costs 15pp, versus an average cost of 50pp from a randomly generated dataset. Assuming linear prize utility, this implies that the average participant violation of ( $Oa$ ) cost a minimum of \$0.39 in expectation while a randomly generated violation costs more than three times as much, with a minimum of \$1.31.<sup>21</sup>

It is difficult to assign a monetary value to the cost of a mistake caused by dynamic inconsistency, because rather than failing to maximize the probability of a prize, a participant failed to minimize their attention costs for a given prize probability. These participants enter

<sup>21</sup>This minimum is based on the unconditional expectation of prize value, but a player in the attention stage knows that none of the prize values for which they chose to guess were chosen for payment, so the expected prize conditional on reaching the attention stage is higher for participants who pay less attention at lower prize values.

the information gathering stage and incur at least 10 seconds of time cost, but ultimately choose to guess, which they could have done in the first stage with no time cost incurred.

**RESULT 4: The majority of attention cost is time cost, and estimated time costs in the first stage are negatively correlated with time spent in the second stage. Average cognitive costs are near zero. Belief-adjusted cognitive costs are negative, implying a preference to manually make a choice.**

I perform within-subject analysis to estimate cognitive costs by comparing analogous choices in the manual and automatic environments. The manual environment requires data interpretation before action selection and so plausibly affects cognitive costs, while the automatic environment only incurs a time cost as there is no potential for making a mistake in action selection.

There are four possible values that the pair  $(I_{pl}^i(auto), I_{pl}^i(manual))$  can take, each providing different information on the cognitive and time costs of the participant:

1.  $(I_{pl}(auto), I_{pl}(manual)) = (0, 0)$  is the “no attention” case with implied high attention costs, implying  $T(\pi_l) \geq \frac{1}{2}u(p)$  and  $C(\pi_l) \geq \frac{1}{2}u(p) - T(\pi_l)$ .
2.  $(I_{pl}(auto), I_{pl}(manual)) = (1, 1)$  is the “always attention” case with implied low attention costs, implying  $T(\pi_l) \leq \frac{1}{2}u(p)$  and  $C(\pi_l) \leq \frac{1}{2}u(p) - T(\pi_l)$ .
3.  $(I_{pl}(auto), I_{pl}(manual)) = (1, 0)$  is the “doesn’t like puzzles” case where the participant will let the computer interpret data but will not do it manually, implying low time costs relative to cognitive costs:  $T(\pi_l) \leq \frac{1}{2}u(p)$  and  $C(\pi_l) \geq 0$ .
4.  $(I_{pl}(auto), I_{pl}(manual)) = (0, 1)$  is the “prefers puzzles” case where the participant will not let the computer interpret data but will do it manually, implying a negative cognitive cost:  $T(\pi_l) \geq \frac{1}{2}u(p)$  and  $C(\pi_l) \leq 0$ .

A costly information representation implies a cutoff strategy over prizes for any given sequence length and environment; the cutoff prize value where a participant begins to gather information can be estimated by the sum of all first stage (ex ante) information choices for a given round’s attention cost  $L$ . This proxy maintains every participant’s intended probability of reaching the information gathering stage without selectively dropping any observation, and identifies the cutoff perfectly for those participant rounds when no observations fail ( $O\pi$ ). If  $\sum_p I_{pl} = n$ , this proxy is interpreted as the participant being willing to gather information for the  $n$  largest prizes and not willing for the  $(20 - n)$  smallest prizes. This allows for upper and lower bound identification of attention costs. In utility terms, it implies  $\frac{1}{2}u(p_{20-n}) \leq T(\pi_4) \leq \frac{1}{2}u(p_{21-n})$ . Assuming  $u(P) = P$  over the prize values, I can estimate a dollar value of time cost from a participant’s first stage choices in the automatic

environment, and by using the expected time in minutes, calculate the implied reservation wage.<sup>22</sup>

Estimated time costs from ex ante first stage decisions are negatively correlated with the ex post time spent gathering information conditional on reaching the information gathering stage ( $\rho = -0.45$ ,  $p = 0.000003$ ). This is further evidence that participants do in fact respond to unobserved personal time costs in a way that aligns with my interpretation of the data.

Figure 1.5 displays the aggregated estimates of attention and time cost. Average time cost is positive and increasing in expected time. Average cognitive cost is near zero, implying indifference between making ones own prediction in the manual environment relative to letting the computer predict in the automatic environment. Participants pay attention equally often in manual and automatic rounds, implying time is the primary driver of choice. However, there is substantial heterogeneity in cognitive cost, with many participants displaying preference for manual predictions over automatic, a negative-valued cognitive cost. A negative-valued cognitive cost may be interpreted as a preference for choice agency despite the weakly greater probability of mistake.

In the manual environment, participants report a belief of their own error rate in an incentive compatible mechanism. This means that the gross benefit of paying attention in the manual environment is weakly less than in the automatic environment because there is a possibility of making a mistake, and this reduces the prize probability conditional on choosing to pay attention. The naive and adjusted cognitive costs are both displayed in Figure 1.5. Cognitive costs are small in all cases, zero in the naive case and slightly negative in the belief-adjusted case. This implies that participants prefer the manual rounds after adjusting the gross benefit to account for the probability that they could make a mistake; this could be interpreted a preference for choice agency or a preference to complete the puzzle over mindlessly clicking while waiting for a computer to do so. An equally plausible explanation for this pattern of choices is that participants are naive to inclusion of their mistake probability in their first stage attention choice and indifferent between the two environments. In either case, it is the time cost rather than the cognitive portion of attention costs acting as the key driver of behaviour in this environment.

## 1.4 Related Literature

A number of models beginning with [Sims, 2003] focus on the optimal choice of information for a consumer with a bounded ability to obtain information. Sims' decision maker makes an information choice modeled as a reduction in Shannon entropy, and [Matejka and McKay,

<sup>22</sup>Linearity over small prizes is not a substantially stronger assumption than a general expected utility representation [Rabin, 2000].

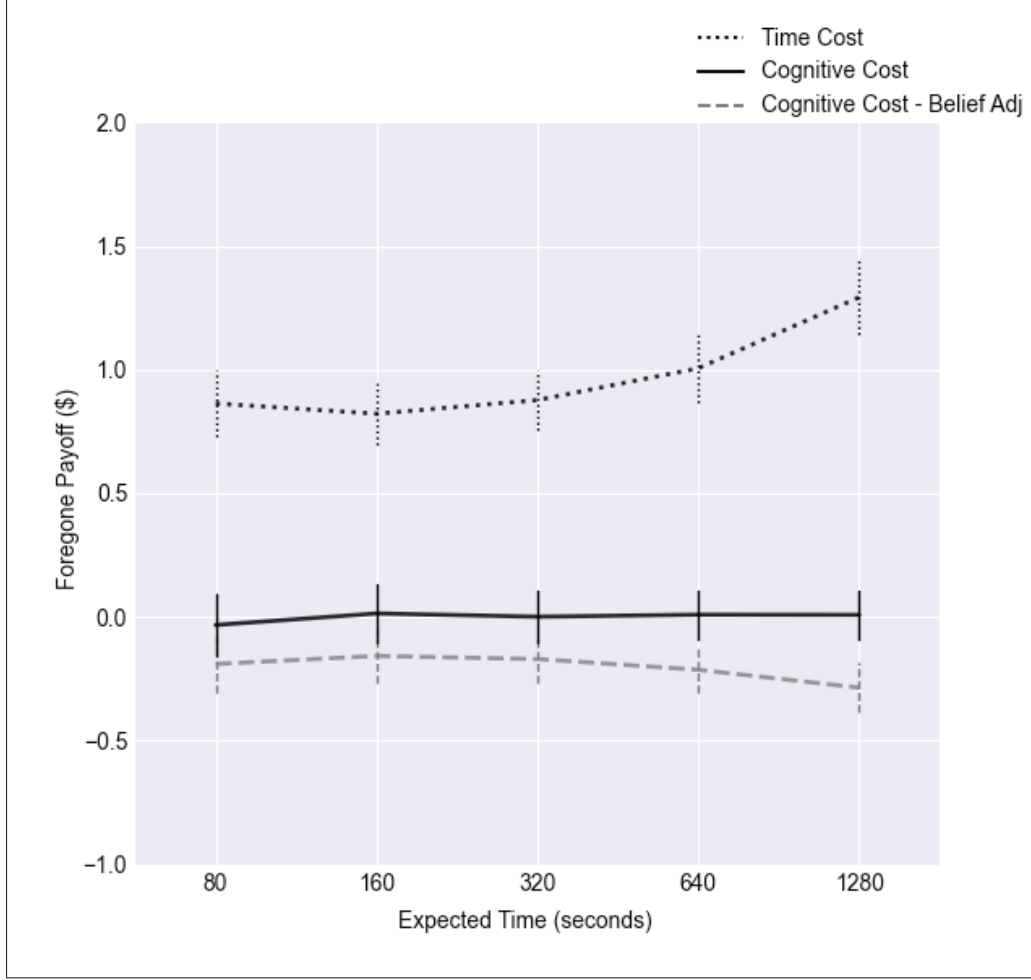


Figure 1.5: Implied Average Time Costs and Cognitive Costs  
*The implied time cost from choices is increasing in the expected attention time by round. Cognitive costs are not a key driver of choices. 10% confidence intervals.*

2014] use this setup to derive new foundations for the multinomial logit as a model of discrete choice probabilities.<sup>23</sup> Modeling information as Shannon entropy however faces the drawback of using a single choice parameter, and more recent theoretical contributions consider broader types of attention choices including random attention [Aguiar et al., 2021, Cattaneo et al., 2020], consideration sets [Manzini and Mariotti, 2007, Manzini and Mariotti, 2014], and sparse attention [Gabaix, 2014, Enke, 2017, Caplin et al., 2018]. I am agnostic to the specific formulation of attention costs and so I build my testable conditions using the approach of [Caplin and Dean, 2015], whose general attention cost function  $K()$  can incorporate all of the above formulations of attention cost.

<sup>23</sup>This contrasts [McFadden, 1973] who had originally founded the multinomial logit model as resulting from IID unobserved errors on consumer utility.



There is a growing literature that uses non-choice data to learn about consumer decision processes, including [Fenig et al., 2018] which is focused on identifying consumer beliefs. [Bartoš et al., 2016] conduct a resume study that directly measured attention through computer clicks and other non-choice data in job and housing markets. Like these authors, I rely on button clicks in the second stage to directly measure attention, which is an inexpensive alternative to eye-tracking and other costly physiological measures of attention.

This experiment was inspired by the approach of [Dean and Neligh, 2019] who also asked participants to discern a binary state using red and blue shapes. This experiment expands on their design; Dean and Neligh infer the intensity of a participant’s attention based on the accuracy of the final action choice, while I instead directly measure the intent to pay attention in the first stage and the information collected through the button clicks in the second stage. The separation of the first two stages allows me to identify dynamic inconsistency as a source of mistakes, and unlike Dean and Neligh I can distinguish between inattention and ineffective attention.

The experiment by [Martínez-Marquina et al., 2018] uses novel treatment variation to separate out two drivers of consumer choice to be inattentive in uncertain situations. They show that complexity (cognitive costs) and certainty are two distinct drivers of a consumer’s choice to be inattentive, my experiment environment is simpler and finds weak effects of cognitive costs. [Dean and Martin, 2016] measure rationality with the minimum cost of violations, I similarly bound attention costs assuming rationality of decision makers and find time cost accounts for the majority of variation in participant choices to gather information.

## 1.5 Discussion

I introduce an experimental methodology that enables isolation of several confounding explanations for mistakes by a consumer who has the choice to trade off attention costs for information that enables a better choice. When all uncertainty about the costs and benefits of attention are removed – the automatic rounds – a participant’s ex ante choices to gather information provide bounds on their personal time costs. The manual rounds introduce a potential for ineffective attention by forcing the participant to interpret information and make an action choice; these choices provide information to bound the participant’s subjective cognitive effort. These bounds allow for the 75% mistakes caused by optimal inattention to be distinguished from those that are caused by ineffective attention (12% of mistakes in this experiment) or suboptimal inattention (5% of mistakes). Dynamic inconsistency causes the remaining 8% of mistakes, when a participant said gathering information was worth the cost ex ante, but then quit after incurring some time cost and before receiving any information. The data supports the notion that many or even most mistakes can be explained as a result of optimal inattention, but even in this simple environment a substantial number of

mistakes are driven by errors at the information gathering and processing stages. Given the simplicity of the experiment, the observed rates of dynamic inconsistency and ineffective attention might be interpreted as lower bounds on these effects in more complicated contexts. Researchers looking to explain or improve choices in domains with substantial information requirements should consider dynamic inconsistency and ineffective attention as potential causes of sub-optimal choice.

## Chapter 2

# Early Completion of a Real Effort Task

David Freeman and Kevin Laughren

### 2.1 Introduction

Many economic decisions involve a trade-off between current and future benefits and costs: how much to consume versus save, to exercise or not, and whether to complete an onerous task today or delay. In such situations a person makes choices at different points of time and cannot commit their future choices. However, most experiments that study intertemporal decision-making study choices over delayed monetary rewards made at a single point in time [Coller and Williams, 1999, Harrison et al., 2002, Andreoni and Sprenger, 2012]. Such experiments are by design uninformative of three crucial facets of intertemporal decisions.

The first standard axiom of intertemporal choice is time consistency: if a person previously chose between two options, then at any point before the options entail different consequences, the person would not wish to revise their initial choice. The second standard axiom is time invariance: if a person would choose an option over another today, they would make the same choice between these options tomorrow if the consequences of each action were shifted a day in the future. A person whose choices are time-inconsistent should not expect their own future choices to be consistent with their current desires, and thus faces a difficult problem of forecasting their own future choices. This introduces a third facet of intertemporal decision-making. Sophistication is the normative axiom for time-inconsistent decision-making that posits that a person correctly forecasts their future choices. With limited exceptions, little existing experimental work addresses the latter two aspects of intertemporal decision-making.

We introduce a new experimental design that studies participants' task-completion decisions for real-effort tasks. A participant faces multiple different choice sets, which we call effort schedules, that list two or three days and a number of chores associated with each

day. For each effort schedule that includes the current day, the participant must indicate their choice to either do the task today or not. If they chose to do the task “today” in the effort schedule that is randomly selected to be played out for real, then they must complete the number of real-effort chores that schedule specifies by the end of today. Otherwise, the next day they face all schedules for which they previously selected “not today” and those for which “today” was not available. Crucially, their initial choices cannot commit their future choices except by completing the task or forgoing an option by selecting “not today”. Thus, we can elicit each participant’s preference to complete or not complete the task each day for each effort schedule that the participant faces before they complete the task. We vary the effort required on each available day across schedules as well as the days available to complete the task, which allows us to identify violations of time consistency, sophistication about time inconsistency, and time invariance.

We identify time inconsistency in an effort schedule  $\{(e_1, 1), (e_2, 2), (e_3, 3)\}$  in which  $e_t$  tasks are required at time  $t$  by looking at the associated two-option schedules of all pairwise combinations of these options. We call the combination of all four two- and three date effort schedules derived from the set  $\{(e_1, 1), (e_2, 2), (e_3, 3)\}$  a “quad”. If in period 1, the participant selects “today” in one such two-option schedule but not the other, we infer their  $t = 1$  preference between  $(e_2, 2)$  and  $(e_3, 3)$  – and the participant is revealed time inconsistent if their  $t = 2$  choice for the schedule  $\{(e_2, 2), (e_3, 3)\}$  is inconsistent with that  $t = 1$  preference. In such a case of time inconsistency, the participant’s initial choice for  $\{(e_1, 1), (e_2, 2), (e_3, 3)\}$  reveals their beliefs about how they will choose at  $t = 2$ ; a participant reveals themselves as sophisticated if their  $t = 1$  today-or-not choice is that same as their choice between  $(e_1, 1)$  and what they would actually choose at  $t = 2$ , and otherwise revealing themselves naive [Freeman, 2021]. We also test time invariance by comparing choice in two effort schedules, where effort values are shifted one day forward in one schedule.

Our participants exhibit a high degree of time consistency in spite of the considerable power to detect violations with the experimental design. We find that 50 of 82 participants are time consistent in every quad, and overall, choices in 84% of all participant-quads are time consistent. We find that this time consistency arises from a strong tendency to choose to complete a task immediately, even when delaying would have reduced the number of required chores. We find that 29 of the time consistent participants always choose “today” on day 1, when available, and 70% of all two-date choices are “today” – including half of pairwise choices in which delaying reduces the number of chores that must be completed. In pairwise choices, a participant’s beliefs about their own future behavior are trivial – and thus the immediate completion tendency we document for real effort tasks does not arise from sophistication about inconsistent preferences.

These findings are broadly inconsistent with the dominant intuition that people tend to prefer to delay unpleasant tasks [O’Donoghue and Rabin, 1999] and thus contradicts the common modelling assumption that bads like aversive and effortful chores entail negative

flow payoffs, but are discounted if they occur in the future.<sup>1</sup> Our findings thus pose a challenge to the literature on intertemporal choice. Our experiment was designed to be essentially similar to existing real-effort experiments that study the preferences governing intertemporal trade-offs [Augenblick et al., 2015, Augenblick and Rabin, 2019] in all respects that behavioral economic theory deems relevant. Our novel design features are primarily intended to distinguish sophistication from naivete about present bias – and should have no bearing on the presence of present bias. Yet not only do we not find the choice patterns that similar past work attributes to present bias – we find an opposing pattern of behavior. This stark finding stands apart from most of the experimental economics literature (although we note some exceptions, including [Aycinena et al., 2020]). However, a recent literature in psychology finds evidence of “precrastination”: in experimental bucket-carrying tasks, participants exhibit a tendency to start the task too soon even when it increases the total amount of effort required [Rosenbaum et al., 2014]. Rosenbaum et al. attribute this to a desire to “reduce working memory loads”. [Haushofer, 2015] formalizes this idea through a “cost of keeping track” and shows it can explain diverse evidence on intertemporal choice through a completely different mechanism than present-biased discounting. Our findings suggest this may be a more important determinant of intertemporal choice in standard economic decisions than previously recognized.

## Related Literature

There is a long experimental literature on intertemporal choice that studies preferences over delayed monetary rewards revealed at one point in time [Coller and Williams, 1999, Harrison et al., 2002, Halevy, 2015]. Since money can be saved and borrowed, in principle, such experiments should not reveal intertemporal preferences if participants broadly bracket their experimental choices with opportunities outside of the lab [Cubitt and Read, 2007]. Thus, some intertemporal choice experiments use less-fungible rewards like snacks that will be consumed immediately [Read and Van Leeuwen, 1998] and real-effort tasks [Augenblick and Rabin, 2019]. Our design uses a real-effort task from [Augenblick et al., 2015].

Much of the literature on intertemporal choice only studies choices made at one point in time and thus cannot directly test time consistency or time invariance – with some exceptions. Read and van Leeuwen (1998) have their participants choose a post-lunch snack both a week in advance, and then ask them to choose again the day they consume the snacks – and find that participants choose unhealthy snacks more frequently in same-day choices than in-advance choices. [Augenblick et al., 2015] study participants’ allocations of real-effort chores between earlier and later dates made both when the earlier date is in the future and when it is in the present – and find that participants tend to delay effortful chores, exhibiting present bias. Relatedly, [Augenblick and Rabin, 2019] elicit participants’

<sup>1</sup>We both tested out our real effort tasks and found them extremely aversive.

preferences over quantities of delayed real-effort chores at different points in time and their beliefs about their future such preferences. These two real-effort experiments both find that participants tend to prefer to delay effortful chores, and exhibit “present bias” by exhibiting a disproportionate preference to delay immediate effort. [Halevy, 2015] uses a design in which participants report their preferences between smaller-sooner versus larger-later monetary payments in successive weeks and tests time invariance and time consistency of preferences for delayed monetary rewards. He finds that over half of participants are time consistent, and roughly half of all participants satisfy time invariance. Compared to Halevy, we study choices involving real-effort tasks – for which previous work has demonstrated higher rates of present bias than for delayed monetary payments [Augenblick et al., 2015].

Existing papers test for a person’s sophistication or naivete about their own time inconsistency by measuring demand for commitment devices – [Ashraf et al., 2006, Augenblick et al., 2015], for a review see [Bryan et al., 2010] – or comparing elicited beliefs to actual future choices [Augenblick and Rabin, 2019]. In contrast, our design elicits choices at different points of time in an environment in which a participant cannot commit their future choices, which allows us to employ the approach of [Freeman, 2021] to test both sophistication and naivete about time inconsistency for each participant. Much of our design is motivated by a literature that commonly finds evidence of time-inconsistent present bias and that has found widely different degrees of sophistication about it [Ashraf et al., 2006, Augenblick et al., 2015, Augenblick and Rabin, 2019]. Yet unlike this literature, we find little evidence of time-inconsistency – and thus have little to say about sophistication and naivete.

After finding that participants frequently prefer to complete the task today even when delaying would reduce the number of chores required, we discovered a conceptual precedent for our findings in the psychology literature [Rosenbaum et al., 2014, Rosenbaum et al., 2019, Wasserman, 2019]. Rosenbaum et al. find that when faced with the choice of two buckets to carry to a fixed endpoint, participants tend to pick up whichever is closer to them – even if it entails a longer carrying distance. [Fournier et al., 2019b] find evidence that suggests this effect arises from a desire to start a task sooner, rather than a desire to complete a subgoal sooner. [Fournier et al., 2019a] show these effects are mediated by working memory – suggesting that precrastination may at least in part arise from a desire to reduce working memory loads. Within economics, existing work has shown that formal models that incorporate memory can explain facets of intertemporal choice that are difficult to explain as aspects of intertemporal preferences alone [Ericson, 2017, Haushofer, 2015].

## 2.2 Theoretical Framework

We design our experiment to study how participants’ decisions of whether to complete a real-effort task depend on future options to complete it. To that end, consider a participant who must complete a real-effort task exactly once. When first confronted with the task,

they are informed of the two or three different days on which they can do the task and how much effort they must exert to complete it on each day. On each day until the last day with an option available, they can either complete the task or not, but they cannot commit their future behaviour except by completing the task. Each such *effort schedule* can be represented as a double or triple of effort-date pairs of the form  $\{(e_t, t), (e_{t+k}, t+k)\}$  or  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$ .

When making a today or not-today decision at time  $t$  when facing the effort schedule  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$ , the participant must consider their intertemporal preferences that they apply to make trade-offs between earlier versus later effort. They must also consider how they would behave in the future should they wait today, since they cannot commit their future behavior.

[Freeman, 2021] shows that if a participant is time inconsistent, their choices in  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$  can reveal their sophistication or naivete about their time inconsistency if their actions change when removing one option.

Specifically, if they delay at  $t$  when facing  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$  and instead completes the task at  $t+k$  (for  $k=1$  or  $=2$ ), but completes it at  $t$  when facing  $\{(e_t, t), (e_{t+k}, t+k)\}$ , then the participant exhibits what Freeman calls a *doing-it-later reversal*, and reveals themselves naive. This is because their choice to do it at  $t$  when facing  $\{(e_t, t), (e_{t+k}, t+k)\}$  reveals that at  $t$ , they prefer completing it at  $t$ ,  $(e_t, t)$ , to waiting until  $t+k$ ,  $(e_{t+k}, t+k)$ . Thus, they would only initially delay when facing  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$  and then complete it at  $t+k$  if they were to incorrectly believe that they would instead complete it on the other date after  $t$  (not  $t+k$ ).

If instead they complete it at  $t$  when facing  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$  but also would delay at  $t$  when facing  $\{(e_t, t), (e_{t+k}, t+k)\}$ , they exhibit a *doing-it-earlier reversal* and reveal themselves sophisticated. This is because their choice to delay at  $t$  when facing  $\{(e_t, t), (e_{t+k}, t+k)\}$  reveals that at  $t$ , they would rather wait to do it at  $t+k$ . Thus they would only do it at  $t$  when facing  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$  if they expect that they would not do it at  $t+k$  were they to wait. That is, doing it at  $t$  reveals that they expect future time inconsistency.

Fixing the triple  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$ , we wish to observe how the participant would choose from each possible pair as well as when all three days are available. We refer to each such quadruple of choice observations as a *quad*. To observe whether a participant would exhibit a possible doing-it-earlier or a doing-it-later reversal involving  $\{(e_t, t), (e_{t+1}, t+1), (e_{t+2}, t+2)\}$ , we need to also observe how they would behave in two of the doubleton subsets in which  $(e_t, t)$  is available. By also observing a participant's choice from  $\{(e_{t+1}, t+1), (e_{t+2}, t+2)\}$ , our experiment can reveal their time inconsistency directly – they reveals themselves *time inconsistent* if either (i)  $(e_{t+1}, t+1) = c(\{(e_t, t), (e_{t+1}, t+1)\})$ ,  $(e_t, t) = c(\{(e_t, t), (e_{t+2}, t+2)\})$ , but  $(e_{t+2}, t+2) = c(\{(e_{t+1}, t+1), (e_{t+2}, t+2)\})$ , or (ii)

$(e_t, t) = c(\{(e_t, t), (e_{t+1}, t+1)\})$ ,  $(e_{t+2}, t+2) = c(\{(e_t, t), (e_{t+2}, t+2)\})$ , but  $(e_{t+1}, t+1) = c(\{(e_{t+1}, t+1), (e_{t+2}, t+2)\})$ .

Standard models of intertemporal choice also restrict how choices vary across quads. Economic models of intertemporal choice assume that choices are *monotonic* in effort: if  $e'_t \leq e_t$ ,  $e'_{t+k} \geq e_{t+k}$ , and  $(e_t, t) = c(\{(e_t, t), (e_{t+k}, t+k)\})$ , then  $(e_t, t) = c(\{(e'_t, t), (e'_{t+k}, t+k)\})$ . Economic models also typically make a less intuitively fundamental assumption about preferences for modeling convenience, known as *time invariance* [Halevy, 2015]: if  $e_t = e_{t+l}$  and  $e_{t+k} = e_{t+k+l}$ , then  $(e_t, t) = c(\{(e_t, t), (e_{t+k}, t+k)\})$  if and only if  $(e_{t+l}, t+l) = c(\{(e_{t+l}, t+l), (e_{t+k+l}, t+k+l)\})$ . Our design allows us to test both of these assumptions.

## 2.3 Experiment Design

We design a real-effort experiment to obtain data on participants' task-completion decisions. After an initial orientation session, our experiment is four days long and took place Monday-Thursday ( $t = 1, 2, 3, 4$  respectively). On each day of the experiment, a participant was required to login to our online interface, make relevant decisions, and complete a single unit of our real-effort task, which we refer to as a *chore*. Each chore requires the participant to transcribe 20 blurry Greek characters by clicking them from their screen; this real effort task was adapted from [Augenblick et al., 2015]. On exactly one of the days, the participant was required to complete  $e_t$  *extra chores*, with the number-date pair  $(e_t, t)$  determined by the participant's decisions. All participants who completed the entire experiment were paid \$25 by electronic transfer on Sunday.

In order to observe a participant's decisions in multiple different effort schedules, we employ a variation on the random incentive system to provide incentives to report their true preferences of whether to work on not on each day. On the first day of the experiment, the participant is presented with all effort schedules for the experiment and is informed that one of these has been randomly chosen and will be implemented – the *schedule that counts*. The participant then makes a “today” or “not today” choice for every effort schedule they face in which a  $t = 1$  option is available. If they chose “today” in the schedule that counts, then they complete the required number of extra chores today. Otherwise, when they log in at  $t = 2$ , they face a “today” or “not today” decision for those effort schedules with two or more options remaining and for which they did not choose “today” at  $t = 1$ ; similarly at  $t = 3$ .

The first two versions of the experiment used quads designed to have power to detect present bias, and thus sophistication and naivete about said bias. For our third version, we also included control quads designed to allow us test whether some participants exhibit a negative discount rate by choosing to exert more effort and complete the task at an earlier date. In all cases, we study quads with  $t \in \{1, 2, 3\}$  and with  $t \in \{2, 3, 4\}$ , where



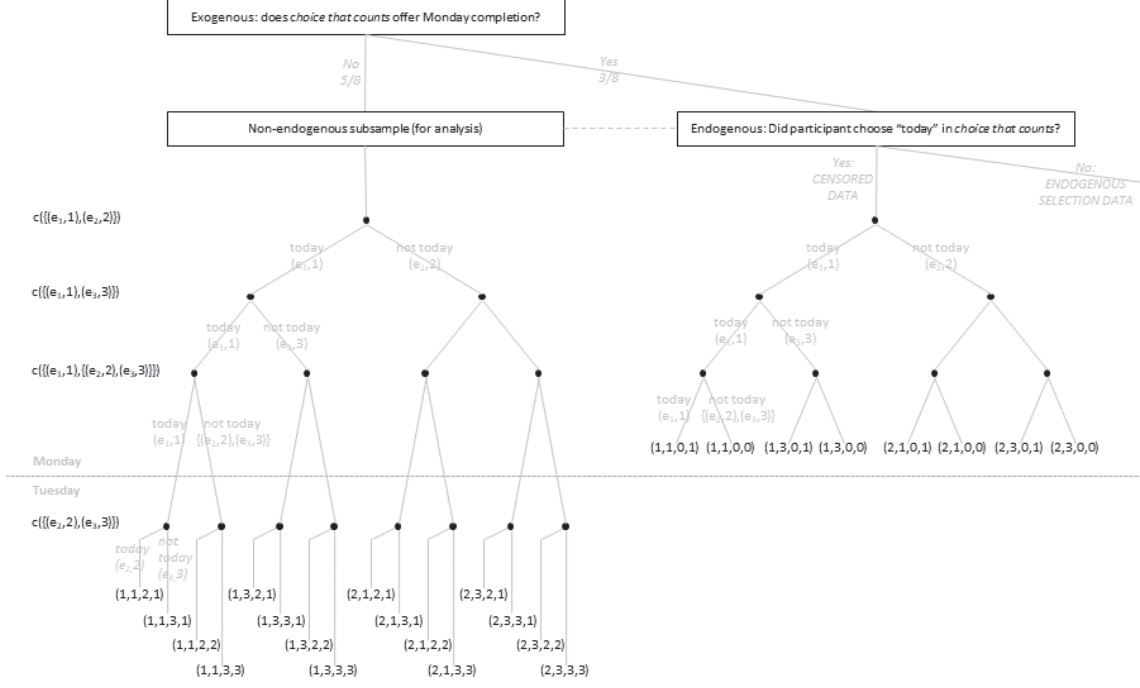


Figure 2.1: Experiment data generating process

The four integers in a quad represent chosen completion dates from the three binary choices in an effort schedule, followed by the triplet.

$$c(\{(e_1, 1), (e_2, 2)\}, c(\{(e_1, 1), (e_3, 3)\}), c(\{(e_2, 2), (e_3, 3)\}), c(\{(e_1, 1), \{(e_2, 2), (e_3, 3)\}\}))$$

the latter effort schedules are obtained by shifting the former by one day.<sup>2</sup> Shifting quads also enables us to test time invariance. We show the effort schedules underlying the choices faced by our experiment participants in Table 2.1. Each participant makes up to 8 choices per effort schedule: one quad with opportunities to work on  $t \in \{1, 2, 3\}$  and one quad with  $t \in \{2, 3, 4\}$ . Some quads are partially censored when participants complete their extra chores before  $t = 3$ .

Figure 2.1 demonstrates the experiment data generating process for a quad with  $t \in \{1, 2, 3\}$ . Note that there are 24 terminal nodes representing a possible quad observation. The probability that a quad is censored is endogenous to the number of times a participant chooses to work “today” on Monday across the entire experiment, and this in turn affects the probability of observing any other quads. Appendix B.1 includes the fully extended decision tree, where the left sub-tree is repeated for the endogenous sample on the right

<sup>2</sup>The presence of the latter quads, combined with the random incentive system and  $\{(e_2, 2), (e_3, 3)\}$  choice sets, implies that we have at least a  $\frac{5}{8}$  chance of observing  $t = 2$  choices.

## Implementation

We recruited participants from the SFU Experimental Economics Laboratory Research Participation System. Experiments were conducted entirely online, with a live introductory Zoom session followed by four days of asynchronous participation at any time of the participants’ choosing. Participants were paid a \$7 (CAD) show up fee for the introductory session and an all-or-nothing completion payment of \$25, both paid by email transfer on the Sunday following the final experiment deadline (Thursday at 11:59pm). The completion payment requires the participant to sign in and complete at least one chore on four consecutive days, complete extra chores on one of those days, and make task-completion decisions when available. Participants were recruited to attend a 30 minute introductory session, held on Fridays via Zoom. Informed consent was collected, instructions were read aloud by an experimenter, and questions were taken via confidential chat. After questions were answered, participants were asked to demonstrate that they were able to sign-in to the online experiment interface and complete one chore, with technical support provided by the experimenter until all participants were successful. Participants were sent a reminder email on Monday morning with a link to the experiment. Participants were required to sign in to the experiment through the University’s centralized authentication service. Upon signing in Monday, participants chose their own email reminder time for the daily reminders for the remainder of the experiment.

101 participants completed the online introduction and received the show up fee. 82 participants completed all of the experiment requirements and received full payment by email transfer. The remaining 19 participants missed a day of participation or chose not to complete their chores on one or more occasion, so they received only the show-up fee and are excluded from analysis. The baseline number of chores (20) and length of chore (40 characters) were chosen so that the session would require participants less than one hour of their time in total over the four days to complete all the chores and make all decisions required for the \$25 variable pay. A 40-character chore requires 40 button clicks with 100% accuracy; the authors required 30-45 seconds to complete each chore on their earliest attempts. It is difficult to estimate an average hourly wage because participants had until 11:59pm each day to complete all decisions and chores and could take breaks before this deadline at no cost. This results in a wide range of observed time spent on extra chores(00:12:37 - 13:38:12), however the median participant completes 20 extra chores in under 25 minutes.

We varied the effort schedules faced across three versions of the experiment. Table 2.1 displays the effort schedules participants faced by version. We conducted Version 1 starting on July 20, 2020 with 23 participants. After observing many “today” choices in Version 1, we added an additional schedule in Version 2 to allow us to detect even small degrees of present bias and conducted that session starting July 27, 2020, with 22 participants. Still

Effort Schedule	Participants Observed	Quads Observed	Versions
14, 20, 28	45	76	V1, V2
16, 20, 25	82	144	All
18, 20, 22	82	144	All
19, 20, 21	22	37	V2
20, 20, 20	82	144	All
22, 20, 18	37	68	V3
25, 20, 16	37	68	V3
	82	681	

Table 2.1: Experiment Effort Schedules

*Each triple describes the number of chores required if working on day 1, day 2, or day 3.*

observing many “today” choices, we changed the set of effort schedules for Version 3 to enable us to detect whether participants would make such choices if doing so increased the number of chores required, which would indicate an opposing preference to those generated by discounting and present bias. We conducted Version 3 with 15 participants starting March 8, 2021 and with 22 participants starting March 29, 2021. In each version, each participant had the opportunity to face each effort schedule twice, once when day 1 was Monday (denoted MTW) and once when day 1 was Tuesday (denoted TWR).

**Data Censoring** We do not always observe two full choice quads from each effort schedule because the day on which a participant completes their extra chores (and thus stops making choices) depends both on their choices and the randomly-selected schedule that counts. When a participant completes their payoff-relevant extra chores on Monday, they no longer make task-completion decisions. In these cases, we obtain no data for TWR effort schedules nor do we obtain Tuesday decisions from MTW effort schedules. This partial censorship also occurs on TWR effort schedules when extra chores are completed Tuesday.

This endogenous censoring is inherent when studying any incentivized ‘when-to-do-it’ choices. However, our design has a  $1/2$  probability that a TWR schedule is the schedule that counts, and a  $1/8$  probability that a MTW schedule with no option to do it on Monday is the schedule that counts. This design results in a  $5/8$  exogenous probability that a participant makes payoff relevant choices on at least two days. Our software randomly assigned 52 of 82 participants such an effort schedule, and this sub-sample is highlighted when discussing results which could be subject to endogeneity. The remaining 30 participants generate data that is subject to endogenous censoring, including 5 participants who (endogenously) generate only censored quads.

## 2.4 Results

We begin by classifying each of the 24 possible observed quad choice profiles in Figure 2.1 into categories. Recall that a quad represents four choices problems from one effort schedule:  $c(\{(e_1, 1), (e_2, 2)\})$ ,  $c(\{(e_1, 1), (e_3, 3)\})$ ,  $c(\{(e_2, 2), (e_3, 3)\})$ , and  $c(\{(e_1, 1), \{(e_2, 2), (e_3, 3)\}\})$ . There are four categories into which a subject’s choice profile in a quad may fall. First, they may demonstrate a consistent preference to complete the task on a certain day, in which case we categorize the quad as time consistent (7 possible choice profiles). Second, a choice profile may meet the definition of a reversal in the theoretical framework, either as a doing-it-earlier reversal (6 choice profiles) or a doing-it-later reversal (2 choice profiles). A third type of choice profile cannot be rationalized by a single  $t = 1$  utility function over when to complete the task, which can be potentially consistent with a preference for commitment or a preference for flexibility; we refer to these as Non-Strotzian (5 choice profiles). The final 4 of 24 choice profiles are censored as there is insufficient observed choice data to reveal a participant’s preferences over that effort schedule. The complete mapping of choice profiles to these five categories is listed in Appendix B.6.

**RESULT 1: Choices in 84% of uncensored quads are time consistent. At an individual level, 50 of 82 participants are time consistent in all of their uncensored quads.**

When all quads are considered regardless of censoring or endogeneity, 500 of 681 observations are Time Consistent. When removing the 84 censored observations, 500 of 597 (84%) uncensored quads are time consistent, 399 (67%) of which exhibit a consistent preference to complete the extra chores on day 1. Table 2.2 provides the classifications by effort schedule, and remarkably over two-thirds (67%) of all quads are time consistent in each row. Among the time inconsistent quads, there are similar quantities of Reversals and Non-Strotzian observations (both in the high single digits). Comparatively if choice data were generated randomly by independently mixing at each choice in Figure 2.1, 36% of uncensored choice profiles would be time consistent, and 39% would exhibit a reversal.

The full set of data in Table 2.2 are subject to endogenous sampling and censoring. Table 2.3 restricts the sample to those participants randomly pre-determined to make at least two choices, and further drops the TWR quads which are subject to endogenous observation of choices at  $t = 3$  (Wednesday). We refer to these data in Table 2.3 as the *non-endogenous subsample*; note these data have zero censored observations by construction, yet still exhibit a very consistent mix of decisions to the “TOTAL excluding Censored” data from Table 2.2.

At an individual level, 30 of 82 participants were randomly assigned a ‘choice that counts’ which allowed them to complete their extra chores on Monday, and this subsample is subject to endogenous selection. On Monday, the participants in this set who chose

ALL Effort Schedules	Time Consistent			Reversal		Non- Strotz	Censored	Quads Observed
	day1	day2	day3	earlier	later			
14, 20, 28	72%	3%	7%	4%	3%	7%	5%	76
16, 20, 25	69%	2%	6%	8%	1%	3%	10%	144
18, 20, 22	65%	4%	5%	8%	1%	7%	11%	144
19, 20, 21	65%	5%	5%	5%	3%	5%	11%	37
20, 20, 20	53%	6%	8%	8%	1%	7%	17%	144
22, 20, 18	37%	7%	24%	7%	0%	9%	16%	68
25, 20, 16	38%	7%	26%	10%	0%	3%	15%	68
TOTAL*	67%	5%	12%	9%	1%	7%		597
TOTAL	59%	5%	10%	8%	1%	6%	12%	681
RANDOM*	14%	11%	11%	28%	11%	25%		
RANDOM	13%	10%	10%	25%	10%	23%	9%	

\*excluding censored

Table 2.2: Classifying Observed Choice Quads by Effort Schedule (All data)

*RANDOM provides the data proportions expected if randomizing independently at each choice in Figure 2.1*

MTW only Effort Schedules	Time Consistent			Reversal		Non- Strotz	Cens.	Quads Observed
	day1	day2	day3	earlier	later			
14, 20, 28	79%	4%	0%	4%	4%	8%	0%	24
16, 20, 25	75%	2%	6%	12%	0%	6%	0%	52
18, 20, 22	73%	4%	4%	10%	2%	8%	0%	52
19, 20, 21	75%	0%	0%	8%	8%	8%	0%	12
20, 20, 20	60%	10%	2%	12%	4%	13%	0%	52
22, 20, 18	36%	14%	29%	4%	0%	18%	0%	28
25, 20, 16	32%	14%	36%	14%	0%	4%	0%	28
TOTAL	63%	7%	10%	10%	2%	9%	0%	248
RANDOM	12.5%	12.5%	12.5%	25%	12.5%	25%	0%	

Table 2.3: Classifying Non-endogenous subsample

*This table only uses the 52 participants whose 'choice that counts' does not have a Monday option, and includes only their MTW quads.*

*RANDOM (and data) in this table condition on reaching the left sub-tree in Figure 2.1*

to work “today” for their ‘choice that counts’ do not make any more decisions at future dates. Over half (17/30) of these participants exclusively generate quads that are time consistent, and another 5 of the 30 only generate censored observations, and thus satisfy time consistency trivially.<sup>3</sup> The remaining 8 of these 30 endogenous participants generated at least one reversal or Non-Strotzian quad of choices.

Within the non-endogenous subsample of 52 participants randomly assigned a ‘choice that counts’ which did not allow them to complete extra chores Monday, 28 are time consistent in 100% of observed choices. 18 of these 28 time consistent participants exclusively generated quads with a time consistent preference to complete the chores on day 1, even on those schedules for which waiting on day 1 would entail completing fewer total chores. Since we observe this subsample make choices on at least two days, all of these tests of time consistency are non-trivial. All remaining tables in the main text of results include only this non-endogenous subsample – though a comparison of Tables 2 and 3 suggests that data censoring does not appear drive our results on time consistency.

Time consistency is a property of quads of choices within a single effort schedule, but an additional consideration is whether a participant’s full set of choices are collectively sensible across effort schedules.

**RESULT 1b: 90% of participants who are time consistent *within* every schedule also demonstrate monotonicity *across* all effort schedules. Overall, fewer than 5% of observations need to be dropped to make every participant consistent with monotonicity.**

Monotonicity links preferences across effort values and requires participants to consistently prefer exerting less effort while controlling for time. A set of choices which violate monotonicity is difficult to rationalize with an economic model, even if those choices are time consistent. We evaluate whether a subject violates monotonicity, considering every binary choice in all different quads in the experiment.

We count the total number of monotonicity violations for each participant, and 58 of 82 participants (71%) demonstrate no violations of monotonicity in their dataset. Of the 50 participants who were time consistent in 100% of their classified quads, only 5 made a choice violating monotonicity, thus 45 of 82 participants were perfectly time consistent and monotonic.

For those participants who do violate monotonicity, we use the Houtman-Maks Index (HMI) to represent the maximal proportion of data which can be collectively consistent

<sup>3</sup>For example, suppose we observe Monday choices of “today” for  $\{(e_1, 1), (e_2, 2)\}$ , “not today” for  $\{(e_1, 1), (e_3, 3)\}$ , and “not today” for  $\{(e_1, 1), (e_2, 2), (e_3, 3)\}$ . If  $\{(e_1, 1), (e_2, 2)\}$  is the ‘choice that counts’, the the participant completes extra chores Monday so  $c(\{(e_1, 1), (e_2, 2), (e_3, 3)\})$  and  $c(\{(e_2, 2), (e_3, 3)\})$  are never revealed, thus the quad is censored. See Appendix B.6 for a complete categorization of possible quads, including those that may be censored.

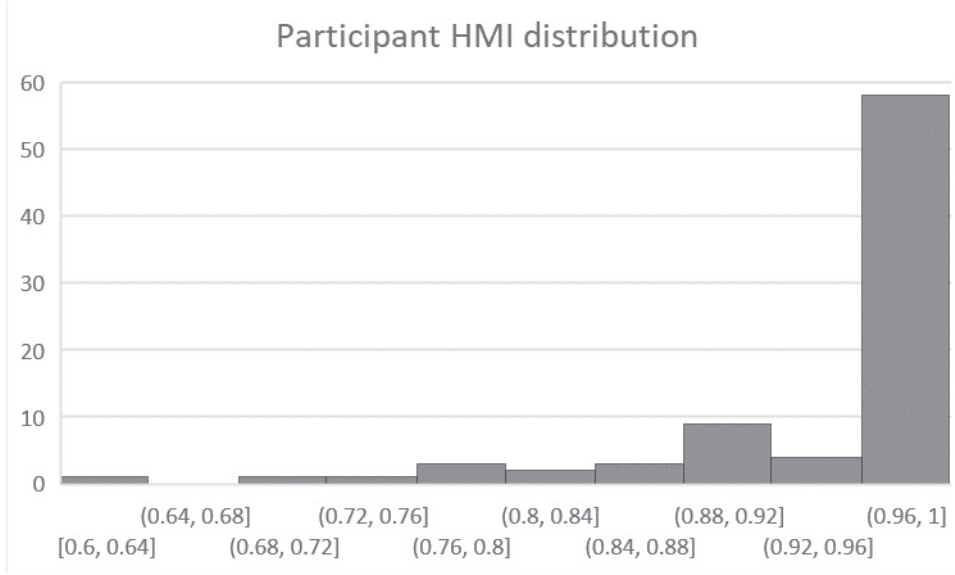


Figure 2.2: Participant Houtman Maks Index - Monotonicity

with monotonicity [Houtman and Maks, 1985, Heufer and Hjertstrand, 2015, Demuynck and Hjertstrand, 2019]. This involves a linear optimization for each participant, minimizing the number of observations removed subject to the constraint that there are zero monotonicity violations in the remaining dataset. In total, 76 of 1726 observations are removed for a weighted mean HMI of 0.955, and the mean HMI among those with at least one violation is 0.86. The distribution of HMI by participant in Figure 2.2 further demonstrates that monotonicity violations are rare and concentrated in a minority of participants.

**RESULT 2: Preferences show a strong immediate completion tendency, with a preference to complete the extra chores on day 1.**

This pattern is evident whether day 1 is  $t = 1$  (in Table 2.3) or day 1 is  $t = 2$  (in Appendix B.2), as both show over 60% of quads are time consistent with a preference for day 1. This suggests the pattern is not simply a day of the week preference. Even for the decreasing effort schedules (22, 20, 18) and (25, 20, 16) the consistent preference to work today is the modal quad choice profile, even though waiting would require less total work for the same payment. Table 2.4 shows that over 75% of the individual binary choices are choices to work “today”, including approximately half of binary choices from the decreasing effort schedules (22, 20, 18) and (25, 20, 16). Given the median participant required 1.2 minutes per chore, this implies a willingness to exert an extra 6 minutes of effort to complete the extra chores early. The median participant completes 24 total chores (20 extra chores, plus a minimum of one each day) implying this preference results in a 20% increase in effort for fixed pay. Remarkably, 18 of 28 participants in the non-endogenous subsample who were

Proportion choosing to work Today			
Effort schedule	$c(\{(e_1, 1), (e_2, 2)\})$	$c(\{(e_1, 1), (e_3, 3)\})$	$c(\{(e_2, 2), (e_3, 3)\})$
14, 20, 28	96%	88%	88%
16, 20, 25	90%	85%	90%
18, 20, 22	81%	88%	88%
19, 20, 21	92%	92%	83%
20, 20, 20	77%	77%	85%
22, 20, 18	57%	50%	61%
25, 20, 16	46%	43%	57%
TOTAL	77%	76%	81%

Table 2.4: Proportion choosing to work Today when a choice is available  
*Non-endogenous subsample*

time consistent in every effort schedule (and 34 of 50 overall participants who were time consistent in every schedule) chose to work on day 1 in every choice day 1 was available.

When the same effort tradeoff is observed on different days, a participant who makes a different today/not today choice has violated time invariance. A violation of time invariance could suggest an unobserved preference to complete the task on a specific day.

**RESULT 3: Violations of time invariance occur in 21% of comparable decision pairs, the majority of violations are consistent with a plan to do chores on a favoured day of the week.**

For two binary choices with fixed values of  $(e_t, e_{t+k}, k)$ , time invariance requires that the choice of “today” or “not today” is identical for  $t = 1$  and  $t = 2$ . Thus time invariance requires us to compare participant choices in MTW quads to their analogous choice in TWR quads. Restricting attention to the non-endogenous subsample of participants who were pre-determined to make choices twice, there are 744 total possible tests of time invariance.<sup>4</sup> The choices between day 2 and 3 –  $c(\{(e_2, 2), (e_3, 3)\})$  – are excluded because observing them for TWR quads requires observing choices on Wednesday, which is subject to endogenous selection. Time invariance is satisfied in 79% of tests. Individual participant rates of time invariance are positively correlated with rates of time consistency ( $\rho = 0.52$ ,  $p < 0.01$ ), even though these two axioms are logically independent [Halevy, 2015].

Table 2.5 displays the proportion of pairwise choices that satisfy time invariance when we observe a participant make choices over the same effort schedule on two different days. Violations of time invariance of the (“not today”, “today”) form are plausibly consistent with a favoured day of the week to complete extra chores that overrides differences in chore

<sup>4</sup>Three tests per effort schedule



First Day Choice, Second Day Choice			
Choice set	Time Invariant	Today, Not Today	Not Today, Today
$c(\{(e_1, 1), (e_2, 2)\})$	79%	8%	14%
$c(\{(e_1, 1), (e_3, 3)\})$	78%	8%	13%
$c(\{(e_1, 1), \{(e_2, 2), (e_3, 3)\}\})$	81%	7%	11%
TOTAL	79%	8%	13%

Table 2.5: Time Invariance Violations by Choice Set  
*Non-endogenous subsample*

quantities, and these make up the majority of violations (95 of 153 violations, 13% of tests).<sup>5</sup> Violations of the form (“today”, “not today”) are plausibly consistent with a least favoured day of the week.<sup>6</sup> Together, the relative scarcity of violations of time invariance suggest specific day of the week preferences are driving choice in at most 21% of tests, and thus are unlikely to be a substantial driver of the observed immediate completion tendency discussed above. The number of chores and interest rate on effort do not appear to systematically affect the rate of failure of time invariance across schedules (Appendix B.1), nor do they appear to systematically affect the rate of time consistency within schedules (Table 2.3).

The small response to a negative interest rate on effort apparent in Table 3 indicates that choices are not well represented by a standard model of intertemporal preferences in which participants discount costly future relative to immediate effort. We next conduct a structural estimation to facilitate a comparison of behaviour in our experiment to existing work.

**RESULT 4: Structural estimation of a model of quasi-hyperbolic discounting yields  $\hat{\beta} > 1$ , capturing a strong tendency to complete tasks immediately.**

We model the probability of choosing “today” as resulting from a latent utility model. Consider only the binary decisions  $(e_t, e_{t+k})$ , with two observable actions at time  $t$ , “today” ( $Y_t = 1$ ), or “not today” ( $Y_t = 0$ ), and suppose that  $Y_t = 1 \iff Y_t^* \geq 0$ , where  $Y_t^*$  represents the time- $t$  (unobserved) utility difference between choosing “today” and “not today”. We specify a structural quasi-hyperbolic discounting model with a linear disutility-of-effort:

$$\begin{aligned}
Y_t^* &= U_t(Y_t = 1, e_t, e_{t+k}) - U(Y_t = 0, e_t, e_{t+k}) \\
U(Y_t = 1, e_t, e_{t+k}) &= -\lambda e_t \\
U(Y_t = 0, e_t, e_{t+k}) &= -\beta \delta^k \lambda e_{t+k}
\end{aligned}$$

<sup>5</sup>For example, a participant who wants to complete their extra chores on Tuesday would choose “not today” for the MTW choices and “today” for the TWR choices.

<sup>6</sup>For example, a participant who wants to complete their extra chores on any day except Tuesday.

The net utility of working today can be written as :

$$Y_t^* = -\lambda e_t + \beta \delta \lambda e_{t+k} \mathbb{I}_{\{k=1\}} + \beta \delta^2 \lambda e_{t+k} \mathbb{I}_{\{k=2\}}$$

With a corresponding regression equation:

$$Y_t = b_0 e_t - b_1 e_{t+k} \mathbb{I}_{\{k=1\}} + b_2 e_{t+k} \mathbb{I}_{\{k=2\}} + \epsilon_t$$

We assume there is some variation in individual values of  $Y_t^*$  due to individual preference shocks, and estimate the regression equation above using a binary logit with no intercept. We recover estimates of  $(b_0, b_1, b_2)$  and use them to estimate  $\hat{\beta} = \frac{(b_1)^2}{b_0 b_2}$ ;  $\hat{\delta} = \frac{b_2}{b_1}$ ; and  $\hat{\lambda} = -b_0$ .<sup>7</sup> We cluster standard errors by participant, and recover asymptotic standard errors for the parameter estimates using the delta method. Parameter estimates and their asymptotic standard errors are in Table 2.6, the underlying logit regression estimates of  $(b_0, b_1, b_2)$  and delta method derivation are in Appendix B.2.

Parameter	Estimate (Std. Error)	Confidence Intervals ( $\alpha = 0.05$ )	
		Lower Bound	Upper Bound
Present Bias $\beta$	1.53 (0.22)	1.11	1.96
Discount Factor $\delta$	0.93 (0.050)	0.83	1.03
Utility of Effort $\lambda$	0.14 (0.042)	0.06	0.22
Observations	743		
Clusters	52		

Table 2.6: Structural Estimates of Time Preference Parameters  $\beta, \delta$

*Non-endogenous subsample Binary choices only. Standard errors of logit regression are clustered by individual participant. Asymptotic standard errors estimated using the delta method (derivation in Appendix B.2)*

Previous studies of intertemporal preference consistently estimate values of  $\beta < 1$ , with the interpretation being that there is additional (non-geometric) discounting of all future

<sup>7</sup>We caution against overly interpreting our point estimates. The parameter  $\lambda$  can be viewed as controlling sensitivity to effort or alternatively the degree of stochasticity, and our estimation cannot separate these interpretations. Similarly, notice that  $-e_t > -\beta \delta^{t+k} e_{t+k}$  if and only if  $-e_t^\gamma > -\beta^\gamma \delta^{\gamma t + \gamma k} e_{t+k}^\gamma$  for every  $\gamma > 0$ . Our design does not identify the curvature-of-disutility-of-effort parameter  $\gamma$ , so our point estimates of  $\delta$  and  $\beta$  cannot be directly compared to those in existing work that does attempt to identify  $\gamma$  (e.g. [Augenblick et al., 2015]). However, this does not affect our tests of the null hypotheses (i)  $\beta = 1$  and (ii)  $\delta = 1$ .

periods relative to the present [Augenblick et al., 2015]. Participant choices in this experiment had a clear disposition to complete the extra chores today, and this is reflected in the estimate of  $\beta > 1$ , as caring more about future utility than today’s utility would result in the observed participant preference to complete the task today (with 580/743 binary choices resulting in the choice to work “today”). The utility of effort has the expected (negative) sign, and  $\delta$ , which is identified from the difference in choices when delay is  $k = 2$  days versus  $k = 1$ , does not appear to be a significant driver of choice.

## 2.5 Discussion

Our novel experimental design was ideally-suited to measure a person’s sophistication or naivete about their own time inconsistency. Yet we discovered far more time consistency than we expected based on prior work from economics experiments [Thaler, 1981, Read and Van Leeuwen, 1998, Ashraf et al., 2006], including other experiments that use designs with similar real-effort tasks [Augenblick et al., 2015, Augenblick and Rabin, 2019]. This appears to be driven by a strong tendency to complete tasks immediately – even when this requires additional effort. In our structural model, this leads us to estimate that our subjects tend to have future-biased preferences. While some previous studies have found evidence of future bias [Takeuchi, 2011], it is the opposite of the present bias found in much prior experimental work and commonly assumed in applications of quasi-hyperbolic discounting [Laibson, 1997, O’Donoghue and Rabin, 1999]. Taken together with prior work, our findings suggest that quasi-hyperbolic discounting with present bias may not be a good descriptive model of task completion. Recent work in psychology that uses a different research design has also found that people tend to exhibit a preference to start tasks sooner [Rosenbaum et al., 2014]. Our findings may result from a behavioural bias distinct from present bias that merits new theory. The cost of keeping track model from [Haushofer, 2015] is one such approach.

We also explored two other standard properties assumed in most models of intertemporal choice – time invariance and monotonicity – and we do not detect systematic failures of either property. This suggests that these are both reasonable properties to retain in modelling intertemporal decision-making.

## Chapter 3

# Selection Effects in a First Year Seminar

Kevin Laughren\*, Lara Aknin, Dai Heide, Panayiotis Pappas, and Milan Singh

\*Corresponding author

### 3.1 Introduction

Existing studies that set out to measure the causal effect of educational programs on student outcomes are often limited to designs other than the randomized control trial (RCT) because research ethics require that a student opt-in to such programs, even when eligibility is randomly assigned. The result is that analysis of student outcomes cannot rely on random assignment to believe that the students participating in the intervention are comparable to those who do not participate. This means that any differences between treatment and control groups that we observe in such analyses could be caused not by the treatment itself, but by unobserved and meaningful differences in the sample of students who select into such programs relative to the rest of their cohort. We refer to significant effects driven by this type of sample selection as *selection effects*. Several experimental design paradigms attempt to alleviate these concerns, including studies that generate a control group by selecting a group of students who match the treated students on observable characteristics such as grades, gender, race, and socioeconomic status. We study the causal effects of a first-year seminar program at a public university using a difference-in-differences framework [Wooldridge, 2010] and find that seminar enrollment has a positive causal effect on GPA when using the entire cohort or a matched sample as the control group. Our linear probability and logistic binary regressions using these same control groups suggest that seminar enrollment positively affects the probability of first year student retention into their second year. Our contribution to the literature is to elicit students' *willingness to enroll* prior to the educational intervention, which we believe is a meaningful and unobserved covariate in most non-RCT studies of educational programs. When we include this control variable

our causal effects are no longer statistically significant, suggesting that the causal effects we observe with our typical control groups are actually selection effects. Many studies of educational interventions risk confounding genuine causal effects with underlying differences between students who opt-in to treatment versus those who do not.

Student satisfaction and retention have been a focus of undergraduate administrators for decades, and along with GPA, have been the outcomes of interest for a large program evaluation literature in higher education research. Evaluation of these programs dates at least as far back as [Pascarella and Terenzini, 1979]. [Pascarella and Terenzini, 1991] provides a survey of early studies, and [Bailey and Alfonso, 2005] provides a survey of interventions specific to community colleges where low retention is even more pernicious than at degree-granting institutions.

Aside from a few older studies such as [Strumpf and Hunt, 1993], research in this literature has not been able to conduct experiments following the RCT methodology due to ethical concerns with forcing students into any educational intervention, whether it be a specific course or an academic support program. Instead, students select themselves into such programs, and so we cannot rely on random assignment to believe that the students participating in the program are comparable to students who do not participate. As a result, any differences in outcomes between treatment and control groups that we observe in such analyses could be caused not by the treatment itself but by selection effects, through unobserved but meaningful differences in the sample of students who select into such programs relative to the rest of their cohort.

Our contribution to the educational program evaluation literature is to collect a novel proxy for a student’s willingness to enroll in a first year seminar (FYS) and demonstrate that results drawn from comparisons to an entire cohort or matched sample may be driven by selection effects, as the results are not robust when comparing to our novel control group. Notably for practitioners this is a low-cost measure but it required contact with the students of interest prior to the FYS intervention, and so while it is a desirable control it is not available to studies of past interventions using only registrar data.

We believe collecting such a proxy provides a superior control to a matched pairs protocol, because generating unbiased estimates with a matched control group requires a *selection on observables* assumption [Wooldridge, 2010]. This assumption states that after we control for the observable characteristics on which a match is generated - demographics and high school grades - two students of the same observable type are equally likely to select into a seminar. However if this assumption is not satisfied, the estimates generated by the matching control protocol are biased.

Considering that FYS courses require substantially different types of work from students - frequent writing and speaking in class, with little or no weight on final exams - we think it is plausible that the selection on observables assumption is not satisfied in this instance. Demographics and high school grades may not be enough information to accurately predict

the probability that a student enrolls themselves in a seminar. A student's desire to be evaluated on their writing and speaking in a seminar relative to their desire to be evaluated on exam performance in a lecture class is an unobserved variable that, when omitted from regression, can introduce bias to estimates of the effects of seminar participation.

To proxy for this unobserved variable, we sent an incentivized email survey to all incoming domestic Faculty of Arts and Social Sciences (Faculty) students at a large public Canadian university prior to the Fall 2017 enrollment period. This survey included detailed information on the seminar program overall (small classes, award winning faculty, emphasis on research, writing, and speaking) as well as titles and synopses of the ten seminars being offered. Students were asked to provide a Yes/No answer to whether they were willing to enroll in each of the ten, and told their responses could affect whether a seminar would run in Fall 2017. 219 eligible students responded to this survey.<sup>1</sup> We use those students who did not actually enroll in a seminar but indicated 'Yes' to being willing to enroll in some seminar as our preferred control group relative to the students who did enroll in a seminar.

In comparisons of seminar students to the entire cohort or a matched sample, we find that the first year cumulative GPA of seminar students is significantly greater. However in our preferred model specifications that control for willingness to enroll, we find enrollment in a first year seminar provides no significant benefit on student well-being, grades, or continuation to second year. We believe this is evidence that selection effects are a plausible alternative explanation for significant results in program evaluation papers that implicitly rely on a selection on observables assumption.

## Related Literature

Broadly, this is a program evaluation study making use of a difference-in-differences design [Wooldridge, 2010]. Some recent studies in higher education have used similar program evaluation techniques including regression discontinuity [Moss and Yeaton, 2013], and propensity score matching [Sneyers and Witte, 2017].

There is a substantial literature with a specific focus on college student retention as a response to an educational intervention using linear probability models (OLS): [Crissman, 2001, Andrade, 2009, Webster and Showers, 2011, Sommet et al., 2015, Swanson et al., 2017], or logistic binary variables regression [Venuleo et al., 2016]; including several which focus on minorities or international students [Barlow and Villarejo, 2004, Andrade, 2009]. The *What Works Clearinghouse* is an online repository of results from educational program evaluation studies with strict criteria for inclusion, and includes a section specific to post-secondary interventions.<sup>2</sup>

<sup>1</sup>And an additional 24 ineligible students who found the survey link through other means

<sup>2</sup><https://ies.ed.gov/ncee/wwc/FWW>

[Shanley and Witten, 1990] and [Fidler, 1991] conducted a 15-year study at University of South Carolina comparing seminar students to all other first years, and found significant improvement in retention that could not be explained by observed demographic and high school variables.

Many of the above papers cite [Tinto, 1975], [Astin, 1993], or [Seidman, 2005] as sources of theoretical hypotheses that the types of activities in a first year seminar could improve retention. We do not explicitly use these models to generate hypotheses but rely on reduced-form models to test intuitive hypotheses that are in line with these models.

The vast majority of studies - including our own - are cases where students opt into treatment, whether it be a seminar, mentor program, math prep course, etc. One exception is [Strumpf and Hunt, 1993] in which the researchers collected motivation to enroll and randomly assigned a subset of the motivated students to a seminar, leaving the rest as a control group. They found significant effects on retention and academic standing. (Random assignment of courses would not be feasible or pass ethics approval in most public universities today). [Angrist et al., 2009] randomly assign *eligibility*, but students had to actively consent to receive academic support services (with 55% consent) or financial GPA incentives (with 87% consent).

Meta analyses of retention studies [Fong et al., 2017, Fike and Fike, 2008, Colton et al., 1999] focus on demographic predictors of retention, reach little consensus, and point out methodological challenges. [Clark and Cundiff, 2011, Reid et al., 2014] are surveys that point out methodological issues in previous studies, particularly the inability of previous causal analyses to rule out potential confounding explanations, in particular unobserved characteristics affecting selection.

Many studies relied on creating a group of control students through matching, i.e., ensuring that the control group had a student who was similar to a student in the seminar based on recorded characteristics such as gender, ethnicity, and high school grades [Campbell and Campbell, 1997, Schnell and Doetkott, 2003, Hendel, 2007, Miller and Lesik, 2014]. Analyses using demographic control groups to measure causal effects implicitly rely on a ‘selection on observables’ assumption [Wooldridge, 2010]. For such regression estimates to be unbiased this theoretically requires that there is no difference in error distributions between the two populations after controlling for observable variables. This matters in the case of a seminar because any regression is still omitting a known difference between the two groups: willingness to enroll in a seminar-style course over a traditional large lecture course.

## 3.2 Methods

The Faculty launched nine first year seminars led by highly regarded permanent faculty members in the Fall 2017 semester.<sup>3</sup> These seminars were open only to domestic first-year non-transfer students in the Faculty. Each seminar enrollment was limited to 25 students.

The rationale for offering the seminars was twofold. First, to determine whether offering students this kind of first-year experience would positively impact their GPA and their overall well-being in addition to reducing the likelihood that they will leave the university without graduating. Second, such seminars were thought to allow students to develop closer bonds with faculty and to feel more closely connected with the university and with their peers at an early stage in their academic careers.

The pilot was meant to examine the long-term effects of participation in a seminar on GPA and student retention. While meaningful data on graduation will not be available for several years, we report on one-year retention and first year GPA. To measure this, we identified a control group by surveying all incoming domestic Faculty students concerning their level of interest in the courses (regardless of whether they in fact enroll). Focusing only on those students willing to enroll in such a seminar helps us to I selection effects, which are inevitable in this setting where we cannot choose which students enroll in a seminar. In our causal effects framework, we compare first year domestic students who enroll in FYS with willing students who did not enroll, and determine whether participation in FYS positively affects GPA, academic status, and retention. If nothing else, our willingness proxy identifies students who are reading and actively responding to emails prior to and early in the enrollment period, which is an often unobserved but plausibly meaningful covariate.

This project also examines the immediate impact of participation in FYS on student engagement and well-being. We collected validated measures of well-being in surveys of students at two times: prior to the start of the semester, and at the end of the semester. We use a difference-in-differences framework to examine whether participation in FYS positively impacts student well-being and engagement, while controlling for student selection, beginning versus end of term effects, and covariates such as entering (high school) GPA, gender, and course load.

In total, 124 domestic first year students enrolled in one of the seminars. These students achieved a 2.57 GPA in their non-FYS courses in Fall 2017, compared to a 2.31 GPA in non-FYS courses for the entire cohort. Of course there are selection issues that inhibit us from making a causal interpretation between these two numbers.

<sup>3</sup>One of the ten seminars listed in the survey was canceled due to lack of enrollment



## **Data**

### **University Registrar**

The university’s registrar provided the majority of our data, including all data on demographics, enrollment, academic status, and grades. Most of our analyses require only these data, except where we controlled for willingness to enroll or wanted to look for effects of FYS on psychological measures of well-being.

### **Willingness to enroll**

We sent an email survey with a small financial incentive to all incoming domestic Faculty students prior to the Fall 2017 enrollment period. This survey included detailed information on the seminar program overall (small classes, award winning instructors, emphasis on research, writing, and speaking) as well as titles and synopses of the ten seminars being offered. Students were asked to provide a Yes/No answer to whether they were willing to enroll in each of the ten, and told their responses could affect whether a seminar would run in Fall 2017 (an incentive to respond truthfully). 219 eligible students responded to this survey. We use those students who did not actually enroll in a seminar but indicated Yes to being willing to enroll in some seminar as our preferred control group relative to the students who did enroll in a seminar. We also compare a matched control group for illustration.

### **Well-being and engagement**

Our surveys of well-being and engagement were conducted online, with domestic first-year Faculty students recruited directly via email. Completion was incentivized with a draw for Visa gift cards in \$50CAD and \$100CAD denominations. Individual responses to a series of Likert-scale style questions are aggregated into individual measures of Depression (CESD) [Radloff, 1977], Loneliness [Russell et al., 1978], Satisfaction with Life [Diener et al., 1985], Social Connection [Lee et al., 2001], Anxiety [Spitzer et al., 2006], Flourishing, and frequency of Positive and Negative Emotions (SPANE) [Diener et al., 2010]. The measures were chosen for the survey based on their common usage in psychology research and strong performance on validity measures such as test-retest reliability, internal consistency, and convergent validity [Russell, 1996, Silva and Caetano, 2013, Aishvarya et al., 2014]. We also asked about a students’ feelings of Belonging at the University, and about their number of close friends and acquaintances.

311 students completed at least one well-being survey, 71 of whom completed it both before and after the semester.

To examine how the measures change over the course of a semester, we restrict attention to the 71 students who completed a survey twice (to control for sample selection). Within these students, we calculate the correlation coefficient between well-being measures, as well

as covariates (high school GPA, gender), treatment status (TREATMENT), and a dummy variable for measurements done post-seminar (POST). In Appendix 3a, we provide results of a t-test on each Pearson correlation coefficient to determine if the measured correlations are significantly different from zero. None of the measures of well-being are significantly correlated with POST, which tells us that the overall average response of these students did not change significantly over the Fall 2017 semester. The correlation matrix also suggests that female students have significantly higher measures of Negative Emotions, Anxiety, Social Connection, and Symptoms of Depression. Notably, high school grades (eGPA) are significantly correlated with Satisfaction with Life.<sup>4</sup>

## Empirical Frameworks

### Difference-in-differences

We implement a difference-in-differences causal effects framework [Wooldridge, 2010] using ordinary least squares (OLS) and fixed effects (FE) frameworks in an attempt to identify causal effects of the seminar on well-being and GPA. The difference-in-differences framework involves running a regression of the form:

$$y_{i,t} = \alpha_i + \beta_1 * TREATMENT + \beta_2 * POST + \delta * TREATMENT : POST + \theta * X_i + u_{i,t}$$

Where  $y_{i,t}$  is the outcome variable  $y$  (e.g., GPA, academic status, anxiety) for individual  $i$  measured at time  $t$ .  $TREATMENT = 1$  for individuals who enrolled in FYS and zero otherwise.  $POST = 1$  for observations made post-seminar (i.e., December 2017 or later) and zero otherwise.  $X_i$  is a vector of individual  $i$ 's covariate characteristics such as their high school grades, home province, gender, etc. In the OLS specification we can estimate  $\theta$ , the effect of covariates on outcomes like GPA because of the assumption that all individuals have the same intercept:  $\alpha_i = \alpha \quad \forall i$ . In the FE specification, the coefficient on any covariates  $X_i$  that do not change across  $t$  (such as gender) are all collapsed into an individual estimate of  $\alpha_i$ . The causal effect of seminar enrollment on outcome  $y$  is measured by  $\delta$ , which can be thought of as the interaction effect on  $TREATMENT*POST$ . Letting  $t \in \{PRE, POST\}$ , algebraic manipulation of the regression equation reveals why  $\delta$  is known as the difference-in-differences estimator:

$$\delta = (\overline{y_{T,POST}} - \overline{y_{T,PRE}}) - (\overline{y_{C,POST}} - \overline{y_{C,PRE}})$$

Where T represents treatment group, C represents the control group, and the over-bar represents the average over the group in the specified time period (PRE or POST). We

<sup>4</sup>There are a number of significant correlations between well-being variables; for example, our measure of Belonging is significantly negatively correlated with Loneliness, Social Connection, and Symptoms of Depression.

could calculate  $\delta$  by hand, but by using regression we obtain standard errors that help us understand whether the estimated value is significantly different from zero.

In all specifications, individuals who enrolled in a seminar in Fall 2017 are identified as the Treatment group, and we vary the control group to demonstrate the lack of robustness of effects to our measure of willingness to enroll. Specifically, each regression table is subdivided into three sections: where the control group are students who responded they were willing to enroll in a seminar but did not, where the control group are students matched to a treatment student on the basis of demographics and high school grades, and finally where the control group is the entire cohort of first year domestic students.

### Non-panel frameworks

While we have measures of well-being and grades prior to seminar enrollment, a number of measures are only observed once (after the seminar) and so a difference-in-differences framework is not applicable. In particular, for grades, academic status, and enrollment we specify a linear model of the form:

$$y_{i,t} = \alpha + \beta_1 * TREATMENT + \theta * X_i + u_{i,t}$$

For binary variables (enrollment and academic status) we estimate the logit regression:

$$Prob(y_{i,t} = 1) = \frac{e^{\alpha + \beta_1 * TREATMENT_i + \theta * X_i}}{1 + e^{\alpha + \beta_1 * TREATMENT_i + \theta * X_i}}$$

$$\log\left(\frac{Prob(y_{i,t} = 1)}{1 - Prob(y_{i,t} = 1)}\right) = \alpha + \beta_1 * TREATMENT_i + \theta * X_i$$

## 3.3 Results

### Baseline Equivalence

In this quasi-experimental methodology, it is important to establish that the treatment and control groups are comparable on observable baseline dimensions. We compare the students enrolled in a seminar in the left column to three potential control groups in Table 3.1. By construction, the *Match* control group is almost exactly the same as the *Seminar* group on observable dimensions. The *Willing* control group (those who took the pre-enrollment survey and indicated they were willing to enroll in a seminar, but did not end up doing so) have similar gender proportions and course load to the seminar group, but have higher high school grades (86.6 versus 84.8) and fewer students identifying as First Nations (0.7% versus 2.4% for the treatment group). These differences are not significant at a 5% level using a Mann-Whitney-U test of means. Using the entire cohort as a control group introduces substantial difference in gender ratio (64.4% female versus 74.2% in the treatment group), as female students had a higher propensity to select into FYS and respond to the willingness

email survey. A number of studies have demonstrated that women achieve higher post-secondary grades, so we should expect a comparison between *Seminar* students and the *Cohort* control group to show greater grades for the *Seminar* group.<sup>5</sup>

Table 3.1: Baseline Measures

	<i>Seminar Group</i>	<i>Control Groups</i>		
	TREATMENT = 1	<i>Willing</i>	<i>Match</i>	<i>Cohort</i>
Number of Students	124	151	124	1014
High School GPA (out of 100)	84.82	86.61	84.87	85.43
% Female	74.2%	74.8%	74.2%	64.4%
% First Nations	2.4%	0.7%	2.4%	1.3%
Fall 2017 Course Units	11.19	11.20	11.25	11.05

## Seminar Effect on GPA

Recall that the causal effect of FYS on GPA in the difference-in-differences framework,  $\delta$  is the coefficient on TREATMENT:POST. Table 3.2 demonstrates that the estimate of this effect is positive and significant when comparing FYS students to the *Match* or *Cohort* control groups, but there is zero effect of FYS when we compare treated students to the *Willing* control group. This is the first evidence we provide of selection effects in frameworks that do not control for willingness to enroll.

## Seminar Effect on Well-being

We conducted two optional surveys prior to and following the seminar in Fall 2017 to collect psychological measures of well-being. If the seminar had significant effects on academic outcomes, these results might help us to identify a mechanism by which this happens (e.g., by increasing feelings of belonging to the school community). We used the same difference-in-differences framework as we did for GPA in an attempt to measure any causal effects of the seminar on these outcomes. The sample is smaller than the GPA analysis however, because we are limited to those students who chose to complete both surveys of well-being, and who identified themselves as willing seminar enrollees either through our pre-enrollment survey or by enrolling themselves. The table below provides our difference-in-differences fixed effects estimates with each column representing a different measure of well-being.

<sup>5</sup>There were fewer than ten students who did not provide gender information to the registrar or who identified as a gender other than male or female. This group does not significantly affect overall results and we do not provide separate figures for a small and plausibly identifiable group, instead they are included among the larger group of students who are not female.

Table 3.2: Seminar Effect on GPA

<i>Dependent variable: GPA_Proxy</i>						
Control Group:	<i>Willing</i>		<i>Match</i>		<i>Cohort</i>	
	<i>OLS</i> (1)	<i>FE</i> (2)	<i>OLS</i> (3)	<i>FE</i> (4)	<i>OLS</i> (5)	<i>FE</i> (6)
TREATMENT	−0.099 (0.077)		−0.022 (0.089)		−0.071 (0.062)	
POST	−1.262*** (0.073)	−1.264*** (0.060)	−1.549*** (0.089)	−1.549*** (0.080)	−1.481*** (0.029)	−1.482*** (0.025)
GENDERF	0.280*** (0.062)		0.395*** (0.072)		0.210*** (0.029)	
TREATMENT:POST	−0.0001 (0.109)	0.002 (0.089)	0.286** (0.126)	0.286** (0.113)	0.219** (0.088)	0.219*** (0.077)
Constant	3.989*** (0.054)		3.941*** (0.066)		3.943*** (0.023)	
Observations	550	550	496	496	2,276	2,276
R <sup>2</sup>	0.510	0.749	0.521	0.717	0.559	0.767

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 3.3 shows that of the nine psychological measures, we see a significant causal effect of the seminar on only one: CESD (a measure of how frequently an individual experiences symptoms of depression). This effect is significant only to a 10% significance level. The probability of observing at least one significant result at a 10% level when conducting nine independent tests is substantial, so we interpret these results as inconclusive across the board. There seems to be no causal effect of enrolling in a first year seminar on any of the following: symptoms of depression, loneliness, flourishing, anxiety, satisfaction with life, social connection, frequency of positive emotions, frequency of negative emotions, or belonging to the school community.

Table 3.3: Seminar Effect on Survey Measures of Well-being

	<i>Dependent variable:</i>								
	Symptoms of Depression	Loneliness	Flourishing	Anxiety	Satisfaction with Life	Social Connection	Frequency of Pos. Emotions	Frequency of Neg. Emotions	Belonging at University
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
POST	0.201 (0.127)	0.625 (0.633)	0.005 (0.150)	0.783 (0.973)	-0.192 (0.216)	0.059 (0.049)	0.080 (0.693)	0.680 (0.836)	-0.522** (0.225)
TREATMENT:POST	-0.373* (0.198)	-0.625 (1.001)	-0.218 (0.233)	-1.253 (1.493)	0.251 (0.340)	-0.099 (0.077)	0.861 (1.090)	0.555 (1.314)	0.584 (0.351)
Observations	82	83	84	83	85	80	85	85	82
R <sup>2</sup>	0.093	0.025	0.036	0.021	0.021	0.052	0.031	0.051	0.128

Note: Observation numbers vary slightly as all questions were optional

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Seminar Effect on Academic Status and Retention

Tables 3.4 and 3.5 provide a linear probability model (OLS) and a logistic binary variable regression on Spring 2018 academic standing and Fall 2018 (second year) enrollment, respectively.

Table 3.4 shows that FYS enrollment has a positive effect on academic standing when comparing enrolled students to the *Match* or *Cohort* control groups, but again there is no significant effect when FYS students are compared to the *Willing* control group.

Table 3.5 shows that FYS students are significantly more likely to be retained to second year (as measured by Fall 2018 enrollment) than the control group generated by matching covariates, but that there is no significant effect of FYS on enrollment relative to the *Cohort* or *Willing* control groups.

## 3.4 Discussion

We are able to replicate significant positive effects of a first year seminar (FYS) on GPA, academic status, and retention when we compare participants to convenient control groups used in previous studies, whether they be a group matched on demographics, or a comparison to the overall cohort. However, when we use our pre-enrollment survey measure and compare FYS students only to others who are interested in such a course but do not

Table 3.4: Seminar Effect on Academic Standing after two semesters

<i>Dependent variable: Spring 2018 Good Academic Standing</i>						
Control Group:	<i>Willing</i>		<i>Match</i>		<i>Cohort</i>	
	<i>OLS</i> (1)	<i>logit</i> (2)	<i>OLS</i> (3)	<i>logit</i> (4)	<i>OLS</i> (5)	<i>logit</i> (6)
TREATMENT	0.040 (0.047)	0.386 (0.332)	0.121** (0.053)	0.726** (0.313)	0.090** (0.041)	0.561** (0.248)
HIGHSCHOOLGPA	0.017*** (0.003)	0.192*** (0.038)	0.008** (0.003)	0.062** (0.030)	0.016*** (0.002)	0.127*** (0.016)
GENDERF	0.101* (0.054)	0.606* (0.340)	0.223*** (0.062)	1.099*** (0.325)	0.066** (0.027)	0.314** (0.142)
Constant	−0.626** (0.280)	−14.899*** (3.243)	0.091 (0.257)	−4.200 (2.595)	−0.634*** (0.172)	−9.785*** (1.345)
Observations	275	275	248	248	1,138	1,138
R <sup>2</sup>	0.118		0.109		0.069	
Log Likelihood		−119.284		−128.981		−624.536

Note: Logistic coefficients (not marginal effects) displayed for comparison to standard errors

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 3.5: Seminar Effect on Second-year Enrollment

<i>Dependent variable: Fall 2018 Enrollment</i>						
Control Group:	<i>Willing</i>		<i>Match</i>		<i>Cohort</i>	
	<i>OLS</i> (1)	<i>logit</i> (2)	<i>OLS</i> (3)	<i>logit</i> (4)	<i>OLS</i> (5)	<i>logit</i> (6)
TREATMENT	−0.018 (0.042)	−0.154 (0.357)	0.113** (0.051)	0.704** (0.324)	0.052 (0.038)	0.366 (0.264)
HIGHSCHOOLGPA	0.005* (0.003)	0.032 (0.022)	0.001 (0.003)	0.003 (0.016)	0.005** (0.002)	0.025** (0.011)
GENDERF	0.043 (0.048)	0.351 (0.386)	0.094 (0.060)	0.543 (0.349)	−0.005 (0.025)	−0.030 (0.159)
Constant	0.435* (0.252)	−0.702 (1.941)	0.709*** (0.248)	0.888 (1.386)	0.408*** (0.157)	−0.750 (0.920)
Observations	275	275	248	248	1,138	1,138
R <sup>2</sup>	0.019		0.030		0.007	
Log Likelihood		−108.269		−123.625		−560.751

Note: Logistic coefficients (not marginal effects) displayed for comparison to standard errors

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

enroll, all significant effects disappear. Controlling only for observable characteristics such as demographics, high school grades, and location does not appear to be sufficient to use such groups for causal effects analysis because the selection on observables assumption is violated. Specifically, after controlling for observable characteristics, it is not true that all students are equally likely to enroll in FYS; most studies omit an unobserved characteristic *willingness to enroll* that captures student preferences for small interactive classrooms with less weight on exams. This unobserved and uncontrolled preference could create spurious results when evaluating first year seminars or other interventions where a student selecting themselves in reveals something about them that is not captured by a registrar database.

**Funding** Research assistance and survey incentives were funded by Teaching and Learning Development Grants (TLDG Projects G0233, G0234) from the Institute for the Study of Teaching and Learning in the Disciplines at Simon Fraser University.



# Bibliography

- [Aguiar et al., 2021] Aguiar, V. H., Boccardi, M. J., Kashaev, N., and Kim, J. (2021). Random utility and limited consideration.
- [Aishvarya et al., 2014] Aishvarya, S., Maniam, T., Karuthan, C., Sidi, H., Jaafar, N. R. N., and Oei, T. P. S. (2014). Psychometric properties and validation of the satisfaction with life scale in psychiatric and medical outpatients in malaysia. *Comprehensive Psychiatry*, 55:S101–S106.
- [Andrade, 2009] Andrade, M. S. (2009). The value of a first-year seminar: International students’ insights in retrospect. *Journal of College Student Retention: Research, Theory & Practice*, 10(4):483–506.
- [Andreoni and Sprenger, 2012] Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–56.
- [Angrist et al., 2009] Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–63.
- [Ashraf et al., 2006] Ashraf, N., Karlan, D., and Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *The Quarterly Journal of Economics*, 121(2):635–672.
- [Astin, 1993] Astin, A. W. (1993). College retention rates are often misleading. *Chronicle of Higher Education*, 40(5):A48–A48.
- [Augenblick et al., 2015] Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115.
- [Augenblick and Rabin, 2019] Augenblick, N. and Rabin, M. (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, 86(3):941–975.
- [Aycinena et al., 2020] Aycinena, D., Blazsek, S., Rentschler, L., and Sprenger, C. (2020). Intertemporal choice experiments and large-stakes behavior. Technical report, Chapman University, Economic Science Institute.
- [Bailey and Alfonso, 2005] Bailey, T. R. and Alfonso, M. (2005). *Paths to persistence: An analysis of research on program effectiveness at community colleges*. Lumina Foundation for Education.

- [Barlow and Villarejo, 2004] Barlow, A. E. and Villarejo, M. (2004). Making a difference for minorities: Evaluation of an educational enrichment program. *Journal of Research in Science Teaching*, 41(9):861–881.
- [Bartoš et al., 2016] Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *The American Economic Review*, 106(6):1437–1475.
- [Bryan et al., 2010] Bryan, G., Karlan, D., and Nelson, S. (2010). Commitment devices. *Annu. Rev. Econ.*, 2(1):671–698.
- [Campbell and Campbell, 1997] Campbell, T. A. and Campbell, D. E. (1997). Faculty/student mentor program: Effects on academic performance and retention. *Research in Higher Education*, 38(6):727–742.
- [Caplin and Dean, 2015] Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *The American Economic Review*, 105(7):2183–2203.
- [Caplin et al., 2018] Caplin, A., Dean, M., and Leahy, J. (2018). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, 86(3):1061–1094.
- [Cattaneo et al., 2020] Cattaneo, M. D., Ma, X., Masatlioglu, Y., and Suleymanov, E. (2020). A random attention model. *Journal of Political Economy*, 128(7):2796–2836.
- [Chen et al., 2016] Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.
- [Clark and Cundiff, 2011] Clark, M. and Cundiff, N. L. (2011). Assessing the effectiveness of a college freshman seminar using propensity score adjustments. *Research in Higher Education*, 52(6):616–639.
- [Coller and Williams, 1999] Coller, M. and Williams, M. B. (1999). Eliciting individual discount rates. *Experimental Economics*, 2(2):107–127.
- [Colton et al., 1999] Colton, G. M., Ulysses J. Connor, J., Shultz, E. L., and Easter, L. M. (1999). Fighting attrition: One freshman year program that targets academic progress and retention for at-risk students. *Journal of College Student Retention: Research, Theory & Practice*, 1(2):147–162.
- [Crissman, 2001] Crissman, J. L. (2001). The impact of clustering first year seminars with english composition courses on new students’ retention rates. *Journal of College Student Retention: Research, Theory & Practice*, 3(2):137–152.
- [Cubitt and Read, 2007] Cubitt, R. P. and Read, D. (2007). Can intertemporal choice experiments elicit time preferences for consumption? *Experimental Economics*, 10(4):369–389.
- [Dean and Martin, 2016] Dean, M. and Martin, D. (2016). Measuring rationality with the minimum cost of revealed preference violations. *Review of Economics and Statistics*, 98(3):524–534.

- [Dean and Neligh, 2019] Dean, M. and Neligh, N. L. (2019). Experimental tests of rational inattention. Technical report, Columbia Economics Discussion Papers.
- [Demuynck and Hjertstrand, 2019] Demuynck, T. and Hjertstrand, P. (2019). Samuelson’s approach to revealed preference theory: Some recent advances. In Cord, R. A., Anderson, R. G., and Barnett, W. A., editors, *Paul Samuelson Master of Modern Economics*, chapter Ch.9. Springer.
- [Diener et al., 1985] Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, 49(1):71–75.
- [Diener et al., 2010] Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-w., Oishi, S., and Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, 97(2):143–156.
- [Enke, 2017] Enke, B. (2017). What you see is all there is. *Working Paper marked revise and resubmit at Quarterly Journal of Economics*.
- [Ericson, 2017] Ericson, K. M. (2017). On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *Journal of the European Economic Association*, 15(3):692–719.
- [Fenig et al., 2018] Fenig, G., Gallipoli, G., and Halevy, Y. (2018). Piercing the ‘payoff function’ veil: Tracing beliefs and motives. Technical report, University of Toronto, Department of Economics.
- [Fidler, 1991] Fidler, P. (1991). Relationship of freshman orientation seminars to sophomore return rates. *Journal of the First-Year Experience & Students in Transition*, 3(1):7–38.
- [Fike and Fike, 2008] Fike, D. S. and Fike, R. (2008). Predictors of first-year student retention in the community college. *Community College Review*, 36(2):68–88.
- [Fong et al., 2017] Fong, C. J., Davis, C. W., Kim, Y., Kim, Y. W., Marriott, L., and Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research*, 87(2):388–424.
- [Fournier et al., 2019a] Fournier, L. R., Coder, E., Kogan, C., Raghunath, N., Taddese, E., and Rosenbaum, D. A. (2019a). Which task will we choose first? procrastination and cognitive load in task ordering. *Attention, Perception, & Psychophysics*, 81(2):489–503.
- [Fournier et al., 2019b] Fournier, L. R., Stubblefield, A. M., Dyre, B. P., and Rosenbaum, D. A. (2019b). Starting or finishing sooner? sequencing preferences in object transfer tasks. *Psychological research*, 83(8):1674–1684.
- [Freeman, 2021] Freeman, D. J. (2021). Revealing naïveté and sophistication from procrastination and preproportion. *American Economic Journal: Microeconomics*, 13(2):402–38.
- [Gabaix, 2014] Gabaix, X. (2014). A sparsity-based model of bounded rationality \*. *The Quarterly Journal of Economics*, 129(4):1661.
- [Halevy, 2015] Halevy, Y. (2015). Time consistency: Stationarity and time invariance. *Econometrica*, 83(1):335–352.

- [Harrison et al., 2002] Harrison, G. W., Lau, M. I., and Williams, M. B. (2002). Estimating individual discount rates in denmark: A field experiment. *American economic review*, 92(5):1606–1617.
- [Haushofer, 2015] Haushofer, J. (2015). The cost of keeping track. Technical report, Cite-seer.
- [Hayek, 1945] Hayek, F. A. (1945). The use of knowledge in society. *The American economic review*, 35(4):519–530.
- [Hendel, 2007] Hendel, D. D. (2007). Efficacy of participating in a first-year seminar on student satisfaction and retention. *Journal of College Student Retention: Research, Theory & Practice*, 8(4):413–423.
- [Heufer and Hjertstrand, 2015] Heufer, J. and Hjertstrand, P. (2015). Consistent subsets: Computationally feasible methods to compute the houtman–maks-index. *Economics Letters*, 128:87–89.
- [Holt and Smith, 2009] Holt, C. A. and Smith, A. M. (2009). An update on bayesian updating. *Journal of Economic Behavior & Organization*, 69(2):125–134.
- [Houtman and Maks, 1985] Houtman, M. and Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve methoden*, 19(1):89–104.
- [Karni, 2009] Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica*, 77(2):603–606.
- [Laibson, 1997] Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478.
- [Lee et al., 2001] Lee, R. M., Draper, M., and Lee, S. (2001). Social connectedness, dysfunctional interpersonal behaviors, and psychological distress: Testing a mediator model. *Journal of counseling psychology*, 48(3):310.
- [Manzini and Mariotti, 2007] Manzini, P. and Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, 97(5):1824–1839.
- [Manzini and Mariotti, 2014] Manzini, P. and Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176.
- [Martínez-Marquina et al., 2018] Martínez-Marquina, A., Niederle, M., and Vespa, E. (2018). Failures in contingent reasoning: The role of uncertainty. *American Economic Review*.
- [Masatlioglu et al., 2012] Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183–2205.
- [Matejka and McKay, 2014] Matejka, F. and McKay, A. (2014). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review*, 105(1):272–298.
- [McFadden, 1973] McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*.

- [Miller and Lesik, 2014] Miller, J. W. and Lesik, S. S. (2014). College persistence over time and participation in a first-year seminar. *Journal of College Student Retention: Research, Theory & Practice*, 16(3):373–390.
- [Moss and Yeaton, 2013] Moss, B. G. and Yeaton, W. H. (2013). Evaluating effects of developmental education for college students using a regression discontinuity design. *Evaluation Review*, 37(5):370–404. PMID: 24662603.
- [O’Donoghue and Rabin, 1999] O’Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American economic review*, 89(1):103–124.
- [Pascarella and Terenzini, 1979] Pascarella, E. T. and Terenzini, P. T. (1979). Student-faculty informal contact and college persistence: A further investigation. *The Journal of Educational Research*, 72(4):214–218.
- [Pascarella and Terenzini, 1991] Pascarella, E. T. and Terenzini, P. T. (1991). *How college affects students*, volume 1991. Jossey-Bass San Francisco.
- [Rabin, 2000] Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem econometrica. *September*, 68:5.
- [Radloff, 1977] Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401.
- [Read and Van Leeuwen, 1998] Read, D. and Van Leeuwen, B. (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational behavior and human decision processes*, 76(2):189–205.
- [Reid et al., 2014] Reid, K. M., Reynolds, R. E., and Perkins-Auman, P. G. (2014). College first-year seminars: What are we doing, what should we be doing? *Journal of College Student Retention: Research, Theory & Practice*, 16(1):73–93.
- [Rosenbaum et al., 2019] Rosenbaum, D. A., Fournier, L. R., Levy-Tzedek, S., McBride, D. M., Rosenthal, R., Sauerberger, K., VonderHaar, R. L., Wasserman, E. A., and Zentall, T. R. (2019). Sooner rather than later: Precrastination rather than procrastination. *Current Directions in Psychological Science*, 28(3):229–233.
- [Rosenbaum et al., 2014] Rosenbaum, D. A., Gong, L., and Potts, C. A. (2014). Pre-crastination: Hastening subgoal completion at the expense of extra physical effort. *Psychological Science*, 25(7):1487–1496.
- [Russell et al., 1978] Russell, D., Peplau, L. A., and Ferguson, M. L. (1978). Developing a measure of loneliness. *Journal of personality assessment*, 42(3):290–294.
- [Russell, 1996] Russell, D. W. (1996). Ucla loneliness scale (version 3): Reliability, validity, and factor structure. *Journal of personality assessment*, 66(1):20–40.
- [Schnell and Doetkott, 2003] Schnell, C. A. and Doetkott, C. D. (2003). First year seminars produce long-term impact. *Journal of College Student Retention: Research, Theory & Practice*, 4(4):377–391.
- [Schotter and Trevino, 2014] Schotter, A. and Trevino, I. (2014). Belief elicitation in the laboratory. *Annu. Rev. Econ.*, 6(1):103–128.

- [Seidman, 2005] Seidman, A. (2005). Minority student retention: Resources for practitioners. *New directions for institutional research*, 2005(125):7–24.
- [Shanley and Witten, 1990] Shanley, M. G. and Witten, C. H. (1990). University 101 freshman seminar course: A longitudinal study of persistence, retention, and graduation rates. *NASPA Journal*, 27(4):344–352.
- [Silva and Caetano, 2013] Silva, A. J. and Caetano, A. (2013). Validation of the flourishing scale and scale of positive and negative experience in portugal. *Social Indicators Research*, 110(2):469–478.
- [Sims, 2003] Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.
- [Sneyers and Witte, 2017] Sneyers, E. and Witte, K. D. (2017). The effect of an academic dismissal policy on dropout, graduation rates and student satisfaction. evidence from the netherlands. *Studies in Higher Education*, 42(2):354–389.
- [Sommet et al., 2015] Sommet, N., Quiamzade, A., Jury, M., and Mugny, G. (2015). The student-institution fit at university: interactive effects of academic competition and social class on achievement goals. *Frontiers in Psychology*, 6:769.
- [Spitzer et al., 2006] Spitzer, R. L., Kroenke, K., Williams, J. B., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the gad-7. *Archives of internal medicine*, 166(10):1092–1097.
- [Stigler, 1961] Stigler, G. J. (1961). The economics of information. *Journal of political economy*, 69(3):213–225.
- [Strumpf and Hunt, 1993] Strumpf, G. and Hunt, P. (1993). The effects of an orientation course on the retention and academic standing of entering freshmen, controlling for the volunteer effect. *Journal of The First-Year Experience and Students in Transition*, 5(1):7–14.
- [Swanson et al., 2017] Swanson, N. M., Vaughan, A. L., and Wilkinson, B. D. (2017). First-year seminars: Supporting male college students’s long-term academic success. *Journal of College Student Retention: Research, Theory & Practice*, 18(4):386–400.
- [Takeuchi, 2011] Takeuchi, K. (2011). Non-parametric test of time consistency: Present bias and future bias. *Games and Economic Behavior*, 71(2):456–478.
- [Thaler, 1981] Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics letters*, 8(3):201–207.
- [Tinto, 1975] Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1):89–125.
- [Venuleo et al., 2016] Venuleo, C., Mossi, P., and Salvatore, S. (2016). Educational subculture and dropping out in higher education: a longitudinal case study. *Studies in Higher Education*, 41(2):321–342.
- [Wasserman, 2019] Wasserman, E. A. (2019). Predestination: The fierce urgency of now. *Learning & behavior*, 47(1):7–28.

- [Webster and Showers, 2011] Webster, A. L. and Showers, V. E. (2011). Measuring predictors of student retention rates. *American Journal of Economics and Business Administration*, 3(2):301–311.
- [Wooldridge, 2010] Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- [Zhong, 2019] Zhong, W. (2019). Optimal dynamic information acquisition.



# Appendix A

## Supplements to Chapter 1

### A.1 Applying Model Axioms

Each action  $a$  is a mapping from the state space  $\Omega$  to a prize space  $X$ . Formally,  $F = X^\Omega$  is the grand set of actions and  $\mathcal{F} \equiv \{A \subset F \mid |A| < \infty\}$  is the grand set of decision problems. I model the DM as choosing an information structure  $\pi : \Omega \rightarrow \Delta(\Gamma)$  which is a stochastic mapping from objective states of the world to subjective signals. Like [Caplin and Dean, 2015], I identify each subjective signal with its associated posterior beliefs  $\gamma \in \Gamma = \Delta(\Omega)$ . Let  $\Gamma(\pi)$  denote the set of possible posteriors when using information structure  $\pi$ .

A state dependent stochastic choice dataset  $(D, P)$  is a collection of decision problems  $D \subset F$  and a related set of state dependent stochastic choice functions  $P = \{P_A\}_{A \in D}$  where  $P_A : \Omega \rightarrow \Delta(A)$ . Denote  $P_A(a|\omega)$  as the probability the DM chooses action  $a$  conditional on state  $\omega$  in decision problem  $A$ , and  $\hat{P}_A(a|\omega)$  its experimental analogue.

Denote the gross payoff of using information structure  $\pi$  in decision problem  $A$  as  $G(A|\pi)$ , which can be very generally defined as:

$$G(A, \pi) = \sum_{\gamma \in \Gamma(\pi)} \left[ \sum_{\omega \in \{R, B\}} \mu(\omega) \pi(\gamma|\omega) \right] \left[ \max_{a \in \{r, b, g\}} \sum_{\omega \in \{R, B\}} \gamma(\omega) u(a(\omega)) \right] \quad (\text{A.1})$$

The first term in brackets captures the probability of a posterior given the information structure chosen, and the second term in brackets captures the gross benefit of choosing the optimal action given the posterior.

[Caplin and Dean, 2015] suppose that a DM has an information cost function  $K()$ , an attention function  $\pi_A$  which captures the DM's choice of information structure in a given decision problem  $A$ , and a choice function  $C_A : \Gamma(\pi_A) \rightarrow \Delta(A)$  which maps posterior beliefs to action probabilities. [Caplin and Dean, 2015] state that a state dependent stochastic choice dataset  $(D, P)$  has a *costly information representation* if there exists  $K(), \pi_A(), C_A()$  such that for all  $A \in D$ :

1. Information is optimal:  $\pi_A \in \arg \max_{\pi \in \Pi} \{G(A, \pi) - K(\pi)\}$



2. Choices are optimal: Given  $C_A(a|\gamma) > 0$ :

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \text{ for all } b \in A$$

3. The data is matched:  $P_A(a|\omega) = \sum_{\gamma \in \Gamma(\pi_A)} \pi_A(\gamma|\omega) C_A(a|\gamma)$

[Caplin and Dean, 2015] show a state dependent stochastic choice dataset  $(D, P)$  has a costly information representation if and only if it satisfies the testable criteria *No Improving Attention Cycles* (NIAC) and *No Improving Action Switches* (NIAS). NIAC is a property of the entire dataset  $(D, P)$  and requires that there could be no gross payoff improvement by reassigning the chosen information structures across decision problems. NIAS is a property of each decision problem, and requires that the observed actions chosen are optimal given the revealed posterior beliefs of the DM.

### NIAS violations

NIAS is a property of each decision problem, and requires that the observed actions chosen are optimal given the revealed posterior beliefs of the DM. Given  $\mu \in \Gamma, A \in D, P_A \in P$ , and  $a \in \text{Supp}(P_A)$ , the revealed posterior  $\bar{\gamma}_A^a \in \Gamma$  is defined by:

$$\bar{\gamma}_A^a(\omega) \equiv \frac{\mu(\omega) P_A(a|\omega)}{\sum_{\nu \in \Omega} \mu(\nu) P_A(a|\nu)}$$

One formulation of the NIAS condition is given in [Caplin and Dean, 2015] equation (3): For any  $A \in D, a \in \text{Supp}(P_A)$ , and  $b \in A$ :

$$\sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega) u(b(\omega)) \quad (\text{A.2})$$

A violation of the NIAS condition in an experimental dataset implies that the DM failed to choose the optimal action given their posterior beliefs  $\gamma$  in some decision problem  $A$ . One example of a NIAS violation is when a posterior belief is strong enough that the optimal action is a prediction (e.g., if  $\gamma(R) = 1$ , the optimal  $a = r$ ) and the DM is observed to choose one of the suboptimal actions  $b$  or  $g$ . I interpret such a violation of NIAS as *ineffective attention*; the choice of  $\pi$  may have been ex ante optimal, but the DMs use of the information was not effective.

NIAC is a property of the entire dataset and requires that there could be no gross payoff improvement by reassigning the chosen information structures across decision problems.

## NIAC violations

[Caplin and Dean, 2015] define the NIAC as follows: Given  $\mu$  and  $u : X \rightarrow \mathbb{R}$ , dataset  $(D, P)$  satisfies NIAC if, for any set of decision problems  $A^1, A^2, \dots, A^J \in D$  with  $A^J = A^1$ :

$$\sum_{j=1}^{J-1} G(A^j, \bar{\pi}_{A^j}) \geq \sum_{j=1}^{J-1} G(A^j, \bar{\pi}_{A^{j+1}}) \quad (\text{A.3})$$

A violation of the NIAC condition (A.3) in an experimental dataset would suggest that the DM failed to allocate attention efficiently across decision problems.

## Costly Information Representation: Translating to this Experiment Context

In the simple context of this experiment the second term in A.1 in brackets is equal to  $u(P)$  when the state is known ( $\gamma(R) \in \{0, 1\}$ ) and is equal to  $\frac{1}{2}u(P)$  when the state is not known ( $\gamma(R) = \frac{1}{2}$ ). The space of posteriors is restricted, with  $|\Gamma(\pi_0)| = 1$  and  $|\Gamma(\pi_L)| = 3$  with  $pi_L(\gamma = \frac{1}{2}|\omega) \rightarrow 0$  as the information sample gets large. Thus, if a participant only enters the attention stage with an intention of waiting for a signal that changes the posterior, the gross attention benefit of any decision problem A is drastically simplified to:

$$\begin{aligned} G(A, \pi_l) &= u(P) \text{ if } \pi_l \neq \pi_0 \\ G(A, \pi_0) &= \frac{1}{2}u(P) \end{aligned}$$

If we consider  $\pi_0$  (guessing) as the default action, then the gross *marginal* benefit of paying attention in this context is:

$$G^*(A, \pi_l) = \frac{1}{2}u(P) \quad (\text{A.4})$$

A DM has a *costly information representation* [Caplin and Dean, 2015] if they both (i) make ex ante optimal choices of attention structure given expected costs and benefits ( $\pi_A \in \arg \max_{\pi \in \Pi} \{G(A, \pi) - K(\pi)\}$ ), and (ii) their chosen action  $a$  is optimal given their observed information ( $\sum_{\omega \in \{R, B\}} \gamma(\omega)u(a(\omega)) \geq \sum_{\omega \in \{R, B\}} \gamma(\omega)u(b(\omega))$  for all  $b \in A$ ). If a participant behaves in this way, any ‘mistake’ must be case of optimal inattention, because (i) rules out suboptimal inattention and (ii) rules out ineffective attention. These conditions are difficult to test directly because the attention cost  $K()$  is unobserved. However in this experiment context these conditions can be simplified to those used in the main text:

$$a = g \iff \gamma(R) = \frac{1}{2} \tag{Oa}$$

$$a = r \iff \gamma(R) = 1$$

$$a = b \iff \gamma(R) = 0$$

$$I_{ple} \leq I_{p*le} \quad \forall p, p* \in \mathbb{P} \quad s.t. \quad p < p*; \quad \forall l \in \mathbb{L} \quad \forall e \in \{\text{auto}, \text{manual}\} \tag{O\pi1}$$

$$I_{ple} \geq I_{pl*e} \quad \forall l, l* \in \mathbb{L} \quad s.t. \quad l < l*; \quad \forall e \in \{\text{auto}, \text{manual}\} \tag{O\pi2}$$

## A.2 Appendix: Additional Screenshots

### Filter type B Beliefs

#### Beliefs: Filter type B only

With each opportunity to use Filter type B, you will be asked to state your belief about the probability you will make a mistake with the data filter.

A mistake occurs when there is enough information to make a correct prediction but you make the wrong prediction.

You will be asked to input your belief about your mistake probability using a slider:

#### Mistake Probability

What do you believe is the probability you make a mistake with Filter 4:



Over 50% (guess) Next

The number you input as your mistake probability can affect the probability you win a prize.

The prize probability is set up so that you **maximize your expected probability of a prize if you input the truth.**

You may input any number between 0% and 49% as your belief about your own probability of making a mistake.

If you think your mistake probability is 50% or greater you can also indicate this, but then the experiment will insist you then Guess with 51% chance of being correct.


The following explains the detail of the belief payoff mechanism. **It is only necessary to understand that it is always in your best interest to tell the truth** as this maximizes the probability you win a prize.

Belief Payoff Details [Click to expand/collapse](#)

Next

### Round 2: Data Filter 4 type B (Practice): Mistake Probability

What do you believe is the probability you make a mistake with Filter 4?



Next

50% or more Clicking this button will Guess Now and be correct with 51% probability.

## Final Results

Thank you for your participation. Please review the results below. You earned a baseline of \$7.00CAD just for participating.

You answered 4 quiz questions correctly for a payoff of \$2.00CAD.

Review the table of your predictions and prizes below.

Round	Round Type	Prize if Correct	Previous Terms	Next Term	Your Choice	Prediction	Prize Won	Type M Payment Method	Mistake Probability	t/100 (lottery entry)	x/100 (lottery win)
1	P	0.5	R R B R	R	Guess	B	0	na	na	0.07	0.10
2	P	3.25	B B R R	R	Predict Blue	B	0	na	0.15	0.72	0.44
3	M	3	B B R R R	R	Predict Blue	B	0	Prediction - loss	0.15	0.37	0.24
4	M	3.75	R R B	B	Predict Blue	B	3.75	Prediction - win	0.2	0.25	0.52
5	M	1	R B B B R B	B	Guess	B	1	Lottery - win	0.25	0.96	0.41
6	M	2	R R B B	B	Predict Blue	B	2	Prediction - win	0.25	0.19	0.61
7	M	0.25	B B R R R R B	B	Guess	B	0.25	Lottery - win	0.2	0.91	0.62
8	A	1.25	B R R B B	R	Guess	B	0	na	na	0.75	0.27
9	A	0.5	B R B	R	Predict Red	R	0.5	na	na	0.86	0.78
10	A	2.25	B B R B R R	R	Predict Red	R	2.25	na	na	0.91	0.44
11	A	0.5	B R B R	B	Predict Blue	B	0.5	na	na	0.92	0.61
12	A	2.25	B R R B B B R	B	Predict Blue	B	2.25	na	na	0.81	0.70

Your total payoff of \$21.50CAD will be emailed to you by 2021-May-29.

## A.3 Appendix: Experiment Instructions

Participants sign on to a live zoom meeting to complete ID check, consent form, and read the instructions aloud and take questions. The meeting ends after the instructions and they have until 23:59 two days ahead to complete the experiment. They can exit and return with saved progress using a secure link through generated by oTree [Chen et al., 2016].

## Instructions



This experiment is conducted using your own internet-connected device. You earn a participation fee of \$7.00 CAD today.






You can sign in and complete the experiment at any time from 00:01 2021-May-25 to 23:59 2021-May-27. If you exit the browser, you can sign back in with the same link and your progress is saved.

It is possible to earn between \$0.00 CAD and \$53.00 CAD in additional prizes based on your decisions and an element of chance.



Any additional prizes you earn during this time will be paid by email transfer along with your \$7.00 participation fee on 2021-May-29.

### Primary Task

To earn additional prizes in this experiment, you must correctly predict whether the next term in a sequence is B = Blue Triangle  or R = Red Circle .

In each round you will be given one specific sequence - for example:    - and asked whether  or  comes next.

Your choices are to (1) Guess, (2) Predict , or (3) Predict .






For any sequence you are given, the next term is randomly selected to be  with 50% probability or  with 50% probability. This randomization occurs independently for each participant, round, and sequence.

If you 'Guess', the computer will choose on your behalf and will be correct with 51% probability, so it can be better to choose Guess than predict randomly.













As an alternative to a Guess, you will be given an opportunity to collect information about the sequence using a tool called a **data filter**.






Every 10 seconds, a data filter randomly draws a sequence and displays the single term which followed, just like your prediction task.

### Example

You will win a monetary prize if you can predict whether the term following    is  or .

You choose to spend time using data filter 3 instead of guessing immediately.

Every 10 seconds, data filter 3 randomly draws one of the possible sequences made of 3  or  terms - such as (  ) , ( ), (  ) , etc. - and displays the randomly selected  or  term that follows.

Eventually, data filter 3 will draw the same 3 term sequence as you are given in your prediction problem (  ), and display whether the next term is  or .






Below is a screenshot of data filter 3 recording a draw of    followed by .

(type M' will be explained on the next page)

## Round 8: Data Filter 3 type M

Filter Type M: You must make a Prediction or 'Guess'

Each sequence is followed by  with 50% probability or  with 50% probability.

You must predict if the next term in the sequence following:     
is  or 

This data filter randomly draws a sequence and displays the following term every 10 seconds.  
You can add these draws to a table below to help you form a prediction.

The current draw is:



























   

You can 'Add to Table' a max of one time per draw (10 seconds).

The table display will only add observations if you click 'Add to Table'.


You have 10 seconds before an observation disappears.


Observation Table

previous terms	followed by 	followed by 
  	0	3
  	0	4
  	1	0
  	0	0
  	0	0
  	1	0
  	2	0
  	1	0






Make your prediction or guess whenever you are ready.






Guess (correct with 51% probability)









Predict 

Predict 

Every 10 seconds filter 3 selects one of the sequences of length 3 and displays the term that follows. You can store this information in the observation table by clicking 'Add to Table'.

The first row of the observation table shows that    has been followed by  zero (0) times, and has been followed by  three (3) times.

The second row of the observation table shows that    has been followed by  zero (0) times, and has been followed by  four (4) times.

In this example round, you are being paid to predict what follows   , so you will be most interested in the row of the table which says that   , has been followed by  one (1) time, and has been followed by  zero (0) times.

## Instructions

This experiment has 12 total rounds: Two practice rounds, which are not paid, followed by ten paid rounds.

There is one practice round with filter type A and one practice round with filter type M.

You then play a set of five paid rounds with filter type A and a set of five paid rounds with filter type M. The order of these two sets is random for each participant.

### Filter Types A and M

Filter type A (A for 'Automatic') has the computer make a correct prediction on your behalf as soon as you collect enough data.

Filter type M (M for 'Manual') has you enter your own prediction, you can use the filter for as long as you wish before making a prediction.

### Example

See Filter type A and Filter type M below.

Filter type A does not have a button for 'Predict R' or 'Predict B'.

In both filters you can click 'Add to Table' to add observations to the table.






In both filters you can 'Guess' at any time to proceed to the next round.



## Round 3: Data Filter 3 type A

Filter Type A: Computer will make a Prediction using the data you 'Add to Table', or you can 'Guess'.

Each sequence is followed by  with 50% probability or  with 50% probability.

You must predict if the next term following:    is  or 

This data filter randomly draws a sequence and displays the following term every 10 seconds.  
You can add these draws to a table below to help you form a prediction.

The current draw is:





























Add to Table

You can 'Add to Table' a max of one time per draw (10 seconds).

The table display will only add observations if you click 'Add to Table'.

You have 10 seconds before an observation disappears.

Observation Table

previous terms	followed by 	followed by 
  	0	0
  	1	0
  	0	0
  	0	0
  	0	1
  	0	0
  	1	0
  	0	0

### Filter Type A

does not allow you to input your own prediction.

You can 'Add To Table' until the computer makes a prediction or you can 'Guess' to proceed.

Make your prediction or guess whenever you are ready.

**Guess** Guess (correct with 51% probability)

Note: With Filter type A, you cannot make a specific prediction.




The "Predict" buttons are disabled but the "Guess" button is active.



Use "Add to Table" button to collect enough observations for the computer to make a prediction on your behalf.

## Round 8: Data Filter 3 type M

Filter Type M: You must make a Prediction or 'Guess'

Each sequence is followed by  with 50% probability or  with 50% probability.

You must predict if the next term in the sequence following:   

is  or 

This data filter randomly draws a sequence and displays the following term every 10 seconds.

You can add these draws to a table below to help you form a prediction.

The current draw is:



























Add to Table

You can 'Add to Table' a max of one time per draw (10 seconds).

The table display will only add observations if you click 'Add to Table'.

You have 10 seconds before an observation disappears.

Observation Table

previous terms	followed by 	followed by 
  	0	3
  	0	4
  	1	0
  	0	0
  	0	0
  	1	0
  	2	0
  	1	0


### Filter Type M


requires you to input your own prediction using one of three buttons.

You can 'Add To Table', 'Guess', or Make a Prediction at any time.

Make your prediction or guess whenever you are ready.

Guess Guess (correct with 51% probability)

Predict 'B' Predict 

Predict 'R' Predict 

# Instructions


## Filter A: Automatic Prediction

Filter type A will alert you as soon as you have added enough observations to the table for the computer to make a correct prediction.

See Filter 3 type A before and after the prediction alert in the two screenshots below.

### Round 3: Data Filter 3 type A

Filter Type A: Computer will make a Prediction using the data you 'Add to Table', or you can 'Guess'.

Each sequence is followed by  with 50% probability or  with 50% probability.  
You must predict if the next term following:    is  or .

This data filter randomly draws a sequence and displays the following term every 10 seconds.  
You can add these draws to a table below to help you form a prediction.

The current draw is:



























   

You can 'Add to Table' a max of one time per draw (10 seconds).

The table display will only add observations if you click 'Add to Table'.

You have 10 seconds before an observation disappears.

Observation Table

previous terms	followed by 	followed by 
  	0	0
  	1	0
  	0	0
  	0	0
  	0	1
  	0	0
  	1	0
  	0	0

Make your prediction or guess whenever you are ready.

Guess (correct with 51% probability)

Note: With Filter type A, you cannot make a specific prediction.

The 'Predict' buttons are disabled but the 'Guess' button is active.

Use 'Add to Table' button to collect enough observations for the computer to make a prediction on your behalf.

Note the alert generated when the observation table adds **R R R R**, followed by **R**:

## Round 3:

Filter Type A: Com

The computer has enough data and has made a Prediction. The computer predicts the next term is: **R**

OK

Each sequence is followed by **B** with 50% probability or **R** with 50% probability.

You must predict if the next term following: **R R R R** is **B** or **R**

This data filter randomly draws a sequence and displays the following term every 10 seconds. You can add these draws to a table below to help you form a prediction.

The current draw is:

**R R R R**

Add to Table

You can 'Add to Table' a max of one time per draw (10 seconds).

The table display will only add observations if you click 'Add to Table'.

You have 10 seconds before an observation disappears.

Observation Table

previous terms	followed by <b>B</b>	followed by <b>R</b>
<b>B B B</b>	0	0
<b>B B R</b>	1	0
<b>B R B</b>	0	0
<b>R B R</b>	0	0
<b>B R R</b>	0	1
<b>R R R</b>	0	1
<b>R R B</b>	1	0
<b>R B B</b>	0	0

Data Filter A will alert you when you have enough observations in the table for the computer to make its prediction.

Click 'OK' to proceed to the next Round.

**Make your prediction or guess whenever you are ready.**

Guess Guess (correct with 51% probability)

Note: With Filter type A, you cannot make a specific prediction.

The 'Predict' buttons are disabled but the 'Guess' button is active.

### Time-Prize Tradeoff

Filters require time to make predictions, but Guessing takes no time.

In each round you will be told the expected time it will take for a filter to reach a prediction, and indicate whether you want to 'Guess Now' or 'Use Filter' for prize increments ranging between \$0.25 and \$5.00.

The expected time required changes in each round and is published at the top of the decision list for that round.

You must decide ahead of time what you would do for each prize level.

After you submit your choice for each prize, one prize will be randomly selected as the 'choice that counts', and that is the prize value you play for in that round.

Below is a screenshot of the list of choices. Note the average time in this round of 80 seconds to draw the sequence of terms that matches your prediction task:

### Round 3: Data Filter 3 type A Choices

Choose an action for each possible prize level: either let the computer Guess, which is correct with probability 51%, or use a data filter to observe the sequence over time.

Any prediction provided by the filter in this round will be 100% correct.

Filter 3 takes an average of 80 seconds (or 01:20 minsec) to draw the sequence of terms that matches your prediction task.

Prize = \$0.25	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$0.50	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$0.75	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$1.00	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$1.25	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$1.50	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$1.75	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$2.00	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$2.25	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$2.50	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$2.75	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$3.00	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$3.25	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$3.50	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter
Prize = \$3.75	<input type="radio"/> Guess Now	<input type="radio"/> Use Filter

The filter available to you will change by round.

The average time a filter requires doubles with each increase in Filter number. So Filter 4 requires twice as much time as Filter 3, and Filter 5 requires twice as much time as Filter 4 (on average).

### Quiz

Following the two practice rounds, there will be a 6 question multiple choice quiz to ensure you understand the experiment. Each quiz question you answer correctly on the first try is worth an additional \$0.50 CAD. Questions answered incorrectly on the first try are not paid but you are encouraged to review the answers and short instructions prior to the paid rounds.

### Total Payment

You face a total possible payoff between \$7 and \$60 based on your choices and the randomly selected prizes in each list.

You earn \$7 for participating regardless of the outcome of the experiment.

You can earn up to \$3 in the Quiz that follows the Practice Round (\$0.50 per correct answer).

In addition to the participation fee and quiz payment, there are 10 paid rounds, 5 with Filter type A and 5 with Filter type M.

For the 5 rounds with Filter type A, the minimum payment is \$0, maximum payment is \$25, and the expected payment when always choosing Guess is \$6.70.

For the 5 rounds with Filter type M, the minimum payment is also \$0, and the maximum payment is also \$25.

When you use Filter M, you will be asked for your belief about your personal probability of making a mistake with the filter. Your probability of winning a prize with Filter type M is affected by your mistakes and your belief. This is explained in more detail when you first use Filter type M.

### Payment Method

Your payment will be electronically transferred to the email you used to sign up for the experiment, on 2021-May-29. Contact econexp@sfu.ca if you have any issues receiving payment.

When you click 'Next', the practice rounds will begin.

Next

## Appendix B

# Supplements to Chapter 2

### B.1 Additional Tables of Results

Effort Schedule	First Day Choice, Second Day Choice		
	Time Invariant	Today, Not Today	Not Today, Today
14, 20, 28	88%	3%	10%
16, 20, 25	86%	7%	7%
18, 20, 22	82%	9%	9%
19, 20, 21	86%	8%	6%
20, 20, 20	71%	13%	17%
22, 20, 18	71%	8%	20%
25, 20, 16	77%	1%	21%
Grand Total	79%	8%	13%

Table B.1: Time Invariance by Effort Schedule  
*Non-endogenous subsample*

TWR only Effort Schedules	Time Consistent			Reversal		Non- Strotz	Censored	Quads Observed
	day1	day2	day3	earlier	later			
14, 20, 28	83%	0%	0%	8%	4%	4%	0%	24
16, 20, 25	77%	2%	0%	8%	2%	2%	10%	52
18, 20, 22	73%	4%	2%	6%	0%	2%	13%	52
19, 20, 21	75%	8%	0%	0%	0%	0%	17%	12
20, 20, 20	67%	2%	6%	10%	0%	0%	15%	52
22, 20, 18	50%	0%	18%	11%	0%	4%	18%	28
25, 20, 16	57%	0%	18%	7%	0%	4%	14%	28
TOTAL	69%	2%	6%	8%	1%	2%	13%	248
TOTAL*	79%	2%	6%	9%	1%	2%		217

\*excluding censored

Table B.2: Classifying choices in TWR quads

*Non-endogenous subsample*

*Only TWR quads included in table for comparison to Table 2.3*

## B.2 Structural Model Parameter Estimates and Standard Errors:

We estimate the logistic regression:

$$y_t = b_0 e_t + b_1 I_{k=1} e_{t+k} + b_2 I_{k=2} e_{t+k} + \epsilon_t$$

For our structural model, we are interested in the values of  $\frac{b_1^2}{b_0 b_2}$  ( $= \beta$ ),  $\frac{b_2}{b_1}$  ( $= \delta$ ), and  $b_0$  ( $= \lambda$ )

If  $b \sim N(b^*, \Sigma)$  then the distribution of  $f(b)$  is  $N(f(b^*), C\Sigma C')$  where  $C = \nabla f(b)$ .

$$\text{Let } f(b) = \begin{bmatrix} \frac{b_1^2}{b_0 b_2} \\ \frac{b_2}{b_1} \\ b_0 \end{bmatrix}. \text{ Then, } C = \nabla f(b) = \begin{bmatrix} -\frac{b_1^2}{b_0^2 b_2} & \frac{2b_1}{b_0 b_2} & -\frac{b_1^2}{b_0 b_2^2} \\ 0 & -\frac{b_2}{b_1^2} & \frac{1}{b_1} \\ 1 & 0 & 0 \end{bmatrix}.$$

The estimated variance matrix of interest is  $C\hat{\Sigma}C'$



Table B.3: Structural Logit Estimates

	<i>Dependent variable:</i>	
	Today1	
EffortToday	-0.140*** (0.042)	
EffortNotTodayK1	0.199*** (0.041)	<i>Non-endogenous subsample</i>
EffortNotTodayK2	0.185*** (0.043)	
Observations	743	
Log Likelihood	-358.575	
Akaike Inf. Crit.	723.150	
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Table B.4: Variance-Covariance Matrix of Structural Logit Estimates

	EffortToday	EffortNotTodayK1	EffortNotTodayK2
EffortToday	0.001803	-0.001716	-0.001747
EffortNotTodayK1	-0.001716	0.001716	0.001725
EffortNotTodayK2	-0.001747	0.001725	0.001826

Table B.5: Variance-Covariance Matrix of Parameter Estimates using Delta Method

	$\beta$	$\delta$	$\lambda$
$\beta$	0.046933	0.070261	0.007851
$\delta$	0.070261	0.163917	0.016764
$\lambda$	0.007851	0.016764	0.001803

### B.3 Data Generation and Classification

Quad Type	$c(\{(e_1, 1), (e_2, 2)\})$	$c(\{(e_1, 1), (e_3, 3)\})$	$c(\{(e_2, 2), (e_3, 3)\})$	$c(\{(e_1, 1), (e_2, 2), (e_3, 3)\})$	Detail
Time Consistent	1	1	0	1	day 1
	1	1	2	1	day 1
	1	1	3	1	day 1
	1	3	3	3	day 3
	2	1	2	2	day 2
	2	3	2	2	day 2
	2	3	3	3	day 3
Reversal	1	3	0	1	earlier
	1	3	2	1	earlier
	1	3	3	1	earlier
	2	1	0	1	earlier
	2	1	2	1	earlier
	2	1	3	1	earlier
	1	3	2	2	later
	2	1	3	3	later
Non-Strotzian	1	1	2	2	pref flex
	1	1	3	3	pref flex
	2	3	0	1	pref commit
	2	3	2	1	pref commit
	2	3	3	1	pref commit
Censored	1	1	0	0	censored
	1	3	0	0	censored
	2	1	0	0	censored
	2	3	0	0	censored

Table B.6: Classifying Observable Choice Quads

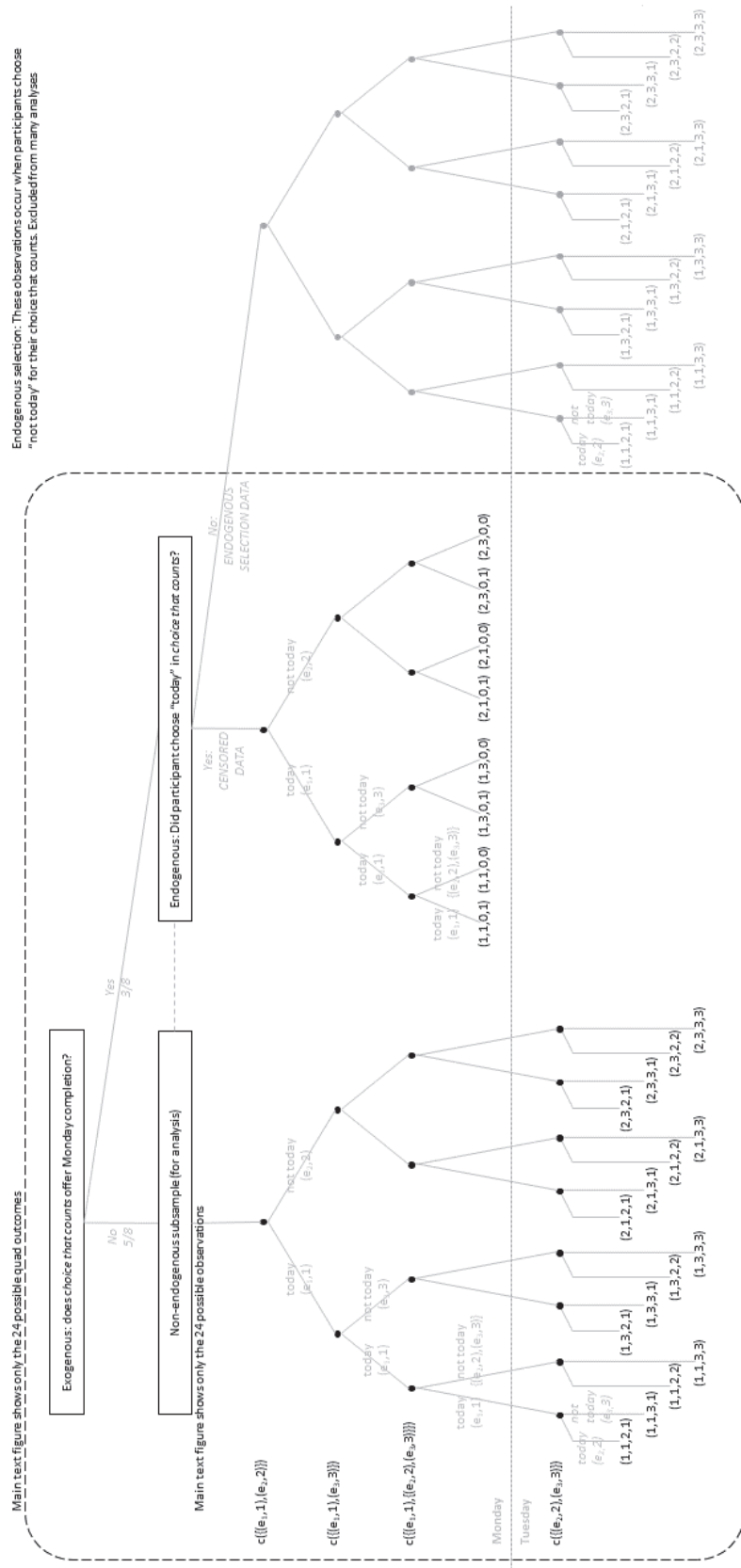


Figure B.1: Complete experiment data generating process

*This figure extends Figure 2.1 from the main text to include the repetitive portion of the decision tree for the endogenous sample. All tables in main text of Chapter 2 except Table 2.2 and Figure 2.2 include only the non-endogenous subsample (the left sub-tree)*

## Appendix C

# Supplement to Chapter 3

### C.1 Correlation of Well-being Measures and Covariates

## Correlation coefficients

TREATMENT	POST	Loneliness	SPAN_E_Pos	SPAN_E_Neg	SWLS	Flourishing	Anxiety	SocialConnection	CESD	BelongsFU	eGPA	FEMALE
TREATMENT	1.00	0.00	0.12	0.01	-0.04	0.07	0.02	-0.08	0.04	-0.12	-0.23	-0.16
POST	0.00	1.00	0.04	0.11	-0.08	-0.11	-0.01	0.03	-0.01	-0.09	-0.02	0.04
Loneliness	0.02	0.04	1.00	0.19	-0.29	-0.39	0.09	0.38	0.19	-0.44	-0.01	0.09
SPAN_E_Pos	0.12	0.05	-0.19	0.19	0.51	0.53	-0.42	-0.30	-0.33	0.10	0.11	0.06
SPAN_E_Neg	0.01	0.11	-0.38	1.00	-0.36	-0.23	0.65	0.29	0.63	-0.14	0.00	0.31
SWLS	-0.04	-0.08	-0.29	-0.36	1.00	0.62	-0.38	-0.40	-0.40	0.15	0.35	-0.06
Flourishing	0.07	-0.11	-0.39	-0.23	0.62	1.00	0.00	-0.50	-0.06	0.19	0.01	0.18
Anxiety	0.02	-0.01	0.09	-0.42	-0.38	0.00	1.00	0.26	0.80	-0.10	-0.06	0.34
SocialConnection	-0.08	0.03	0.38	0.29	-0.40	-0.50	0.26	1.00	0.32	-0.34	0.14	0.31
CESD	0.04	-0.01	0.19	-0.33	0.63	-0.06	0.80	0.32	1.00	-0.33	-0.10	0.38
BelongsFU	-0.12	-0.09	-0.44	0.10	-0.14	0.19	-0.10	-0.34	-0.33	1.00	0.14	-0.15
eGPA	-0.23	-0.02	0.11	0.00	0.35	0.01	-0.06	0.14	-0.10	0.14	1.00	0.32
FEMALE	-0.16	0.04	0.09	0.31	-0.06	0.18	0.34	0.31	0.38	-0.15	0.32	1.00

## P-values of test: is correlation coefficient = 0 ?

TREATMENT	POST	Loneliness	SPAN_E_Pos	SPAN_E_Neg	SWLS	Flourishing	Anxiety	SocialConnection	CESD	BelongsFU	eGPA	FEMALE
TREATMENT	1.0000	0.8492	0.3332	0.9111	0.7391	0.5389	0.8621	0.5090	0.7217	0.2972	0.0559	0.1754
POST	0.7451	1.0000	0.5607	0.3427	0.5169	0.3764	0.9511	0.7842	0.9349	0.4537	0.8694	0.7633
Loneliness	0.7451	0.6607	0.1170	0.1013	0.0135	0.0008	0.4606	0.0009	0.1047	0.0001	0.9116	0.4382
SPAN_E_Pos	0.3332	0.3427	1.0000	0.0009	0.0000	0.0000	0.0003	0.0104	0.0044	0.3931	0.3654	0.6269
SPAN_E_Neg	0.9111	0.5169	0.1013	0.0018	0.0018	0.0568	0.0000	0.0129	0.0000	0.2529	0.9721	0.0076
SWLS	0.7391	0.5169	0.0135	0.0000	0.0000	0.0000	0.0009	0.0005	0.0004	0.2082	0.0024	0.6266
Flourishing	0.5389	0.3764	0.0008	0.0000	0.0000	0.9778	0.9778	0.0000	0.6385	0.1025	0.9306	0.1379
Anxiety	0.8621	0.9511	0.4606	0.0000	0.0000	0.0000	0.0000	0.0279	0.0000	0.3905	0.6120	0.0033
SocialConnection	0.5090	0.7842	0.0009	0.0104	0.0005	0.0000	0.0000	0.0069	0.0069	0.0038	0.2553	0.0075
CESD	0.7217	0.9349	0.1047	0.0044	0.0004	0.6385	0.0000	0.0048	0.0048	0.0048	0.3840	0.0010
BelongsFU	0.2972	0.4537	0.0001	0.3931	0.2529	0.1025	0.3905	0.0038	0.0048	0.2435	0.2435	0.0067
eGPA	0.0559	0.8694	0.9116	0.3654	0.9721	0.0024	0.9306	0.6120	0.3840	0.2435	0.2435	0.0067
FEMALE	0.1754	0.7633	0.4382	0.6269	0.0076	0.6266	0.1379	0.0033	0.0010	0.2176	0.2176	0.0067

Figure C.1: Correlation of well-being measures and covariates

The top matrix of Figure C.1 provides the raw correlation coefficient between measurements. The bottom matrix provides the  $p$ -value of a  $t$ -test of the null hypothesis that the correlation is zero. Stars are omitted,  $p$ -values closer to zero suggest statistically stronger correlation. The data in these tables is restricted to sample of individuals measured twice with complete responses.