From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Psychology

2013

The Role of Delayed Consequences in Human Decision-Making

White, John Myles

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01v118rd619

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

The Role of Delayed Consequences in Human Decision-Making

John Myles White

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Psychology
Adviser: Jonathan D. Cohen

June 2013

© Copyright by John Myles White, 2013.

All Rights Reserved

Abstract

People make many decisions with consequences that are delayed, rather than immediate. Of particular interest are decisions in which long-term gains must be balanced against short-term costs. Such time trade-offs can be advantageous or deleterious to the decision-maker: the decision to abstain from immediately entering the labor force and instead pursuing a lengthy education benefits the educated in the long-term although their short-term wages are lowered. In contrast, the decision to overeat increases the short-term enjoyment of food but decreases long-term health. A large body of research in psychology has shown that the ability to delay gratification and elect long-term over short-term gains leads to superior life outcomes.

Expanding on this tradition, my thesis examines time-tradeoffs in two domains: first, I examine the resolution of time-tradeoffs in settings in which people are asked to explicitly decide between short-term and long-term gains. This line of work is closely connected to economic models of decision-making that account for the role of time in shaping decisions. I then transition to examining the resolution of time-tradeoffs in settings in which time trade-offs are implicit. Specifically, I examine the way in which people explore unfamiliar environments in order to maximize information. Maximizing information represents a time-tradeoff because the goal of obtaining information generally requires the decision-maker to eschew known sources of short-term rewards in order to explore new options whose benefits will be reaped only in the long-term.

Collectively, I describe a large body of experiments that examine these two classes of decision-making and put forward two new models of decision-making, the ITCH model of intertemporal choice and the MaxInfo model of exploratory decision-making, that account for the data from these experiments and extend the state of the art.

Acknowledgements

This dissertation would not exist without the mentoring I've received from my advisor, Jon Cohen. Jon's willingness to forgive my frequent digressions and his encouragement of my academic interests, even when they strayed outside of traditional psychological research, is the primary reason that this dissertation exists. I am also grateful to many other members of the faculty at Princeton who helped me throughout the past five years, including Yael Niv, Dan Osherson, Nick Turk-Browne, Phil Holmes, Naomi Leonard and Dave Blei. Much of this work was done as part of a long-standing (and long distance) collaboration with two economists, David Laibson and Keith Ericson, who have had an enormous impact on my work. My initial thinking about intertemporal choice was also heavily influenced by our lab's collaboration with George Loewenstein.

Beyond these faculty advisors, I am endebted to all of my collaborators and coworkers. In particular, the work I have done on exploration and exploitation would not exist without Bob Wilson and Paul Reverdy. Outside of formal collaborations, I've benefited greatly from working alongside Sam Gershman, Mike Todd, Elliot Ludvig, Andra Geana, Amitai Shenhav, Damon Tomlin and many others.

The development of the ITCH model in particular benefitted from an early comment from Chris Wiggins, who originally inspired me to think about the role of purely numeric heuristics in shaping human decisions in unfamiliar environments.

Of course, my last and largest thanks go to my parents and friends, without whose support this research would not have been completed.

Contents

Abstract iii

Acknowledgements iv

1 Introduction **1**

1.1 Impatience and the Theory of Intertemporal Choice 3

1.2 Curiosity and the Role of Information in Decision-Making 6

2 Decision Theory **8**

2.1 Core Concepts: Decision-Makers, Goods, Actions 9

2.2 Neoclassical Utility Theory 10

2.3 Classical Utility Theory 13

2.4 Expected Utility Theory 14

2.5 Subjective Expected Utility Theory 17

2.6 Random Utility Theory 18

2.7 Discounted Utility Theory 20

3 Intertemporal Choice Theory **21**

3.1 Synopsis of DU Theory 21

3.2 DU Theory in Detail 24

3.2.1 DU Theory in a Binary Choice Problem 24

3.3 Invariances of DU Theory 28

3.3.1 Dynamic Consistency 29

3.3.2 Domain Invariance 32
3.3.3 Magnitude Invariance 33
3.3.4 Partition Invariance 35
3.4 Models of the Discount Function 35
3.5 Probabilistic Extensions of DU Theory 37

4 Intertemporal Choice Experiments **39**
4.1 The Money Earlier or Later (MEL) Task 40
4.2 A MEL Task on a Representative Sample of Americans 47
4.3 The Banking Study 51
4.4 The IRA Study 56
4.5 The Relative and Absolute Gains Study 62
4.6 The First Time Frame Study 66
4.7 The Second Time Frame Study 70
4.8 The External Validity of MEL Tasks 72
4.8.1 The BIS/BAS Study 73
4.8.2 The R^2 Study 74

5 ITCH: A New Model of Intertemporal Choice **86**
5.1 Derivation of the ITCH Model 87
5.2 Extensions of the ITCH Model 91
5.3 The Empirical Performance of ITCH Model 92
5.4 The Gains vs. Losses Study 93
5.5 Interpreting ITCH 96
5.6 Model Fitting and Comparison Procedures 98

6 Theory of Exploration and Exploitation **107**
6.1 The Exploration-Exploitation Dilemma 107

7 MaxInfo: A New Model of Exploration and Exploitation 114

7.1 The Multi-Armed Bandit (MAB) Problem 115

7.2 The Inference Process 117

7.3 The Decision Process 121

7.4 Inference and Decision Combined 125

8 Experiments on Exploration and Exploitation 127

8.1 Designing Exploration-Exploitation Experiments 127

8.2 The Effects of Manipulating Horizons 133

8.3 The Effects of Manipulating Incentives 141

8.4 The Design of Many-Bandit Tasks 149

8.5 The Grid Task Paradigm 153

8.6 The Effects of Manipulating Structure 157

8.7 The Effects of Manipulating Incentives 160

8.8 Interpreting the MaxInfo Model 162

9 Conclusion 169

Chapter 1

Introduction

Human beings make many decisions in which some consequences are immediate and other consequences are delayed (Mischel, Shoda, and Rodriguez, 1992). Psychologists and economists have come to refer to such decisions as *intertemporal choices* in order to emphasize the central role that the specific timing of consequences plays in shaping decision-making (Frederick, Loewenstein, and O'Donoghue, 2002; G. Loewenstein, 2003).

Intertemporal choices are of particular interest to scholars of decision-making because they allow researchers to explore situations in which a decision-maker's immediate goals conflict with the decision-maker's long-term goals (Whittle, 1980). For example, an economist would say that a smoker's decision to smoke a cigarette now *reveals the smoker's preference* for immediately relieving nicotine withdrawal at the expense of the smoker's long-term health. Because a revealed preference such as this one depends upon the specific timing of the two enumerated consequences, it is traditional to refer to such preferences as *time preferences* (G. Loewenstein, 2003).

Time preferences are a very popular research topic in both psychology and economics (Berns, Laibson, and Loewenstein, 2007) because our example of a smoker's decision to forgo long-term gains in order to obtain short-term gains seems to reflect

a much broader concern in decision-making: many, if not almost all, problems of self-control seem to involve a decision-maker making decisions in which maximizing immediate gains is unduly prioritized over long-term gains. Indeed, related conflicts seem to be central to our understanding of issues ranging from the personal problem of drug addiction to the global problem of making plans for dealing with climate change (Baker, Johnson, and Bickel, 2003; Stern et al., 2007). Because so many of the problems that we grapple with as individuals and societies seem to involve this general conflict between the short-term and the long-term, the theoretical analysis of intertemporal choice has come to occupy a central place in theories of decision-making.

But the conflict between immediate gains and long-term gains is far more broadly applicable than is suggested by consideration of problems of self-regulation or lack of forethought. The conflict between the short-term and the long-term also plays a central role in shaping strategies for exploring uncertain environments. When we decide whether to stick with our old favorites or to voyage out into the unknown, we are making a particular form of intertemporal choice in which the gains from exploration can only be obtained in the long-term, whereas the potential costs of exploration are always immediate (Whittle, 1980).

This thesis will therefore describe two distinct lines of research that are united by a shared focus on the conflicts that can arise between the short-term and the long-term. In the first half of this thesis, we will focus upon the problem of **impatience**, in which we seek to understand why humans make decisions that seem to place *excessive* emphasis on the short-term (Koopmans, 1960). I will review the core theoretical tools for reasoning about intertemporal choice and will describe both experiments and modeling work that collectively raise substantial doubts about the validity of traditional methods for studying impatience. In the second half of this thesis, I will focus on the trade-offs that must be made when deciding whether to explore an

unfamiliar environment or to stay in a familiar one. Whereas the first half of this thesis is concerned with impatience, the second half can be viewed as an examination of the normative reasons for human beings to exhibit **curiosity** and an assessment of the extent to which humans exhibit a normatively defensible assessment of the value of information (G. Loewenstein, 1994; Raiffa, 1968; Behrens, Woolrich, Walton, and Rushworth, 2007).

1.1 Impatience and the Theory of Intertemporal Choice

In Chapters 3-5, we will focus upon the kinds of decisions that people make when forced to explicitly choose between short-term gains and long-term gains. By observing the patterns of choices made by experimental participants in different settings, we will try to understand the *qualitative* strategies that humans employ when reasoning about trade-offs between the short-term and long-term.

I will then describe a formal model of the *quantitative* patterns of choice that we can expect to observe when human beings are explicitly asked to choose between a smaller amount of money that is available in the short-term and a larger amount of money that is available in the long-term. I will refer to the specific use of questions involving time-dated monetary goods as the *Money Earlier or Later* paradigm – or MEL for short (Frederick et al., 2002). The MEL paradigm will provide a tightly focused setting that I will use to articulate the core aspects of the theory of intertemporal choice and to test the predictions of the primary theory of intertemporal choice, DU Theory (Samuelson, 1937).

As we will see, the theory of intertemporal choice can be understood with reference to a single theoretical question: "how does the value of a good change as its date of receipt is delayed into the future?" I will approach this question from the standpoint

pioneered by microeconomists, who have conceptualized this question in terms of a **discount function**, a mathematical construct that specifies, quantitatively, how much value a good will have if it is going to be received at time t.

By framing our discussion in terms of discount functions, I will be able to articulate several important psychological questions in very precise terms. In particular, we will be interested in the following questions:

- Can we articulate the differences between people in terms of their capacity to delay gratification? Do individuals have a generalized level of impatience that characterizes their behavior and influences the trajectory of their life? If so, how can we measure these differences?

- How do we define the notion of self-control? Many people seem to feel that they will not follow through on their own plans because of the temptations posed by immediate gains. As such, they choose to precommit to behaviors that benefit the long-term at the expense of the short-term. Can we understand this propensity for precommitment in terms of the qualitative properties of a hypothetical discount function?

I will discuss both of these issues in some depth. The issues of self-control and precommitment lie at the heart of the theoretical literature on intertemporal choice (Frederick et al., 2002; O'Donoghue and Rabin, 2001). Indeed, it is possible to make the concepts of self-control and precommitment rigorous by making assumptions about the functional form of the discount function (Laibson, 1997; Prelec, 2004). I will therefore review the literature on such functional forms in order to explain why many researchers have come to believe that mathematical models of intertemporal choice provide insight into problems of self-regulation – and why others have not (e.g. Gul and Pesendorfer, 2003).

The relationship between theoretical discount functions and observable behaviors

will form the core content of this first half. As we will see, it is unclear whether researchers can plausibly measure a discount function using the kinds of experiments that have been traditionally employed in the literature. In particular, it will become clear that the MEL paradigm conventionally used for measuring discount functions is largely inadmissible as a quantitative measure of the discount function, although it may still be the case that qualitative patterns of choice can be uncovered using the MEL paradigm. On this second matter, my work is partially silent, although two experiments that I will discuss will suggest that the relationship between behavior in MEL behaviors and impatient behaviors in the field is both real and very weak.

To make sense of the failures of MEL to measure discount functions, I will develop a non-standard model of intertemporal decision-making that is not predicated upon the assumptions of a coherent discount function. Instead, I will assume that decision-making is primarily driven by surface features of the descriptions of time-dated goods. This stands in sharp contrast to standard economic theories, which are based on the underlying properties of goods and should not be affected by any incidental details of their verbal presentation (Tversky, Kahneman, et al., 1981). I will make this appearances-based theory of decision-making explicit in a model called the InterTemporal Choice Heuristics (ITCH) model. The ITCH model assumes that psychophysical properties of the way humans reason about numbers account for the types of choices seen in the MEL paradigm, rather than impatience. By focusing on the non-normative aspects of human information-processing, the ITCH model is better able to match the quantitative patterns of observed choices. Given its usefulness in the domain of intertemporal choice, I suggest the ITCH model may embody an approach to decision-making that could be useful as an alternative to classical interpretations of experimental data.

1.2 Curiosity and the Role of Information in Decision-Making

Having enumerated a large range of problems with attempts to measure the compromise between short-term and long-term gains by asking experimental participants to make explicit trade-offs about time-dated monetary payments, we will turn to a different approach in which all of the long-term gains are made entirely implicit (Hertwig, Barron, Weber, and Erev, 2004).

In this new setting, we will focus on one strategy for increasing long-term gains at the expense of short-term gains: curiosity (G. Loewenstein, 1994). When a person decides to explore the world, they foresake the possibility of maximizing immediate reward in order to maximize the amount of information they have about the world (Auer, 2003). I will show that this decision is motivated by the possibility of exploiting newfound knowledge in future situations – and therefore represents a long-term gain chosen at the expense of short-term gains.

In order to formally model the processes that lead one to decide to explore unfamiliar goods, I will take it for granted that human beings are concerned with the long-term consequences of their actions, but that they face a problem of imperfect information (Stiglitz and Weiss, 1981). In other words, people lack information that is essential to their ability to make decisions that will lead to maximal reward. As such, people must attempt to maximize information at the same time that they maximize reward because information is necessary for maximizing reward. This will lead us to a discussion of heuristics for reasoning about the trade-offs between maximizing reward and maximizing information. Finally, I will formalize these heuristics in a new model of exploratory decision-making that I call the MaxInfo model.

Just as we focused our interest on intertemporal choice into a simple paradigm that employed decisions between two time-dated monetary payments, we will focus our

discussion of imperfect information on the problem of information-seeking exploration in the context of the Multi-Armed Bandit problem (Thompson, 1933; Robbins, 1952). In this setting, individuals make decisions between various uncertain options in order to maximize cumulative reward over a small block of sequential decisions. The Multi-Armed Bandit problem provides the simplest possible setting in which exploratory decision-making can be observed and allows me to demonstrate the power of the MaxInfo model for understanding the role of curiosity in shaping decisions.

Collectively these two research lines suggest that people employ complex and sometimes non-normative heuristics for reasoning about time-tradeoffs. They understand that long-term gains often require taking short-term losses and do their best to balance the two, although these attempts are sometimes corrupted by non-normative factors. Nevertheless, their behavior is systematic and consistent. As such, it is amenable to formal modeling. The ITCH model and the MaxInfo model represent two mathematical approaches to understanding the patterns of choice we can observe when delayed consequences are involved.

Chapter 2

Decision Theory

The theoretical basis for all of the research described in this thesis comes from a branch of mathematics called decision theory. In order to familiarize the reader with the mathematical tools that I will employ in later sections, I will briefly review the basics of decision theory (Luce and Raiffa, 1989; Gilboa, 2010). Readers already familiar with decision theory can skip or skim this chapter: it is included only to give clear and precise definitions to technical terms that I will use later on, such as utility, risk and ambiguity.

Furthermore, this brief review will only provide the most basic of outlines of the mathematical theory of decisions. Because our interests as psychologists are ultimately not normative, but descriptive, I will describe the theory of decisions only as an source of inspiration for the construction of experiments and the development of descriptive models of choice. Just as understanding the mechanisms one might use to construct a pump makes it easier to understand how the heart works (Glimcher, 2004), an understanding of the core concepts that arise in the normative theory of decisions can help to elucidate the mechanisms that might drive observed behavior.

2.1 Core Concepts: Decision-Makers, Goods, Actions

Essentially all of decision theory is an exposition of one simple scenario:

- There is a decision-maker.
- There are goods that the decision-maker values.
- There are actions that the decision-maker selects among in order to acquire desirable goods.

Throughout this thesis, we will be considering a finite setting in which our decision-maker must select one of n actions at each point in time. When we say that a decision-maker values goods, we mean that the decision-maker selects among goods/actions on the basis of a reward signal that we will assume the decision-maker seeks to maximize, but which we, as experimenters, generally cannot observe directly.

The assumption that decision-making is driven by a drive to maximize reward is really the core element of mathematical theories of decision-making: it makes the task of constructing mathematical models viable because we can assume that the decision-maker behaves as if *all of decision-making reduced to the task of solving specific optimization problems.* This "decision-making as optimization"assumption provides a template for mathematical modeling, because researchers can construct models that explain behavior simply by constructing models that makes observed decisions as close to optimal as possible. In many ways, reward maximization plays a role in models of decision-making comparable to the role of maximum likelihood estimation in statistics (Fisher, 1922).

Given that modeling decision-making as an optimizing process makes mathematical modeling tractable, we will review the basic cornerstones of the decision theoretic vision of reward maximization as applied to human beings. In decision theory, unlike

psychology, it is traditional to speak of utility instead of reward, but the terms are largely equivalent. I therefore begin with defining utility rigorously.

2.2 Neoclassical Utility Theory

The simplest model of decision-making that we will consider is called Utility Theory. I will describe Utility Theory in its neoclassical form, which stems from the seminal work of Paul Samuelson (reviewed in Houthakker, 1950).

In Neoclassical Utility Theory, we assume that:

- The decision-maker does not need to select actions: instead, the decision-maker selects directly among goods.
- There are only n goods to be selected among. We can therefore use the notation g_i to refer to the i^{th} good.
- Each of the n goods being selected among can be assigned a utility, $u(g_i)$.
- When faced with a choice between a subset of these n goods, the decision-maker will always choose the good, g^*, that has the highest utility.

In mathematical notation, we say that the chosen good, g^*, satisfies the equation,

$$g^* = \arg\max_{g \in S} u(g)$$

when S is any subset of these n goods.

The implied decision-making scenario is shown in Figure 2.1, in which a decision-maker selects from 4 goods. In this diagram, an arrow leaving one good and heading to another good indicates that the latter good is valued more highly than the first good. Neoclassical Utility Theory amounts to the assertion that the good with no arrows coming out of it will always be chosen. Thus, if Good 1 is present, it will be

chosen. If Good 1 is removed, Good 2 will be chosen. Similarly, if Goods 1 and 2 are removed, Good 3 will be chosen. Good 4 is chosen only when no other goods are available.

Although this statement may seem trivial, it is worth pointing out that Utility Theory, as a theory of decision-making, is a more demanding than it first appears:

- Although the axioms stated above can seem to be the bare minimum assumptions required to express the notion of value, Utility Theory makes quite sharp and falsifiable predictions about behavior that can be readily tested. These predictions essentially amount to context-independence and transitivity: context-independence guarantees that the presence of undesirable goods cannot make a desirable good less desirable, whereas transitivity insures that pairwise choices in one setting will easily generalize to pairwise choices in new settings. Psychologists have subsequently found strong evidence against context-independence (Tversky and Simonson, 1993, Hsee, 1996) and weak evidence against transitivity (Bar-Hillel and Margalit, 1988). As we will see, Utility Theory, as stated in its formal form, is too stringent to be a reliable description of human behavior and is falsified by most experimental data.

- Surprisingly, even though Utility Theory is falsifiable, it nevertheless does not define an identifiable model of utility: even if we could observe infinitely many choices that are consistent with Utility Theory, we could never hope to determine the specific utilities associated with a set of goods. Loosely stated, Samuelson's non-identifiability result states that observable behavior can only providing a ranking of goods, but cannot provide a reliable measure of the specific utility of any good.

In many ways, Samuelson's Utility Theory occupies a place in economics similar to the role played by Behaviorism in psychology (Skinner, 1963): Utility Theory

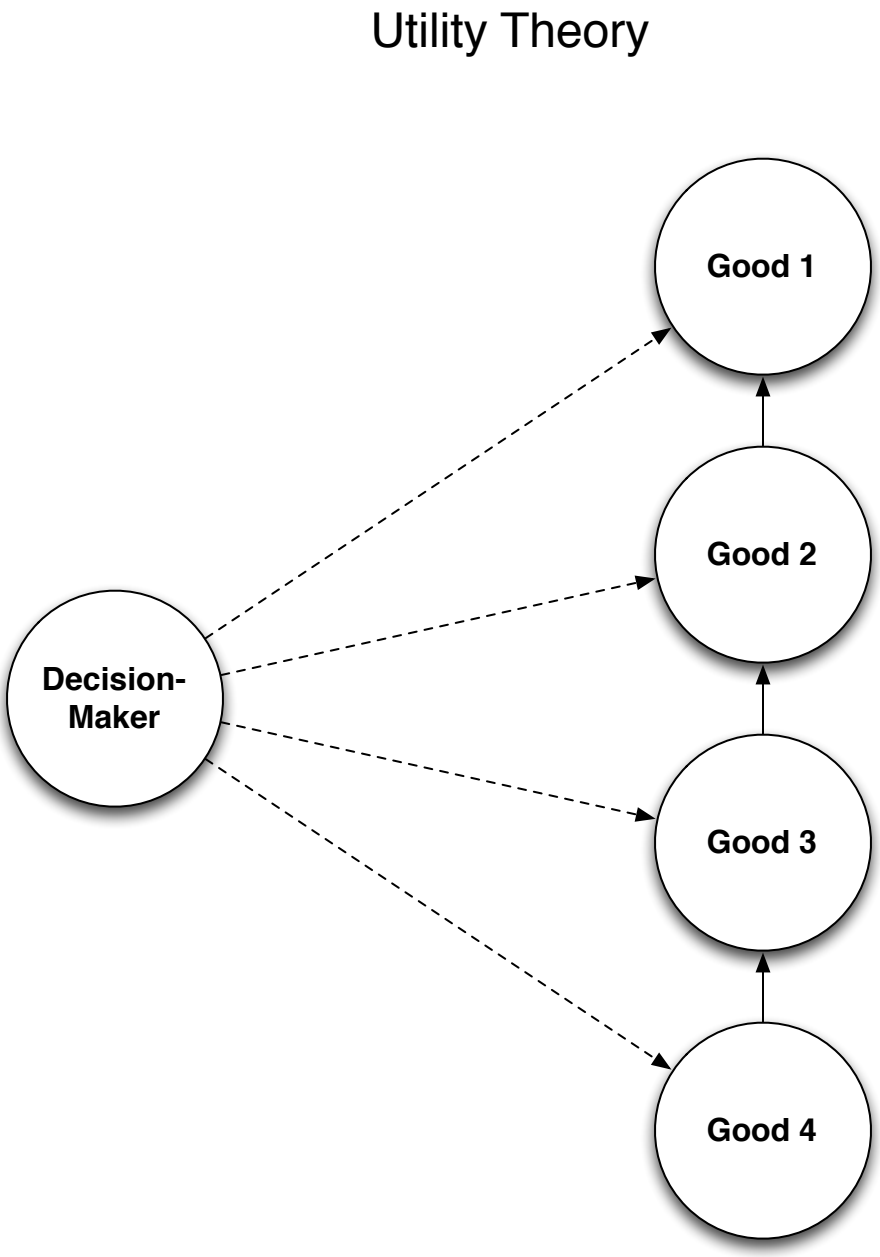


Figure 2.1: Neoclassical Utility Theory

drove economists to focus on falsifiable claims about observable behaviors, but, in the process, implied that observable behavior can never be used to identify the mental experiences of reward that the earlier generation of economists were most interested in studying.

Just as psychology has moved beyond Behaviorism, at present most economists view Samuelson's work primarily as evidence for a strong form of philosophical positivism in regard to choice data (Ayer, 1966): it is standard for economists to claim that utilities are nothing more than an assignment of numbers to goods that is consistent with observable choices (Kreps, 1988). In other words, utilities, which were the core theoretical construct of classical economics (Edgeworth, 1881; Kahneman, Wakker, and Sarin, 1997), are now treated as a convenient mathematical shorthand used to summarize a set of testable claims about observable behavior.

2.3 Classical Utility Theory

Instead of adopting the perspective taken by Samuelson, we will typically assume that *there is* a simple parametric model that relates the objectively measurable properties of a good to a *uniquely* defined utility for that good. For example, we might wish to know the exact (and hence cardinal) utility of x dollars. To do this, we might assume a parametric model in which utilities differ from dollars by a scaling factor, a, so that we can write the equation,

$$U(x) = ax.$$

In the studies of intertemporal choice that I will describe in the next two chapters, I will almost always assume exactly this mathematical relationship between dollars and utilities in order to simplify the task of modeling choice data. This linear relationship between dollars and utilities is generally called the **linear utility function**. The use of the linear utility function has been central to previous research on in-

tertemporal choice. Furthermore, I will use the linear utility model in all of my work on exploratory decision-making. My decision to use a parametric form that relates observable quantities of money to utilities does not represent an attempt at falsifying Samuelson's results about non-identifiability: rather, I will have to adopt additional assumptions in order to be able to estimate parameters like a reliably.

Fortunately, there are two popular variants of Utility Theory that will allow me to make my usage of parametric models of utility mathematically rigorous. These are:

- Expected Utility Theory (EU Theory)
- Random Utility Theory

Both of these variations of Neoclassical Utility Theory are discussed briefly in the following sections.

2.4 Expected Utility Theory

As noted above, theoretical economics was revolutionized by Samuelson's work on Utility Theory. It is hard to overstate the sweeping impact of Samuelson's work. But Samuelson's work did not remove economists' desire to possess a cardinal measure of utility that could be observed in the world. In many quarters, Samuelson's work was not viewed as a cleansing force, but as simply a problem of non-identifiability that would have to be worked around.

Economists were therefore delighted by the creation of Expected Utility Theory, which was first described in the preface to Von Neumann and Morgenstern's groundbreaking book, "Theory of Games and Economic Behavior" (Von Neumann and Morgenstern, 2007). Expected Utility Theory, which I will call EU Theory for short, extends Neoclassical Utility Theory by an additional assumption about the way that a decision-maker will decide between probabilistic goods. The simplest example of this idea comes from considering two goods:

- The first good is the certain receipt of \$20.
- The second good, in contrast, is probabilistic. For the sake of argument, assume that there is 50% chance of receiving \$40 and a 50% chance of receiving \$0.

Under Neoclassical Utility Theory, both of these goods must have some utility that can be assigned to them to explain choices. But what is the utility one should assume for the probabilistic good? This is not obvious from the axioms used by Samuelson nor is it implied by the linear model of utility that I described above. Thankfully, EU Theory provides a simple method for assigning utilities to probabilistic goods.

To describe the answer, I will generalize our example of a probabilistic good and refer to an abstract probabilistic good as a lottery (or prospect). I can then introduce a special notation for lotteries that helps to make describing EU Theory simpler. Specifically, I will write out a lottery as a list of outcomes along with their associated probabilities. In our example above, the choice is between a deterministic lottery, L_1, that I can write as,

$$L_1 = ((1.0, \$20)),$$

and a probabilistic lottery, L_2, that I can write as,

$$L_2 = ((0.5, \$0), (0.5, \$40)).$$

It is conventional to remove the parentheses in these specifications of lotteries and simply write,

$$L_2 = (0.5, \$0, 0.5, \$40).$$

In general, I can write out a lottery with n outcomes as

$$L = (p_1, o_1, \ldots, p_n, o_n),$$

where p_i is the probability that the lottery pays out the i^{th} outcome, o_i.

Given this notation, I can state the assumptions of EU Theory succinctly:

- Every deterministic good, g_i, has a utility, $U(g_i)$.
- The utilty of a non-deterministic lottery, L, is a sum of the utilities of its possible outcomes weighted by their probabilities:

$$U(L) = p_1 U(o_1) + \ldots + p_n U(o_n) = \sum_i p_i U(o_i)$$

In other words, EU Theory states that the utility of a lottery is its **expected value**, which is calculated with respect to the utilities of deterministic goods.

The virtue of this formulation is twofold:

- EU Theory extends Utility Theory and allows it to be immediately applied to probabilistic settings with only one additional axiom.
- EU Theory offers scholars of choice a mechanism for recovering uniquely defined utilities from observed choices.

This second claim is not necessarily obvious, but it is demonstrated in several textbooks (Kreps, 1988). Essentially, the introduction of probabilities provides a cardinal quantity that one can measure reliably and against which one can judge goods. By comparing goods relative to any cardinal quantity with a known measurement system, all goods can be themselves assigned unique cardinal measurements.

This second property is a compelling theoretical argument for using EU Theory, but my real interest in EU Theory is that it provides one with a simple tool for reasoning about decision-making under uncertainty. The type of uncertainty involved in EU Theory is what I will call **risk**, which refers to the objective probabilities that make outcomes non-deterministic. For example, if I flip a coin, the outcome is uncertain, but for purely objective reasons rather than subjective ones about my

lack of understanding of how coins work. Subjective reasons for uncertainty take one beyond EU Theory into the work of Savage, who developed Subjective Expected Utility Theory or SEU Theory. I now describe SEU Theory in order to contrast the form of uncertainty I call **risk** with another form of uncertainty I will call **ambiguity**.

2.5 Subjective Expected Utility Theory

EU Theory provides a model for how decision-makers make decisions about uncertain outcomes whose probabilities are assumed to simply be given to decision-maker. Subjective Expected Utility Theory extends this theory to consider cases in which the probabilities of outcomes are not known with certainty, but form instead the beliefs of the decision-maker. Notationally, I replace my previous description of a lottery as,

$$L = (p_1, o_1, \ldots, p_n, o_n),$$

with the notation,

$$L = (\pi_1, o_1, \ldots, \pi_n, o_n),$$

where π_i is a subjective probability that reflects a belief about the world rather than an objective fact about the non-determinism of the lottery. I do not exclude the possibility that subjective probabilities are equal to objective probabilites: I simply do not require such equality.

With the introduction of subjective probabilities as a mechanism for describing beliefs, SEU Theory reduces to the assertion that,

$$U(L) = \sum_i \pi_i U(o_i).$$

In economics, the full requirements for adopting SEU Theory are known as the Savage Axioms (Savage, 1972). They are arguably the most elegant component of the theory

of decisions, but their full description would lead us astray.

For my work, the value of SEU Theory is that it will make my later discussion of exploration and exploitation more precise, because I will need to articulate two distinct notions of uncertainty in order to contrast objective non-determinism from beliefs encoded as subjective probabilities. Whereas the probabilities considered in EU Theory were objective facts about the world that led to **risk**, the probabilites considered by SEU Theory may reflect purely subjective uncertainty, which I will call **ambiguity**.

The terms **risk** and **ambiguity** derive from the work of the economist Frank Knight (reprinted in Knight, 2006). Distinguishing between risk and ambiguity is central to an understanding of the value of information. I will elaborate on this distinction in some detail when I describe the Ellsberg Paradox in a later chapter, but for now I note the following crucial distinction between risk and ambiguity:

- Risk is an objective fact about the world that cannot be modified by the decision-maker's actions. It is inescapable fact about a situation that the decision-maker must reason about, but cannot resolve.

- Ambiguity is a reflection of the decision-maker's ignorance about the world. By taking actions that are maximally informative, ambiguity can be minimized. In other words, ambiguity, unlike risk, is an actionable form of uncertainty. This point will be central to my MaxInfo model of exploratory decision-making in which I will assume that people behave so as to minimize ambiguity.

2.6 Random Utility Theory

In general, all of the theories described above will fail to account for some of the choices one observes in the lab. This problem can be partially resolved by developing subtler models of utility, but the primary problem is that an experimental participant may

reverse their stated preference between two similar goods under repeated questioning. Unlike all of the models described above, empirical decisions between well-specified goods do not elicit perfectly reliable responses.

To cope with the inconsistencies observed in real choice data, I will assume that choices are probabilistic and that, given a specific choice set, there is a unique probability of selecting each good. For example, in choosing between goods A and B, these probabilities might be defined in such a way that they act like utilities: thus, $p_{A,B}(A) > p_{A,B}(B) \Leftrightarrow u(A) > u(B)$.

The standard way to construct such probabilities is to assume that unique cardinal utilities do exist, but that they are perceived by decision-makers after being corrupted by random noise. After such noise is added to the raw utilities, decision-makers select actions that maximize noisy utilities. For example, a good A is chosen over B if, after adding noise η_A and η_B,

$$u(A) + \eta_A > u(B) + \eta_B.$$

An equivalent formulation of this statement is that A is chosen over B if,

$$u(A) - u(B) > \epsilon,$$

where ϵ is the difference of the two random noise terms, η_A and η_B.

If one assumes that ϵ has a logistic distribution with mean 0 and variance 1, this noisy utility model produces a classic model of choice,

$$\mathbb{P}(\text{Choose A over B}) = \frac{e^{U(A)}}{e^{U(A)} + e^{U(B)}} = \frac{1}{1 + e^{-(u(A)-u(B))}},$$

which is a standard formula for computing the probability that ϵ is less than $u(A) - u(B)$ (McFadden, 1973; Luce, 1977; Gul and Pesendorfer, 2006).

I will therefore use a random utility model with logistic noise in all of the modeling

work that follows. In binary choice paradigms, this will produce models that behave like (potentially non-linear) logistic regression; in many-good choice paradigms, this will produce models that behave like (potentially non-linear) multinomial regression. This simplification makes the statistical estimation of model parameters tractable (McCullagh and Nelder, 1989).

2.7 Discounted Utility Theory

A substantial addition to Neoclassical Utilty Theory, EU Theory and SEU Theory comes in the form of Discounted Utility Theory (DU Theory for short), in which we assume a further axiom which specifies how time-dated goods are valued. This material will form the content of the next chapter as it provides the entire content.

Chapter 3

Intertemporal Choice Theory

3.1 Synopsis of DU Theory

Just as EU Theory extends Utility Theory to situations in which outcomes are uncertain because of non-determinism in the world and SEU Theory extends Utility Theory to situations in which outcomes are uncertain because of the decision-maker's ignorance about the world, DU Theory extends Utility Theory to outcomes that are delayed in time (Samuelson, 1937; Frederick et al., 2002).

As a simple example, consider the possibility of receiving \$10 immediately. As I stated earlier, one can assume that the utility of \$10 is $10a$ for some value a. But what if a person were going to receive \$10 in 1 day? Or in 5 days?

DU Theory answers these sort of questions in a simple way: under DU Theory, one assumes that the utility of \$10 decreases as we delay its receipt into the future. In the simplest case, one can assume that every day's worth of delay diminishes the utility of \$10 by 10%, so that \$10 today has $10a$ utility, \$10 in 1 day has $(10 * 0.9)a$ utility and \$10 in 2 days has $(10 * 0.81)a$ utility.

More abstractly, the devaluation of delayed goods can be defined in terms of a discount function, $D(t)$, which specifies the *percentage of utility* remaining at time t.

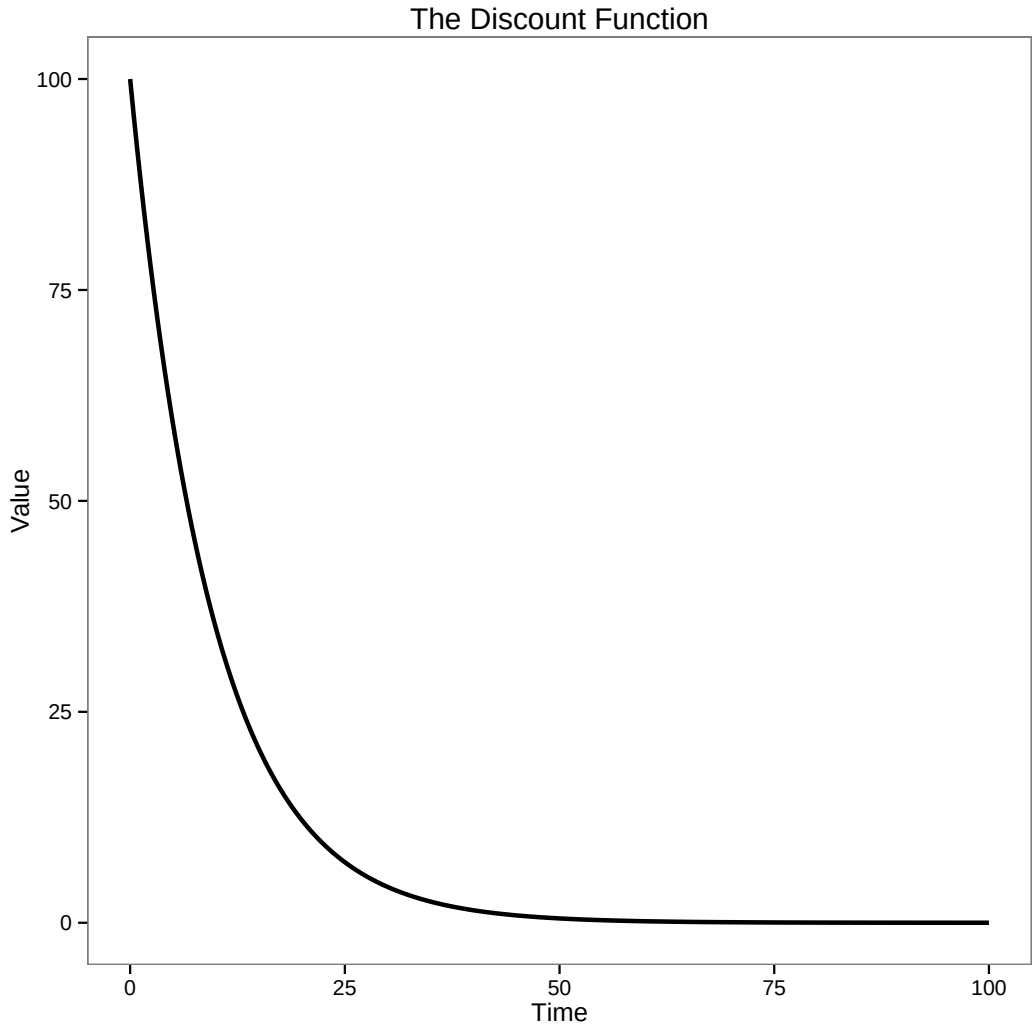


Figure 3.1: A hypothetical discount function

One possible discount function, which decreases geometrically over time, is shown in Figure 3.1. I will describe this specific discount function in detail in the sections that follow because this geometric discounting rule occupies a central place in the theory of intertemporal choice.

Given the definition of a discount function, I can state the assumptions of DU Theory precisely. To do so, I adopt the notation (g, t) for a time-dated good g that will be received at time t. Expanding on this notation for a single good, I can specify sequences of time-dated goods using a notation like that used for lotteries: a

sequence S of time-dated goods, $S = ((g_1, t_1), \ldots, (g_n, t_n))$, will be written as $S = (g_1, t_1, \ldots, g_n, t_n)$. With this notation and the definition of the discount function, DU Theory is summarized entirely by the assumption that *the utility of a sequence,* $S = (g_1, t_1, g_2, t_2, \ldots, g_n, t_n)$*, of time-dated goods is*,

$$U(S) = \sum_{i=1}^{n} U(g_i)D(t_i).$$

This equation makes it clear that DU Theory's mathematical structure is similar in many ways to EU Theory's. The central difference is that the $D(t_i)$ are not constrained to sum to 1.

This simple mathematical structure has remarkable power as a theoretical framework for understanding choices involving delayed consequences. For example, one could define **impatience** in terms of a person's discount function: if Person A has a discount function that devalues goods faster than Person B's discount function, we would say that A was more impatient than B (Ainslie, 1975, Shamosh et al., 2008). This conception of impatience would make clear and falsifiable predictions about behavior: it would predict that person A would be more likely to become addicted to drugs or to neglect climate change than person B because the source of their impatience is a discount function that affects decisions about all goods – a prediction I will soon formalize as DU Theory's **domain invariance** property. If DU Theory were true, the discount function would be a sweeping individual preference parameter that determined behavior in all settings and explained many important individual differences.

Perhaps more remarkably, the assumption that discount functions are used to make decisions involving delayed consequences makes strong predictions about human behavior even without making any quantitative comparisons between individuals. For example, we will soon see that the structure of the discount function could be used

to explain why some individuals suffer from self-control problems in which they make plans to behave patiently, but fail to live up to those plans (Ainslie, 1975, Laibson, 1997; but also see Gul and Pesendorfer, 2003; Fudenberg and Levine, 2006 for a different approach). Going further still, we could account for people's propensity to precommit in terms of their knowledge of their own discount function and their attempts to make up for the deleterious role that their discount function plays in shaping their decisions (Strotz, 1955).

3.2 DU Theory in Detail

Because DU Theory has so much potential to provide a theoretical framework for explaining both (a) the differences between individuals that seem to cause different people to lead different lives and (b) the highly generic failures of self-regulation that affect nearly all human beings, DU Theory has received considerable attention in the theoretical lierature in economics (Frederick et al., 2002). In what follows, I will highlight only the most salient aspects of this literature.

3.2.1 DU Theory in a Binary Choice Problem

The simplest setting in which DU Theory applies involves two time-dated goods: an earlier good and a later good. Because one generally assumes that delayed goods are devalued, this binary choice setting is interesting only when the earlier good is a good with smaller utility than the later good. In this case, we can learn a great deal about the discount function from binary choices because DU Theory says that a decision-maker will be indifferent between an earlier good $e = (g_1, t_1)$ and a later good $l = (g_2, t_2)$ precisely when $U(e) = U(l)$. Because of the core assumption of DU

Theory, this occurs precisely when,

$$U(g_1)D(t_1) = U(g_2)D(t_2).$$

Moreover, if we knew that this equality held for an alternative set of goods, then we would know that the decision-maker would be indifferent between those goods as well.

This specific equality can be quickly related to observable choices if we assume that (a) g_1 is a monetary good providing x_1 dollars and g_2 is also a monetary good providing x_2 dollars and that (b) the utility function is linear. In this case, this equality can be restated as

$$\frac{ax_1}{ax_2} = \frac{x_1}{x_2} = \frac{D(t_2)}{D(t_1)}.$$

In short, we can measure a ratio of the discount function's value at two points in time by simply observing binary choices between time-dated monetary goods. If we can find two time-dated goods that make a decision-maker indifferent between an earlier good and a later good offered at times t_1 and t_2, then we can determine this ratio exactly.

Because these ratios are potentially easily observable in the lab, we will refer to the specific ratio,

$$\frac{D(t_2)}{D(t_1)},$$

as the **discount factor** over the interval of time $[t_1, t_2]$. The equality shown above does not do enough to emphasize the central role of discount factors in DU Theory: under DU Theory, the discount factor over the interval of time $[t_1, t_2]$ is the only thing that determines the role of time in shaping decision-making. For any two goods, g_1 and g_2, DU Theory implies that, if the ratio, $\frac{U(g_1)}{U(g_2)}$, is less than the discount factor, then the later option will be chosen. Alternatively, if the ratio, $\frac{U(g_1)}{U(g_2)}$, is larger than the discount factor, the earlier option will be chosen. Thus, if DU Theory were

correct and we knew the discount factor for every period in time, we could predict all intertemporal choices *perfectly*. For this reason, the discount factor is the primary object that we will attempt to measure experimentally. As will be seen in the next chapter, I believe that current research has been far from successful in measuring discount factors.

To elaborate on the centrality of the discount factor as the object that must be measured to fit DU Theory to real data, I note that the ability to measure discount factors using only indifference points is particularly important for two reasons:

1. If the earlier good is available immediately (i.e. $t_1 = 0$), then the discount factor is exactly equal to the value of the discount function at t_2. This is because one generally assumes that $D(0) = 1$. Under this assumption, the discount factor over the interval of time $[0, t]$ is $\frac{D(t)}{D(0)} = D(t)$. In other words: even though I have defined DU Theory in terms of a discount function, I could work exclusively with discount factors – which are, under the assumptions stated earlier, directly observable in a binary choice paradigm.

2. Even if I do not attempt to observe discount factors that start at time 0, one can still learn a great deal about the properties of the discount function from studying its discount factors because the discount function's values are uniquely determined by its discount factors. We will see in a moment that the discount function can always be written in terms of its discount factors by taking products.

To summarize, the binary choice case makes a central fact about DU Theory clear: *we can build an entire theory of decision-making that predicts how all delayed consequences will be reasoned about out of nothing more than observations from a binary choice paradigm in which we find two time-dated goods,* $e = (g_1, t_1)$ *and* $l = (g_2, t_2)$ *that make the decision-maker indifferent for every interval of time,* $[t_1, t_2]$. As

we will see later, the classical parametric models of the discount function make this econometric problem even simpler: one needs only to measure the discount factor over exactly one interval – namely the interval $[0, 1]$ –, in order to calibrate an entire theory of intertemporal choice. Even if one considers several different parametric models of the discount function, one will need only to measure discount factors over a very small number of intervals of time to fully calibrate DU Theory. In fact, none of the classical models of intertemporal choice ever require that one measure discount factors over more than three intervals of time in before completely pinning down the model's functional form.

The remarkably small number of measurements required to develop a theory of intertemporal choice is what makes both DU Theory and the experiments that have been used to test it so compelling. Before I move on, I will make clear how these measurements could be carried out. This will lay the foundation for the Money Earlier or Later (MEL) paradigm that I will examine in detail in the next chapter.

As mentioned earlier, one can begin this measurement process by considering behavior over the interval of time $[0, 1]$. One then selects a fixed monetary good that will serve as the earlier, smaller good in the binary choice setting described earlier. For now, let us assume that this good is \$10. Under these assumptions, one can measure the discount factor, $\frac{D(1)}{D(0)}$, by locating a second monetary good, x, such that the experimental participant is indifferent between receiving $(\$10, 0)$ and $(\$x, 1)$. To find this good, one can start by offering \$11 and counting up until the participant stops choosing the earlier, smaller good and starts to choose the larger, later good. After reaching this point, the procedure can backtrack until it hones in on the exact value of x that makes the decision-maker indifferent between $(\$10, 0)$ and $(x, 1)$.

Once x has been determined, the discount factor can be computed as $\frac{10}{x}$ based on the equations shown earlier. As such, this simple choice experiment will completely calibrate a sweeping theory of decision-making after determining only one number –

namely x – that makes a person indifferent between one specific immediate good and one specific delayed good. This simplicity explains much of the theoretical appeal of DU Theory and also the appeal of the binary choice paradigm that has been used to calibrate models of the discount function. Moreover, even if the researcher is not confident that they know the functional form of the discount function well enough to only need to measure a single discount factor, they can always try to measure a decision-maker's discount factors over several intervals of time – say $[0, 1]$, $[0, 2]$ and $[1, 2]$. In general, a very small sample of discount factors will provide sufficient data for calibrating a general theory of intertemporal choice. As will be seen, the converse is also true: only a small number of measurements are required to falsify DU Theory.

As should perhaps go without saying, it will become clear as we go on that real human behavior is radically more complicated than the line of reasoning I have just described would suggest. Determining the exact ways in which attempts to calibrate DU Theory fail will constitute the major work of the next chapter. For now, I set the stage for this discussion by highlighting some of DU Theory's testable predictions.

3.3 Invariances of DU Theory

DU Theory, as thus far described, is a very powerful theory that requires minimal data to be calibrated to make precise quantitative predictions about human behavior. In practice, these predictions depend strongly on the exact functional form of the discount function that one assumes decision-makers possess. I will discuss the classical functional forms for the discount function in the next section. But for now I will focus on a secondary issue: what are the predictions that DU Theory makes regardless of the form of the discount function?

Surprisingly, DU Theory makes several very strong predictions about behavior that either (a) do not depend at all on the discount function's form or (b) impose

extremely tight constraints on the discount function. In particular, I describe four properties of DU Theory that are relevant for understanding both the previous literature on DU Theory as well as the so-called "anomalies of DU Theory" that my ITCH model is meant to account for.

3.3.1 Dynamic Consistency

The first prediction of DU Theory involves a property called **dynamic consistency**. A decision-maker is said to be dynamically consistent if they make plans for the future that are consistent with the actions that they will be willing to carry out when the time comes to implement their plans (Strotz, 1955). Failures of dynamic consistency are omnipresent in human life: we plan to go to the gym as a New Year's resolution and stop going after a few weeks, we plan to wake up early and get a head start on an impending project until we decide to sleep in after hearing the alarm clock, etc. (Laibson, 1997).

In DU Theory, there is only one possible discount function that ensures dynamic consistency: it is the geometric discount function mentioned earlier in which each unit of time removes a fixed percentage of value from a good. This function is generally called the **exponential discounting model** and is written as $D(t) = \delta^t$ (Samuelson, 1937).

To see intuitively why the exponential discounting function is the only discount function that produces dynamic consistency, I will make two points:

1. Dynamic consistency is entirely determined by the consistency of discount factors as we shift an interval of time into the future. Two possible shifts in time are shown in Figure 3.2.

2. Any discount function whose discount factors are not a constant when shifted in time will produce inconsistencies.

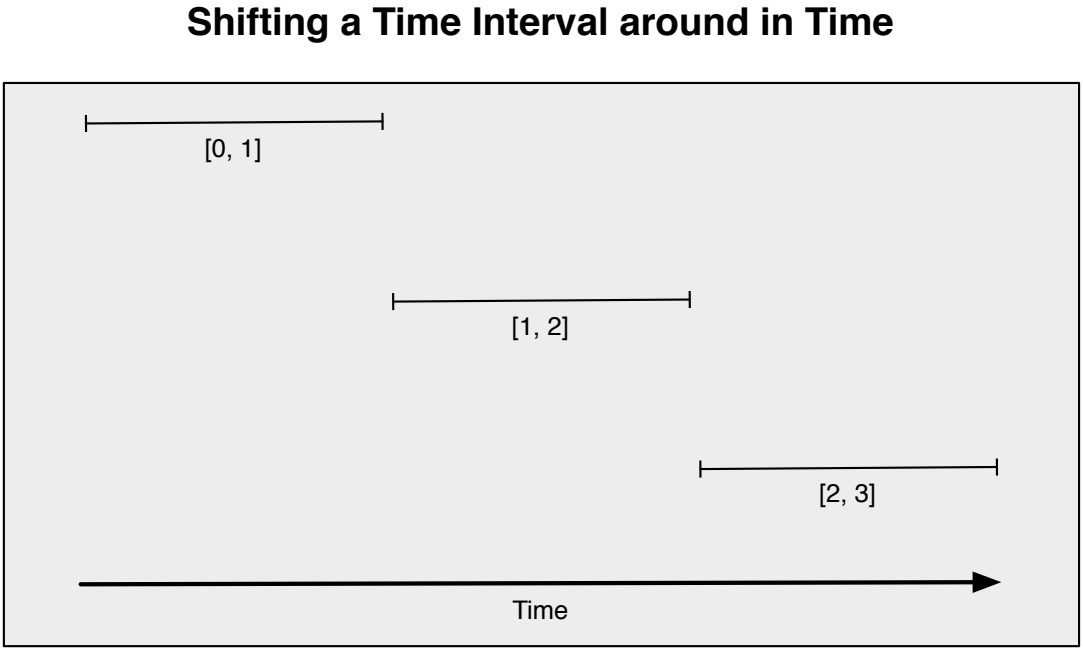


Figure 3.2: Shifting an interval of time around

Giving a rigorous argument here would take us rather far afield, but the general line of argument is as follows: a decision-maker will successfully implement a plan that was made for the future if its preferences between all of the time-dated goods involved in that plan never reverse at any point in time. Because preferences between goods only depend on the discount factor, this implies that the discount factor must be a constant when we shift an interval of time around. Formally, this requirement means that if (g_1, t_1) is preferred over (g_2, t_2), then one will also require that $(g_1, t_1+\tau)$ be preferred over $(g_2, t_2+\tau)$ for all τ. But preferences like this can be consistent for all goods g_1 and g_2 only if the relevant discount factors, $\frac{D(t_2)}{D(t_1)}$ and $\frac{D(t_2+\tau)}{D(t_1+\tau)}$, are identical. If these numbers are not identical at any point in time τ, there will be at least one plan that will not be carried out by the decision-maker.

To see why this property implies a specific discount function, we need to note something else: the discount factors over a sequence of intervals of time can be multi-

plied together to produce the discount function's value over that whole block of time. If we consider a set of time points, τ, 2τ, ..., $n\tau$ all separated by an inteval of time of length τ, this multiplication rule means that

$$D(n\tau) = \frac{D(\tau)}{D(0)}\frac{D(2\tau)}{D(\tau)}...\frac{D(n\tau)}{D((n-1)\tau)}.$$

Because we require that the discount factors at each point in time, $\frac{D(i\tau)}{D((i-1)\tau)}$, be a constant (say δ) to ensure dynamic consistency, this factorization means that,

$$D(n\tau) = \delta * \delta \ldots \delta = \delta^n.$$

Because this line of logic holds for any τ, no matter how small, one has,

$$D(t) = \delta^t,$$

as the limiting case. In other words, DU Theory ensures dynamic consistency only when the discount function has exactly one functional form: the exponential discounting model.

Although the details required to make this line of argument rigorous are quite technical (see Strotz, 1955 for one rigorous proof), the conclusion that I have stated has sweeping qualitative ramifications for DU Theory. The statement that only the exponential discount function guarantees dynamic consistency means that, if DU Theory were correct and one were to find *any evidence* for deviations from exponential discounting, researchers could use DU Theory to both (a) explain why people make plans that they do not live up to and (b) predict in precisely which situations people will fail to live up to their plans for the future.

As such, a major part of the empirical literature on intertemporal choice has focused on finding deviations from exponential discounting as evidence for dynamic

inconsistency (Ainslie, 1975; R. Thaler, 1981). In addition to finding evidence for dynamic inconsistency, the literature has also been concerned with how decision-makers cope with their knowledge that they are not exponential discounters. If people know that they will not live up to their plans, how can they plan intelligently for the future?

To address this question, theorists have focused on parametric models of the discount function in which the discount factors in the far future imply that decision-makers will make extremely patient plans, whereas the discount factors near the present imply that the same decision-maker will exhibit powerful impatience when the time comes for them to implement their plans. If decision-makers are aware of this inconsistency and wish that they would fulfil their plans, they may develop a taste for precommitment (Laibson, 1997; O'Donoghue and Rabin, 2001). Thus, from nothing more than a technical argument about the form of the discount function, economists have constructed an entire theory of self-control and precommitment.

I will describe more of these arguments in some detail later on, especially because the ITCH model I will propose implies that we, as experimenters, will observe decisions in binary choice settings that provide no evidence for or against people's general ability to follow through with their plans. For now, though, I will move on to three other broad predictions of DU Theory.

3.3.2 Domain Invariance

The argument given above regarding dynamic consistency has such sweeping ramifications because the effects of time on choice in DU Theory are domain general: the same discount function guides all intertemporal choices, including retirement savings, educational investments, and health investments. I will refer to this property as **domain invariance**. Domain invariance is an intrinsic property of DU Theory because qualitative properties like dynamic consistency are exclusively derived from the func-

tional form of the discount function and do not depend on the identities of the goods being considered – only on their timing. As such, a believer in DU Theory expects that patterns of choice like dynamic inconsistency will show up in all domains of behavior. If one observes dynamic inconsistency in regard to decisions about money, DU Theory requires that the theorist predict that dynamic inconsistency will also be unobserved in decisions about health and nutrition as well.

To see this, consider two families of goods, $((x_1, t_1), (x_2, t_2))$ and $((y_1, t_1), (y_2, t_2))$. If one assumes that $U(x_1) = U(y_1)$ and $U(x_2) = U(y_2)$, then DU Theory guarantees that $U(x_1, t_1) = U(y_1, t_1)$ and $U(x_2, t_2) = U(y_2, t_2)$ for all times t_1 and t_2. Because equated utilities entirely determine behavior, all patterns of choice will be identical across domains if one finds a pair of goods with matching utilities. In short, DU Theory absolutely forbids the possibility that people will be impatient about food, but patient about education. Research (e.g. Chapman and Elstein, 1995) has, unsurprisingly, found evidence against this claim. In my work, I have not focused upon this prediction of DU Theory, but the evidence against it found by other researchers is relevant to my critique of DU Theory.

3.3.3 Magnitude Invariance

In addition to invariance across domains of goods, DU Theory imposes a specific invariance within a domain of goods. To see why DU Theory makes any prediction within a domain, consider a choice between two time-dated monetary goods: (g_1, t_1) and (g_2, t_2). We will say that a model is **magnitude invariant** if indifference between the goods (g_1, t_1) and (g_2, t_2) implies indifference between the rescaled goods $(\theta g_1, t_1)$ and $(\theta g_2, t_2)$ for all values of the rescaling term θ.

Any parametric model of DU Theory possesses this Magnitude Invariant property whenever the utility function U has the property that $\frac{U(x)}{U(y)} = \frac{U(\theta x)}{U(\theta y)}$. The reason for

this is that indifference between (g_1, t_1) and (g_2, t_2) is equivalent to the claim that,

$$U(g_1)D(t_1) = U(g_2)D(t_2).$$

Shifting pieces around to describe indifference in terms of discount factors, this gives,

$$\frac{U(g_1)}{U(g_2)} = \frac{D(t_2)}{D(t_1)},$$

which says that indifference is determined only by the equality of (a) the ratio of utilities and (b) the discount factor over $[t_1, t_2]$. Because the discount factor is a fixed property determined only by t_1 and t_2 and is not affected by the goods involved, this indifference-defining equation will still hold and will still imply indifference if $\frac{U(x)}{U(y)} = \frac{U(\theta x)}{U(\theta y)}$ because,

$$\frac{U(\theta g_1)}{U(\theta g_2)} = \frac{U(g_1)}{U(g_2)} = \frac{D(t_2)}{D(t_1)}.$$

I will say that any utility function that satisfies this requirement has the **ratio preserving** property. Many popular utility functions possess this property including the linear utility function that I will use in all of my analyses. Moreover, merely adding concavity or convexity to the utility function does not remove the ratio preserving property: many concave functions like logarithms and powers also have the ratio preserving property.

In short, the linear utility formulation of DU Theory that I will use will makes the strong prediction that intertemporal decisions will be consistent between small and large magnitudes of money. This prediction is robustly falsified by the empirical data on intertemporal choice (G. Loewenstein and R. H. Thaler, 1989). The ITCH model that I will describe predicts exactly this sort of violation. *Moreover, the ITCH model asserts that failures of magnitude invariance arise from the exact same psychophysical properties that induce the appearance of dynamic inconsistency.*

3.3.4 Partition Invariance

An important consequence of the argument I used to demonstrate that dynamic consistency is the additional prediction of DU Theory that the discount factors over the components of any interval combine naturally to produce the discount factor over the entire interval. For example, DU Theory implies that,

$$\frac{D(2)}{D(0)} = \frac{D(1)}{D(0)}\frac{D(2)}{D(1)}.$$

This property is often called additivity – a name that makes sense when one works with the logarithms of the discount factors and replaces the product in the equation above with a sum. I will call this property **partition invariance**. Partition invariance is powerfully violated by the empirical data (Read, 2001) and cannot be accomodated by any manipulations of the discount function or utility function. In contrast to DU Theory's inability to predict violations of partition invariance, it will emerge as a natural consequence of some parameterizations of the ITCH model.

3.4 Models of the Discount Function

As noted earlier, the most typical model of the discount function is the exponential discount model: $D(t) = \delta^t$. This model is the only model under DU Theory that provides dynamic consistency.

Thus other models are chosen on the basis of how deviations from dynamic consistency are expected to appear. The most popular models of dynamic inconsistency under DU Theory fall in the hyperbolic family, which includes the hyperbolic model, quasi-hyperbolic and generalized hyperbolic models (Ainslie, 1975; Laibson, 1997; G. Loewenstein and Prelec, 1992). These models assert that the discount function has one of the functional forms shown below:

1. Hyperbolic: $D(t) = (1 + \alpha t)^{-1}$

2. Quasi-Hyperbolic: $D(t) = \beta^{I(t>0)}\delta^{t}$

3. Generalized Hyperbolic: $D(t) = (1 + \alpha t)^{-\frac{\beta}{\alpha}}$

The core distinction between these models is that the hyperbolic model begins to produce discount factors that imply almost perfectly patient decision-making in the near-term future (e.g. over the interval $[6, 7]$), whereas the quasi-hyperbolic and generalized hyperbolic models do not make this prediction. The quasi-hyperbolic model also has the distinction of being much easier to work with analytically, but the generalized hyperbolic is often easier to fit to data because it is continuous everywhere. Regardless of the specific functional form used within the hyperbolic family, these models all exhibit a property called decreasing impatience by Prelec (Prelec, 2004), which asserts that shifting a fixed-length interval of time into the future will produce less impatience (i.e. increase the discount factor). This prediction is a direct and targeted violation of dynamic consistency which predicts that plans will become increasingly patient as they come to refer to events that will occur further and further into the future.

Several other DU Theory-consistent models exist in the literature, including the two models shown below (for examples, see Mazur, 1989; Rachlin, Raineri, and Cross, 1991; Tanaka, Camerer, and Nguyen, 2006):

1. Extended Hyperbolic: $D(t) = (\gamma + \alpha t)^{-1}$

2. The Tanaka Model: $D(t) = \beta^{I(t=0)}(1 - (1 - \theta)rt)^{\frac{1}{1-\theta}}$

These models are not very common in the literature, but are included for completeness. The Tanaka model is more complex than most parametric models of the discount function, but it has the virtue that it includes the quasi-hyperbolic, hyperbolic and exponential models all as special cases under appropriate parametrizations.

Moving outside of DU Theory for a moment, there are several models that have emerged in the recent literature including the following (Benhabib, Bisin, and Schotter, 2010; Glimcher, Kable, and Louie, 2007; Read, Frederick, and Scholten, 2011):

- Fixed Cost Model
- Kable-Glimcher Model
- DRIFT Model

These models were all designed to explain some of the anomalies that occur in the literature, especially failures of partition invariance. We will elaborate upon them later when comparing them to ITCH. Because they are not based on DU Theory, they cannot be expressed in terms of a discount functions. Moreover, the qualitative predictions they make are, in general, still not yet understood.

3.5 Probabilistic Extensions of DU Theory

DU Theory as presented in this chapter is a form of Classical Utility Theory and is therefore fully deterministic. Because this determinism implies perfect reliability under repeated questioning, simple models based on DU Theory are falsified in essentially any laboratory experiemnt.

One therefore needs to convert the models described in this chapter into Random Utility Theory models by adding noise terms (e.g. noise with a logistic distribution). In some of the experiments that I will describe, this logistic noise assumption has the undesirable property that assuming moderate noise when low stakes decisions are made implies nearly zero noise will be observed when high stakes are made. Alternatively, assuming moderate noise is present when decisions are made at high stakes implies that one would observe completely random responding at low stakes. Experimental data does not corroborate these predictions.

In general, it is difficult to find a noise model for DU Theory that performs well in predicting the choices in the experiments that I have used to calibrate the ITCH Model. I therefore note before closing this chapter that DU Theory has another failing as a theory of intertemporal choice: because it is not explicitly probabilistic, DU Theory requires additional assumption before it can be fit to real data sets and these assumptions are insufficiently specified in the literature to make DU Theory a viable model of observed decisions.

Chapter 4

Intertemporal Choice Experiments

In this chapter, I summarize the most salient experiments I have conducted on intertemporal choice. The overarching conclusion from these studies is that traditional experimental designs for studying intertemporal choice are not able to measure a discount function that is consistent with DU Theory. Instead, the choices posed in intertemporal experiments are generally answered using idiosyncratic heuristics that reflect interesting psychological phenomena, but are not consistent with the kind of quantitative forms of impatience that drive DU Theory as an abstract theory of intertemporal choice. While these heuristics are interesting objects of study, they are not equivalent to any economic theory of systematic discounting of future goods. Moreover, the choices observed in these experiments can be powerfully shifted by manipulating the format of normatively equivalent questions. Much of my work has focused on assessing how far behavior can be shifted by minor manipulations of format, but I have also attempted to determine whether intertemporal choice experiments can be constructed in a way that is more likely to measure the kinds of impatience that we see in the field.

4.1 The Money Earlier or Later (MEL) Task

The traditional intertemporal choice experiment expands upon the example of the binary choice setting described from a theoretical perspective in the previous chapter. In the typical intertemporal choice experiment, the experimenter asks the participant to make a sequence of binary choices, each of which requires a decision be made between two time-dated monetary goods: an earlier, smaller good and a later, larger good (Frederick et al., 2002). I will refer to this generic experimental paradigm as the **Money Earlier or Later** task – or MEL for short.

Based on results from searching Google Scholar for papers containing terms like "time preference", "money AND intertemporal choice" and "intertemporal choice AND experiment", I estimate that more than 5,000 published papers have employed a variant of the MEL task that I analyze in this chapter. The MEL task has grown in popularity for many reasons:

- MEL tasks are extremely easy to administer. If the experimenter does not wish to use monetary incentives, a MEL tasks requires only that the experimenter design a simple questionnaire in order to study intertemporal choice.

- The MEL task is the simplest possible setting in which DU Theory makes testable predictions. Moreover, if one assumes that experimental participants make intertemporal choices using a known utility function, then it is possible to measure the precise functional form of the discount function by determining, for every point in time, how much additional money must be offered at the later date in order to induce the participant to prefer the later, larger good over the earlier, smaller good. The procedure described in the previous chapter can be used to determine the appropriate larger, later good.

As in Chapter 3, I will use the notation (x_1, t_1) to denote the earlier, smaller good and the notation (x_2, t_2) to denote the later, larger good. With this notation, the

generic procedure described above can be used to trace out the indifference curve between all pairs of time-dated goods. For example, given the goods ($\$10, 0$ days) and ($\$x_2, 1$ day), one could produce an estimate of the value of x_2 that made the participant indifferent between these two goods. The value of x_2 would then allow one to compute a discount factor and fit a parametric model of the discount function. Below I discuss several popular mechanisms for estimating the value of x_2.

At the start of this discussion, I again note that the simple models of intertemporal choice presented in the previous chapter all share the property that only a very small number of indifference points needs to be estimated to uniquely determine the exact discount function being used by participants. For example, the exponential and hyperbolic discounting models only require that a single indifference point be estimated in order to estimate the entire discount function. The efficiency with which a broad model of intertemporal choice can be estimated from MEL data makes the MEL task very appealing to researchers.

There are two primary variants of the MEL task in widespread use:

- **The MPL Design**: Economists typically conduct MEL tasks in which a sequence of closely related options are presented in order (Andersen, Harrison, Lau, and Rutström, 2006). For example, the participant might first choose between ($\$10, 0$ days) and ($\$11, 1$ day) and immediately after decide between ($\$10, 0$ days) and ($\$12, 1$ day). By ordering questions so that either money or time values strictly increase, experimenters try to make the MEL task more comprehensible. The strict ordering of questions also increases the likelihood that the participant will produce responses that satisfy the basic transivity requirements of Utility Theory. Any ordered list in which only one attribute of a good changes across questions is typically referred to as Multiple Price List. I will abbreviate this as MPL. In the MPL design, it is traditional to present all of the related questions at once in an ordered table in which the rows vary

along one of the four dimensions relevant for a MEL task – namely, x_1, t_1, x_2 or t_2. An example of this specific tabular structure is shown in Figure 4.6 later in this chapter.

- **The Randomized Choice Design**: Psychologists typically wish to randomize the order of questions across participants in order to control for contextual cues that might be provided by the MPL design. As such, psychologists generally take the questions that would be used to construct an MPL design experiment and randomize their order of presentation so that only one binary choice is shown at a time. This design has several benefits beyond the removal of the contextual cues provided by an MPL design: the isolation of each binary choice makes it easier to determine the participant's current focus of attention in experiments that involve eye-tracking or other methods for measuring local attention such as MouseLab (Johnson, Payne, Bettman, and Schkade, 1989). Moreover, the isolated binary choice paradigm is generally more amenable to fMRI experiments in which the experimenter would like to assert that imaged brain responses are triggered by the content of specific options being shown at a specific time. The primary cost of the randomized choice design is that typically produces data in which intransitivies are more frequent, which makes the statistical estimation of economic models more difficult.

These two variants of the MEL paradigm share one virtue: they can be incentivized by offering to pay participants in accord with their expressed preferences. If the participant is told at the start of an experiment that their choices will be enacted as stated, the experiment is said to be **incentive-compatible** because the participant is incentivized to state their true preferences (Grether and Plott, 1979; Becker, DeGroot, and Marschak, 2006). Economists typically require that experiments be incentive-compatible. As a psychologist, I have not employed incentivization except in a subset of my experiments. As will be seen, adding incentivizes has not affected

the substantive conclusions of the experiments in which they were provided.

The possibility of incentivizing both the MPL and randomized choice forms of MEL tasks contrasts strongly with another experimental design similar to MEL that has also been used in the literature on intertemporal choice. In this **matching** paradigm, the participant is asked to fill in a blank item in an incomplete description of two time-dated goods. For example, a participant might be asked to type in the value of x_2 for which the participant would be indifferent between ($\$10, 0$ days) and ($\$x_2, 1$ day). I mention the matching paradigm only for completeness. Although I have employed the matching paradigm in some of my work, I will not describe any experiments in which it was used in this thesis.

In particular, note that the matching paradigm is not incentive-compatible because the participant has no reason to state their true preferences. If they are going to be paid in accord with one of values they produce, the participant can an incentive to say that x_2 is a million dollars because this statement cannot possibly have any negative consequences for them. This is distinct from the binary choice MEL paradigm in which selecting a non-preferred option can only harm the participant.

Setting aside all of these concerns, these three experimental designs account for nearly all experimental work on intertemporal choice (Frederick et al., 2002). In large part, these three designs have been widely used because they directly embody the minimal design required to estimate a parametric model of DU Theory. That said, in practice, the choices observed in a MEL task cannot be interpreted without recourse to several additional assumptions that are generally left implicit by researchers:

1. **No Intertemporal Arbitrage**: Participants have no access to capital markets that would allow them to shift money around in time by either (a) investing in interest-bearing accounts or (b) taking out loans. If participants have access to either savings or loans mechanisms, then MEL tasks should only measure participants' access to capital markets because these financial tools can be used

to profit at the expense of the experimenter (for details, see Frederick et al., 2002; Cubitt and Read, 2007)

2. **Flypaper Consumption**: In order to use the specific indifference equations described in the previous chapter to estimate discount factors, researchers must expand upon the assumption of no intertemporal arbitrage and further assume that receiving money in a MEL task is treated as a direct source of utility. This is equivalent to assuming that the money received in a MEL task is consumed at the time of receipt. This assumption is almost certainly not accurate: if I receive a gift card for $5 today, I can consume that amount any time after today. Because monetary payments could sit idle for many weeks, months or years, the experimenter's meticulous time-dating of goods could be irrelevant to the participant's decisions (but, for evidence for such an assumption, see Hines and R. H. Thaler, 1995)

3. **No Effects of Baseline Wealth**: Finally, the standard analysis of MEL task data requires that we assume that the response of a participant depends only on the utility of a time-dated good relative to $0. If this is not true, then we need to measure the socioeconomic status of every participant and calibrate our calculations to account for their baseline level of wealth. As we will see in a moment, the assumption that baseline wealth is not relevant seems to be plausibly correct based on my work. The assumed irrelevance of baseline wealth for MEL experiments is consistent with Prospect Theory's explicit assumption that baseline wealth is not important to decision-making (Kahneman and Tversky, 1979).

In the rest of this section, I will follow the traditions of the field and make all of the assumptions outlined above. These assumptions make the analysis of data from MEL tasks mathematically tractable. The almost certain falsity of some of these

assumptions may help to partially explain some of the results that I will present, but will generally not be sufficient to account for the majority of my findings. Rather, the problematic assumptions I have just listed are simply an additional set of reasons for doubting the validity of MEL as a paradigm for studying intertemporal choice. With that said, I now move on to elucidating another set of concerns that are empirical rather than conceptual.

The broadest motivation for my experimental work is simple: MEL tasks typically estimate that normal people discount goods by 30%-50% per year or more. From the perspective of DU Theory, these numbers are implausibly large: if people were not willing to wait one year for a good that is twice as large as an immediate good, then naive DU Theory would predict that essentially all banks would go out of business because no customers would be interested in investing in accounts that offer 1%-5% interest per year. (Note that this argument depends upon the domain invariance property of DU Theory highlighted in the previous chapter.) Clearly, most Americans do in fact invest in the low-interest bearing accounts provided by real credit markets. As such, something must be wrong with the discount factors that have been estimated in most MEL tasks. My work has been an attempt to determine precisely what is wrong. With that aim in mind, I will describe a series of eight experiments that collectively provide evidence for robust failures of DU Theory to account for observed choices in MEL tasks.

These eight experiments can be characterized by the types of manipulations they employ, all of which are designed to make questions in MEL tasks appear more like banking decisions. The first experiment presents delayed goods in terms of the effective interest rate that would be received if the later option were chosen. this interest rate information is presented in the annualized-per-year format (APY) that banks typically use to describe bank accounts. This APY information is presented alongside the monetary values of the earlier and later goods that would be shown in

a standard MEL task.

As will be seen, we find that the inclusion of interest rate information has a small effect on behavior. This leads to the manipulations employed in the seven other experiments described. In the second experiment, we (a) removed all mention of the monetary value of the later good and described in solely in terms of the percentage increase it represents over the smaller good and (b) described decisions in terms of deposits into bank accounts. These two changes lead to substantially more patient behavior. In the third experiment, we attempted to assess how much of this effect is driven simply by describing choices in terms of deposites into bank accounts. In the fourth experiment, we attempted to determine how much of this change is determined by (a) describing the later good in terms of its relationship to the earlier good and (b) describing the later good in terms of percentages rather than absolute quantities.

These first four experiments demonstrated that behavior in MEL tasks is extremely labile to changes in the verbal description of goods. In the fifth and sixth experiments, we assessed whether this lability was sufficient to induce internally inconsistent decisions. We did this by returning to the standard MEL format, but varying the units of time used to describe goods. For example, we demonstrated that a good available in 7 days is perceived as less valuable than a good available in 1 week. This effect occured whether participants made purely hypothetical decisions (Experiment 5) or made decisions that would pay out the quantities being decided between (Experiment 6).

The collective demonstrations of extreme lability in responses to MEL questions to changed descriptions suggested that MEL tasks do not measure a stable discount function. In the seventh and eight experiments, we therefore assessed whether MEL tasks measure a property of individuals that is a useful predictor of other behaviors that might reflect impatience. We find that MEL is predictive, but that its predictive power is very small: the median R^2 of MEL measurements of impatience with other

behaviors is 0.005.

Thus, collectively the experiments I will now describe demonstrate that (a) many verbal manipulations that make MEL tasks sound more like banking decisions are able to push behavior towards decisions that are more patient, (b) these manipulations can be understood in terms of their explicit references to banking terminology and in terms of their ability to focus attention on the relative or absolute gains associated with patience, (c) the results of MEL tasks are largely unable to predict other real-world behaviors.

4.2 A MEL Task on a Representative Sample of Americans

Introduction: In this first study, I examined the performance of MEL tasks when extended to a large-scale survey of Americans using the RAND Corporation's American Life Panel (ALP). The ALP is designed to allow social scientists to conduct experiments on a representative sample of Americans. By working with the ALP, I was able to test the extent to which behavior in MEL tasks is correlated with socioeconomic status (SES), gender and age. In particular, an assessment of the role of SES is essential for testing the influence of baseline wealth on intertemporal choice. As mentioned in the previous section, DU Theory predicts that baseline wealth should have a strong effect on intertemporal choices. In addition to measuring correlations between MEL behavior and demographic traits of our participants, I also tested the influence of four framing manipulations on intertemporal choice. These framing manipulations defined four conditions to which participants were randomly assigned. The four conditions were specifically framed as they were in order to elicit progressively greater amounts of patience moving from Condition 1 to Condition 4.

Methods: In Condition 1, I presented standard MEL-like binary choices involving

two time-dated goods. Unlike traditional MEL tasks, I presented the larger, later good as the first option in each binary choice. The larger, later good was presented in a format that made salient both its absolute amount (i.e. $1.33) and its size relative to the earlier good. In particular, its relative size was described in terms of an interest rate that was presented in an annualized per yer (APY) format that is comparable to the format used to describe bank accounts in American credit markets. Thus, a typical question in Condition 1 involved a binary choice like the following: "$1.33 in 48 weeks (40% APY) OR $1.00 in 4 weeks". I expected that the additional APY information would make participants realize that I was offering higher interest rates than were available in normal credit markets, which should lead to a higher frequency of selecting the larger, later goods being offered.

In Condition 2, I presented the same choices as in Condition 1, except that I removed the interest rate information. Thus, a typical binary choice looked like, "$1.33 in 48 weeks OR $1.00 in 4 weeks". I expected that the removal of interest rate information would make participants less likely to choose a delayed option relative to Condition 1.

In Condition 3, I used a similar format as in Condition 2, except that the earlier option was always available in 4 weeks. Thus a typical question looked like, "$1.33 in 48 weeks OR $1.00 in 4 weeks" and all questions ended with the words "4 weeks". Based on the As Soon as Possible hypothesis of Glimcher et al. (Glimcher et al., 2007), I expected that consistently presenting questions in which the earlier option was available in 4 weeks would transform 4 weeks into a surrogate for the present and make participants less likely to choose a delayed option relative to Conditions 1 and 2.

In Condition 4, I used the same questions as were used in Condition 3, except that the earlier option, which was again always available in 4 weeks, was now presented as being available "at no additional delay" and the later option was presented as being

available with "an additional delay". Thus a typical question in Condition 4 looked like following: "$1.33 after 44 weeks additional delay OR $1.00 at no additional delay". I expected that this manipulation would produce a stronger As Soon as Possible effect than in Condition 2 because the suppression of information about the earlier good's timing would mean that 4 weeks of delay would come to seem like a proxy for the present. I hypothesized that this manipulation would make participants less likely to choose a delayed option relative to Conditions 1, 2 and 3.

After assigning each participant to one of these four conditions uniformly at random, the survey presented 400 MEL questions to each participant using an MPL design in which 20 related rows of questions were presented on the screen at one time. Because of anomalies in RAND's condition assignment mechanism as well as participant attrition, I recruited 237 participants in total before reaching my quota of at least 50 participants in each of the four framing conditions. In the end, the enrollment in the four conditions placed 52 participants in Condition 1, 61 participants in Condition 2, 64 participants in Condition 3, and 59 participants in Condition 4.

Results: The results of the ALP Study were inconclusive: an ANOVA revealed that the framing conditions had large and significant effects on behavior as can be seen in Figure 4.1 and Table 4.1 ($F(3) = 365.7$, $p < 2^{-16}$). Nevertheless, these effects did not always operate in the expected directions: I expected that Conditions 1-4 would exhibit decreasing patience across conditions. Although this prediction was true for Conditions 1-3, it was not confirmed in Condition 4, which was in fact the condition in which the greatest patience was observed.

Importantly for future work, no significant effects of income were found: a logistic regression analysis of choice data in which annual income in dollars was included as a covariate produced a coefficient for income of -2.155^{-7}, which was not statistically significantly different from 0 ($z = -0.954$, $p < 0.34$). A second logistic regression analysis that used log income as a covariate also found no significant effect.

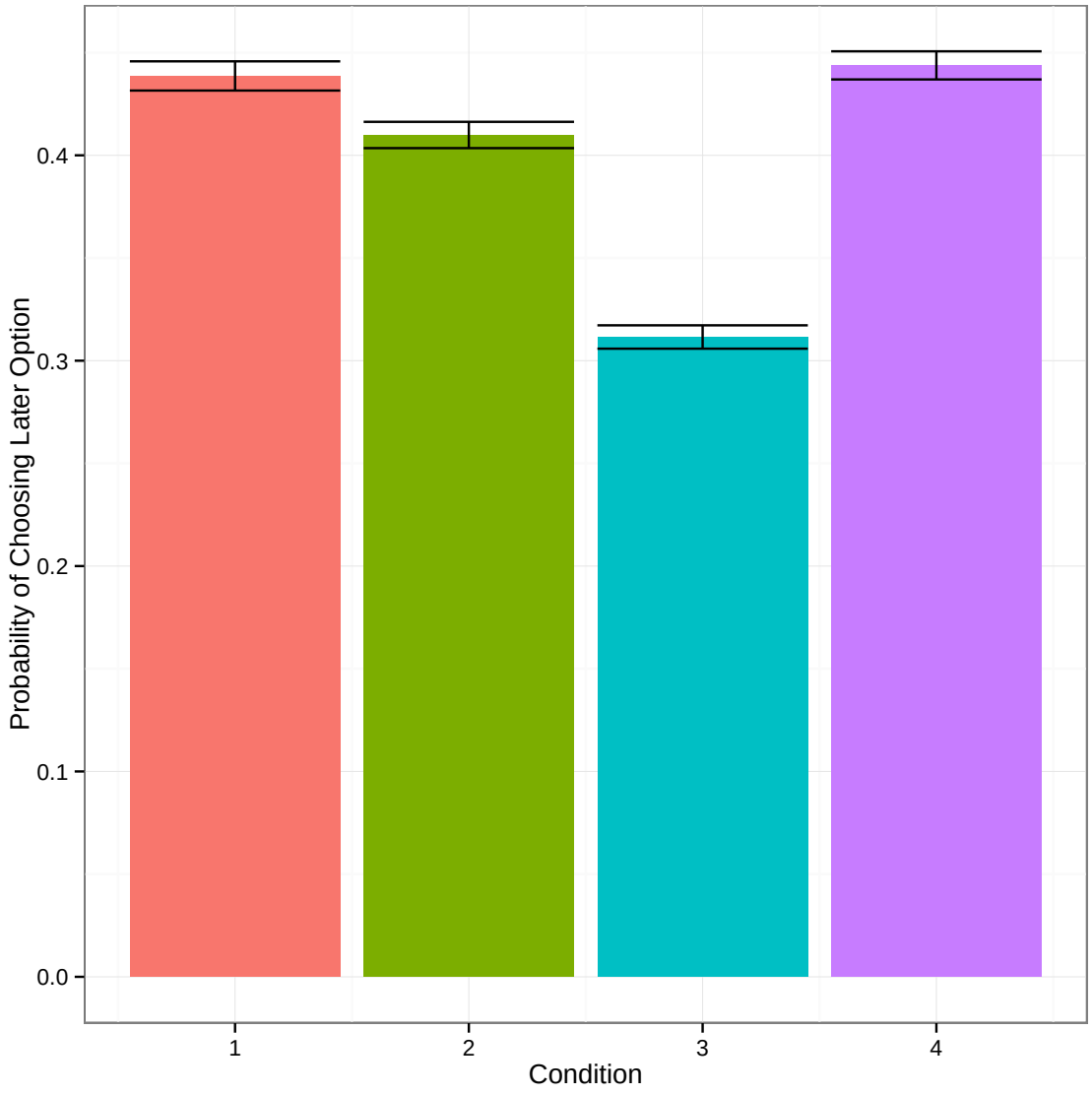


Figure 4.1: The framing manipulations in the ALP study produced strong, but unexpected, effects. (Error bars reflect 95% confidence intervals based on the nonparametric bootstrap.)

Table 4.1: Effects of the framing manipulations employed in the ALP study with 95% confidence intervals

Condition	Frequency of Choosing Delayed Good
1	43.9% [43.2%, 44.6%]
2	41.0% [40.3%, 41.6%]
3	31.1% [30.6%, 31.7%]
4	44.4% [43.7%, 45.1%]

In contrast, an ANOVA revealed a significant gender difference between men and women in which women exhibited greater patience on average than men ($F(1) = 82.64$, $p < 2^{-16}$), although the measured change of 3% increased frequency of selecting delayed goods was very small in substantive terms.

Table 4.2: Evidence of gender differences in the ALP study

Gender	Frequency of Choosing Delayed Good
Female	42.0% [40.5%, 43.4%]
Male	37.9% [37.4%, 38.4%]

Finally, a logistic regression analysis of the relationship between age in years and observed choices revealed a statistically significant effect of age on choice ($z = 11.15$, $p < 2^{-16}$), although the strength of this relationship was extremely weak. Specifically, I estimated a regression coefficient of 0.00515 per year of age, which implied that being 50 years older only increased the probability of selecting a delayed good by at most 6%.

Discussion: Taken as a whole, the results of the ALP study were inconclusive. The only finding that I have pursued in subsequent work was the observation that the inclusion of interest rate information in an APY format increased measured levels of patience. Further examinations of this manipulation are described in the next section. This observation also provided inspiration for the ITCH model that I have developed.

4.3 The Banking Study

Introduction: The ALP study provided weak evidence that the inclusion of interest rate information in an APY format could increase patience. Because observed discounting in MEL tasks is too extreme to be consistent with field behaviors such as the use of low-interest bank accounts, I have conducted several experiments in

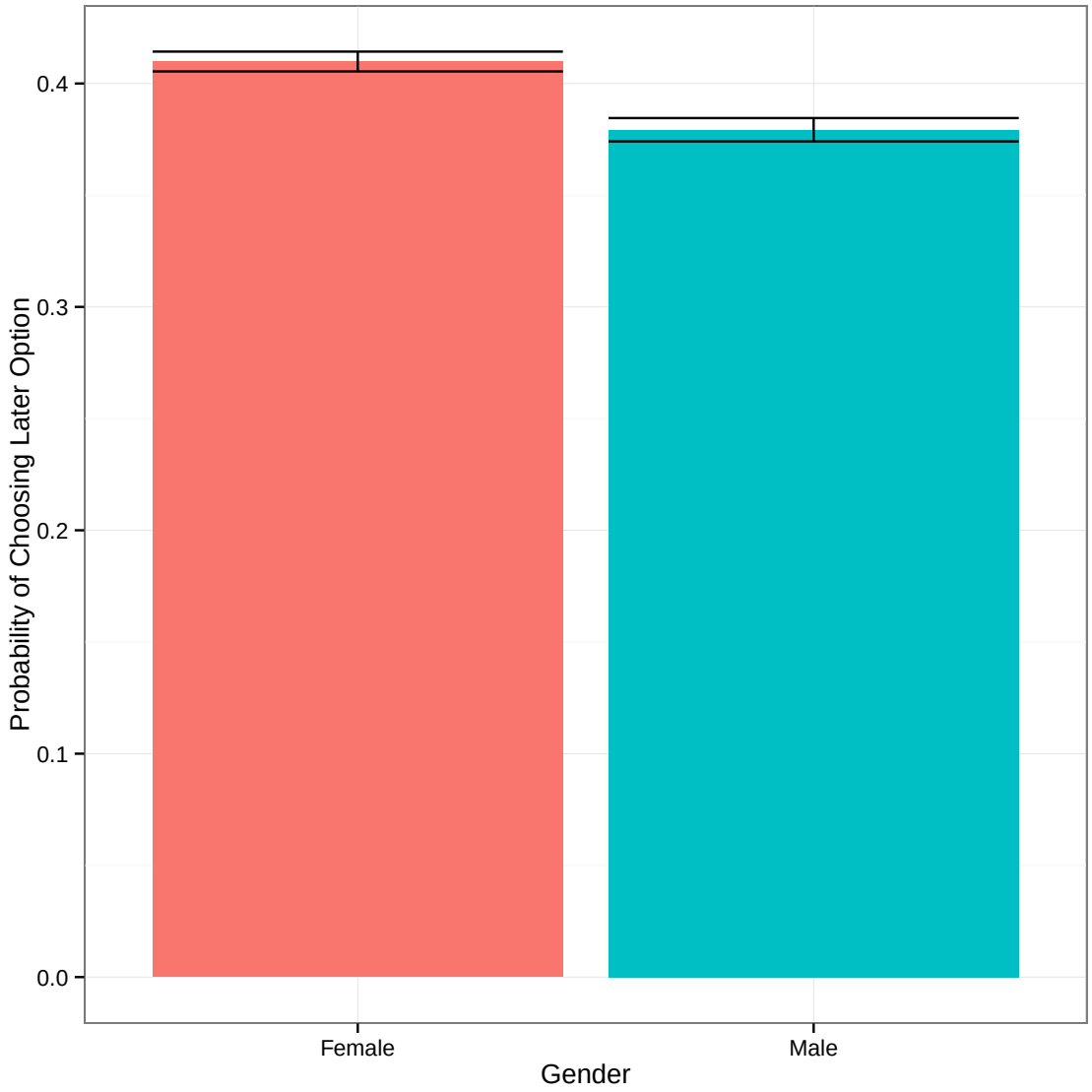


Figure 4.2: In the ALP study, women showed greater patience than men. (Error bars reflect 95% confidence intervals based on the nonparametric bootstrap.)

which I attempted to bring observed discounting closer to the levels predicted by an examination of American credit markets, in which typical accounts offer less than 5% APY interest.

One experiment, which I call the Banking study, explicitly tested if a combination of interest rate information and language that evoked a banking mindset were collectively sufficient to lower observed discounting to the levels that I expected based on field behaviors.

Methods: In the Banking study, I tested two manipulations simultaneously: (1) I presented decisions not in terms of the direct receipt of money, but in terms of depositing money into an investment account; and (2) I presented decisions using only percentage increases relative to the earlier, smaller good rather than the absolute outcomes typicalled used in MEL experiments. By removing the absolute size of the later, larger good, the Banking Study differed from the ALP study in which annualized percentage increases and absolute monetary information had both been present.

To test the manipulations described above, I randomly assigned 100 participants recruited through Amazon's Mechanical Turk, a popular online labor market (Mason and Suri, 2012), to one of two framing conditions, each of which employed a randomized binary choice design in which participants made a sequence of decisions between an earlier, smaller good and a larger, later good.

In Condition 1, participants saw questions like the following:

> Which would you rather have?
>
> $10.00 today
>
> OR
>
> $15.00 in 6 months

I refer to Condition 1 as the "Classical Discounting" condition because the two good are both presented as absolute outcome quantities.

In Condition 2, in contrast, participants saw questions like the following:

Which would you rather have?

Have $10.00 put directly into your checking account today

OR

Have $10.00 invested into a bank account for 6 months that will provide 50% interest over 6 months.

I therefore refer to Condition 2 as the "Banking Frame" condition. Note that the Banking Frame condition combined both (i) a reframing of questions in terms of direct investments into a bank and (ii) the removal of any information about the absolute magnitude of the larger, later good. Unlike traditional MEL tasks, the larger, later good in this condition was defined only relative to the earlier, smaller good. Moreover, this comparison was made using an interest rate rather than an absolute difference.

As noted above, participants in the Banking study were assigned uniformly at random to one of the "Classical Discounting" or "Banking Frame" conditions. Except for this verbal framing manipulation, the conditions were otherwise identical. Importantly, all of the participants in this experiment saw the same questions, although each participant saw these questions in a randomly permuted order. The specific questions I used are shown in Table 4.3.

As can be seen by examining 4.3, the questions I used in the Banking study were designed to assess hypothetical choices over both (a) the small magnitudes typically employed in MEL tasks and (b) a much larger set of magnitudes than are typically used in MEL tasks. Because all choices were hypothetical, it was possible to ask questions about very large magnitudes like $1,000,000. In this thesis, I do not present analyses based on the effects of magnitude, but the main effect of magnitude was large and statistically significant, consistent with previous findings (G. Loewenstein and R. H. Thaler, 1989; G. Loewenstein and Prelec, 1992).

Results: An ANOVA contrasting the "Classical Discounting" and "Banking

Table 4.3: Questions used in the Banking study

($10.00, 0)	($10.10, 6 Months)
($10.00, 0)	($10.50, 6 Months)
($10.00, 0)	($11.00, 6 Months)
($10.00, 0)	($12.00, 6 Months)
($10.00, 0)	($15.00, 6 Months)
($100.00, 0)	($101.00, 6 Months)
($100.00, 0)	($105.00, 6 Months)
($100.00, 0)	($110.00, 6 Months)
($100.00, 0)	($120.00, 6 Months)
($100.00, 0)	($150.00, 6 Months)
($10,000, 0)	($10,100, 6 Months)
($10,000, 0)	($10,500, 6 Months)
($10,000, 0)	($11,000, 6 Months)
($10,000, 0)	($12,000, 6 Months)
($10,000, 0)	($15,000, 6 Months)
($1,000,000, 0)	($1,010,000, 6 Months)
($1,000,000, 0)	($1,050,000, 6 Months)
($1,000,000, 0)	($1,100,000, 6 Months)
($1,000,000, 0)	($1,200,000, 6 Months)
($1,000,000, 0)	($1,500,000, 6 Months)

Frame" conditions revealed a highly statistically significant main effect of condition ($F(1) = 87.65$, $p < 2^{-16}$), which is consistent with the striking 20% increase in the observed odds of selecting a delayed good in the "Banking Frame". The specific results are shown in both Figure 4.3 and Table 4.4.

Table 4.4: Effects of the Banking frame manipulation in the Banking study with 95% confidence intervals

Condition	Frequency of Choosing Delayed Good
Classical Discounting	43.8% [41.0%, 46.6%]
Banking Frame	65.0% [61.5%, 68.4%]

Discussion: The main result of the Banking study was the demonstration that the use of percentage increases to describe the later, larger good in a MEL task could substantially promote measured levels of patience. I have since replicated the main result of this study in five follow-up replications on Mechanical Turk that are not presented here: these replications consistently found that the manipulations that define the Banking frame produce around a 20% change in the probability of selecting a delayed good. Unfortunately, even a 20% increase is not sufficient to reduce observed levels of impatience (and hence discounting) in MEL tasks to the levels seen in American credit markets. I therefore developed a more heavy-handed manipulation in a study that I call the IRA study, which is described in the next section. The IRA study was designed to assess the rate of pure error responding, which I defined as normatively indefensible preferences for earlier options in MEL tasks. As will be seen, my estimate of the error rate is quite high: as high as 30% of responses may reflect errors that can be defended under DU Theory.

4.4 The IRA Study

Introduction: Given that participants reacted strongly to the banking framing ma-

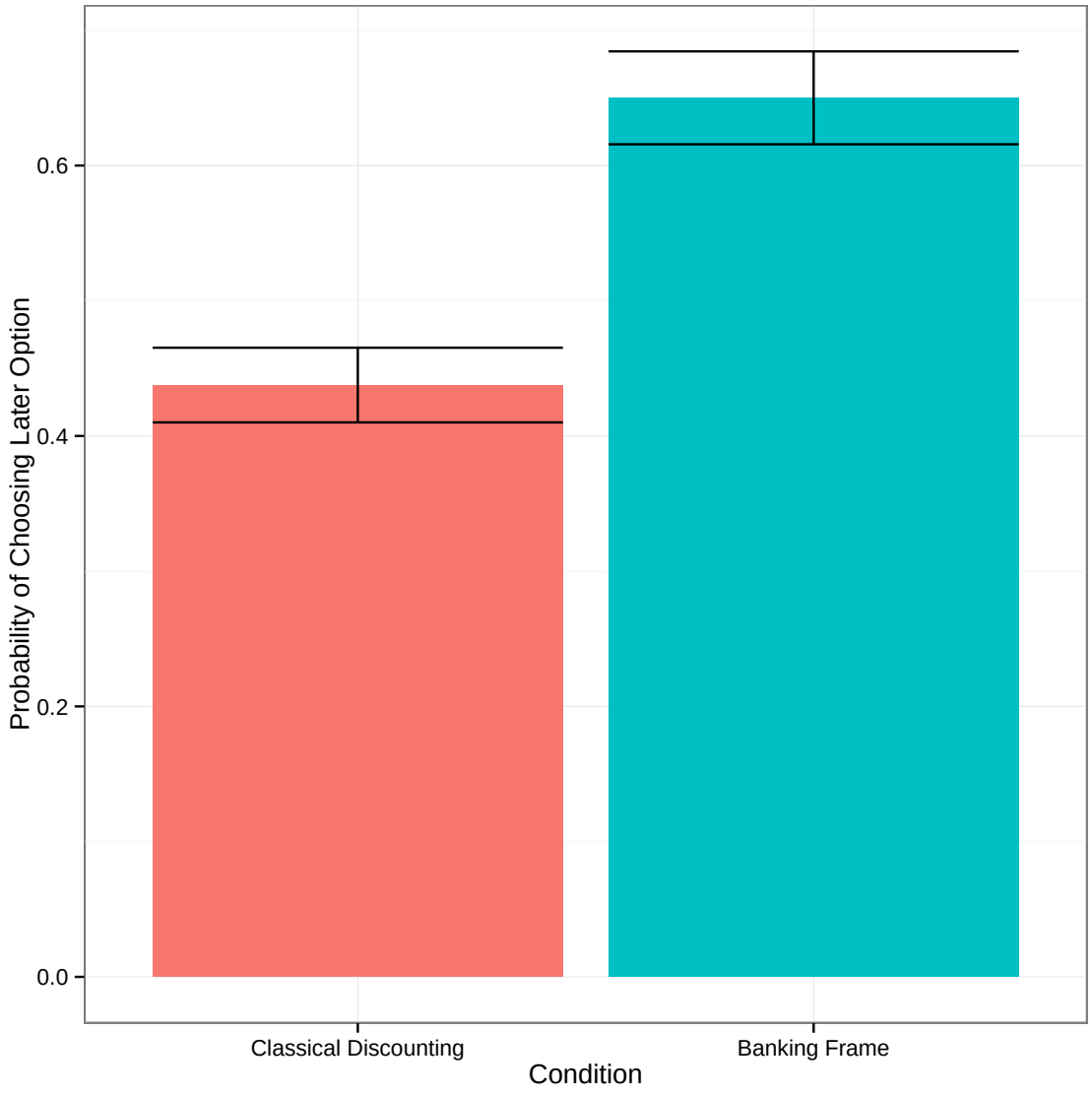


Figure 4.3: In the Banking study, framing choices using banking terminology substantially increased measured levels of patience. (Error bars reflect 95% confidence intervals based on the nonparametric bootstrap.)

nipulation described in the Banking study, but still exhibited excessive impatience, I wondered how I should interpret the anomalously large odds that a participant in a MEL task will select an earlier, smaller good. To answer this question, I developed the IRA study in which I probed the frequency of observing normatively indefensible preferences for the earlier options in MEL-like questions.

In the IRA study, I asked participants to make choices about time-dated deposits of money into their IRA – an account to which they will have no access until they retire. For participants who are not near retirements – and this is the vast majority of MTurk users (Ipeirotis, 2010) –, the long delay before the participant would have access to the funds in an IRA account should imply that the time-dating of any deposits into their IRA will be irrelevant to them. Specifically, for the questions I used in this study, being more than five years away from retirement implies that the ultimate time of consumption of any money I offered participants would be identical for both the nominally earlier and the nominally later options. Thus the choice of a earlier, smaller good over a later, larger good in the IRA study is a choice of a strictly dominated option in which less money is received on the same day on which more money could have been received if the participant had made different decisions. As such, the only possible explanations for participants' preferring the earlier options would be either (a) pure errors in responding or (b) a normatively indefensible preference for what I will call "holding the money in one's own hands". The hypothesis that there is a preference for "control" is consistent with a long-standing concern in the literature that MEL experiments primarily measure the participant's lack of trust in the experimenter rather than actual discounting. Such concerns about trust effects are corrobrated by findings from several groups (e.g. Andreoni and Sprenger, 2010; Epper, Fehr-Duda, and Bruhin, 2011; Benzion, Rapoport, and Yagil, 1989) and may explain the results of the IRA study.

Methods: For the IRA Study, participants were randomly assigned to one of 3

framing conditions. Within each framing condition, participants answered 4 MPL questions, each of which contained 5 rows. In Condition 1, these questions were framed in the "Classical Discounting" language used in typical MEL tasks. Thus, a typical question in Condition 1 looked like the following:

> Which of the following options do you prefer?
>
> $10.00 in 1 year
>
> OR
>
> $12.00 in 2 years

In Condition 2, a distinct set of questions was employed. These questions asked participants to reason about making hypothetical deposits into their retirement account. The nature of the retirement account being considered was explained in the initial instructional phase of the experiment, after which I presented questions like the following:

> Which of the folllowing options do you prefer?
>
> Have $10.00 deposited into your retirement account in 1 year.
>
> OR
>
> Have $12.00 deposited into your retirement account in 2 years.

Thus, in Condition 2, there is no normatively defensible reason for selecting the earlier, smaller good unless the participant will retire within the next 2 years.

In Condition 3, I heavy-handily discouraged selecting the earlier option by adding additional text before participants saw any questions. Specifically, before asking any of the five questions that had been used in Condition 2, I showed participants the following text to "nudge" them into understanding that the nominal dating of deposits into their accounts would not be relevant to their decisions:

> For each of these five pairs of options, please indicate which of the options you prefer. Please remember that you will not be able to withdraw any

money from your retirement account until you are 65 years old. You should also assume that the current return on your retirement account is 3% per year.

After showing the participants in Condition 3 this "nudge" text, they then saw the very same binary choices that were used in Condition 2, an example of which is repeated below for clarity:

Which of the folllowing options do you prefer?

Have $10.00 deposited into your retirement account in 1 year.

OR

Have $12.00 deposited into your retirement account in 2 years.

In total, I recruited 303 participants through Amazon's Mechanical Turk. Each participant was assigned with equal probability to one of the 3 framing conditions. Each of these participants made 20 binary choices in total. Each group of five of these binary choices was linked into a single 4 MPL question involving 5 rows of goods. These four groups of binary choices each offered an increasing sequence of implicit interest rates that varied between 20%, 50%, 100%, 150% and 200% and were shown increasing in that order. The baseline amounts that were offered across the 4 MPL groups were $10, $100, $10,000 and $1,000,000.

Table 4.5: Summary statistics for each of the three conditions used in the IRA study with 95% confidence intervals

Condition	Frequency of Choosing Delayed Good
Classical Discounting	63.3% [61.0%, 65.5%]
IRA Choices	67.1% [65.0%, 69.3%]
IRA Choices with Nudge	72.1% [69.9%, 74.3%]

Results: An ANOVA comparing the three conditions for a main effect of condition found a significant difference across conditions ($F(2) = 15.16$, $p < 2.74^{-7}$.) More

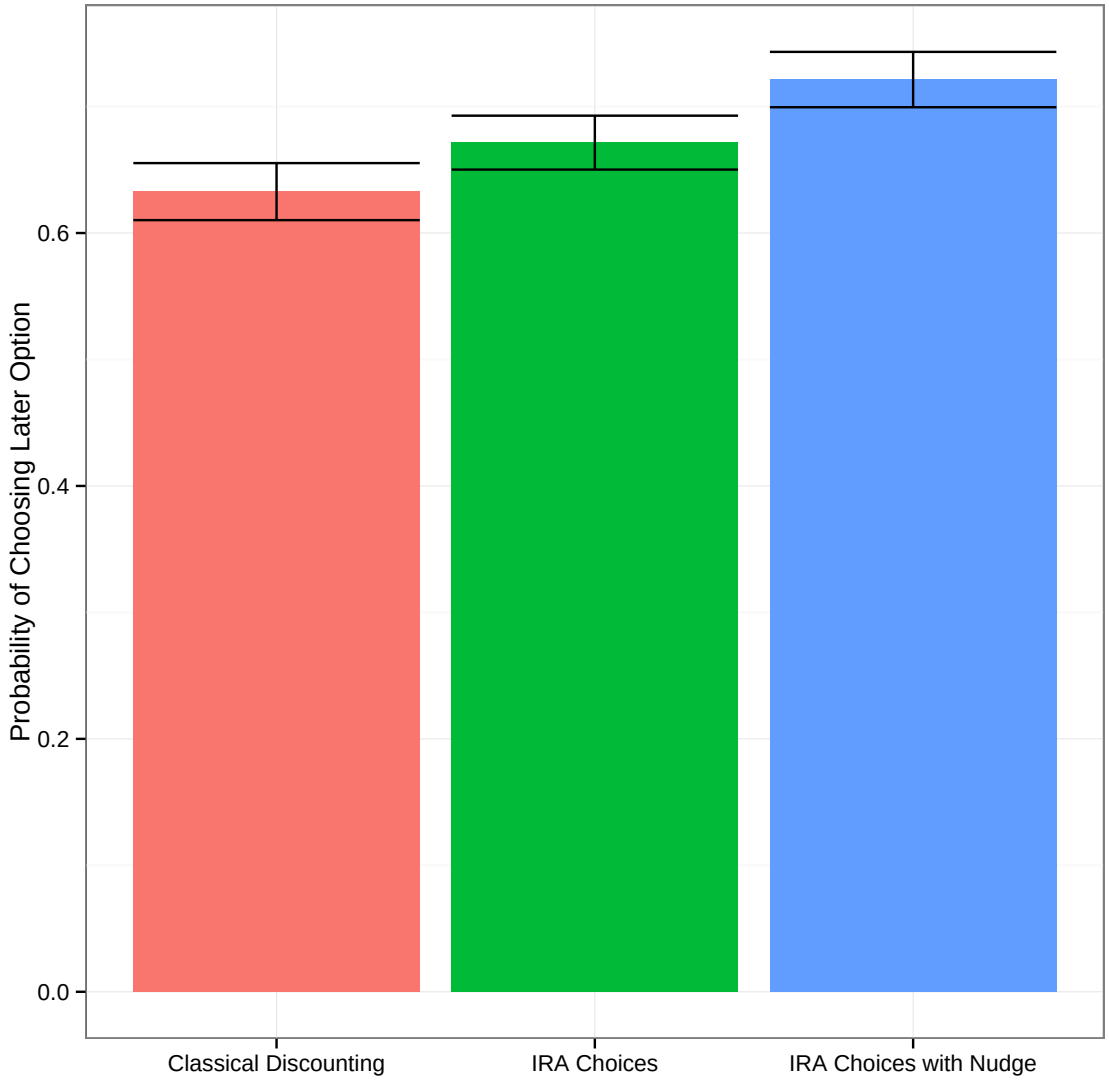


Figure 4.4: In the IRA study, framing decisions in terms of investment increased patient responding, but still produced a large proportion of normatively indefensible choices for the earlier, smaller good. (Error bars reflect 95% confidence intervals based on the nonparametric bootstrap.)

interestingly, Figure 4.4 and Table 4.5 both demonstrate that at least 30% of choices were made in favor of strictly dominated options. Moreover, the fact that decisions in favor of the earlier options can be seen as errors is confirmed by the fact that a reminder that (a) IRA's do not allow withdrawls until retirement at age 65 and that (b) real-world interest rates are much lower than 20% was able to substantially decrease selection of the dominated option – although this "nudge" was not sufficient to remove such choices entirely. Whether the remaining preference for dominated options is a result of a drive to "control" one's own money is a question for future research.

4.5 The Relative and Absolute Gains Study

Introduction: The IRA study found that participants exhibited irrational impatience even when this impatience could not be defended in terms of heterogeneity of tastes (Tabuchi and Thisse, 2002). In other words, manipulations of banking terminology alone were insufficient to remove the excessive levels of impatience seen in MEL tasks. To what extent can the excessive amount of discounting be explained in terms of idiosyncrasies of numeric processing? The interest rate manipulation introduced I employed in the ALP, Banking and IRA studies collectively suggested that focusing attention on the relative gains from waiting (articulated as percentage increases) could make MEL participants more likely to choose delayed options. This led to a natural question: to what extent was this effect merely an artifact of presenting the delayed option as a gain with respect to the earlier good?

Methods: To answer that question, I conducted a study that I will refer to as the Relative and Absolute Gains study. For this study, I recruited 600 participants through Amazon's Mechanical Turk and randomly assigned them to one of 3 conditions in which 25 normatively equivalent questions were framed using three types of

language.

Specifically, the conditions were framed to highlight the monetary gains associated with choosing the delayed good. In Condition 1, which I will call the "Absolute Gains Frame," a typical question looked like a choice between "$10.00 today or $10.00 plus an additional $5.00 in 2 weeks." I will refer to $5.00 in such a question as the **absolute gain** from waiting for 2 weeks. In Condition 2, which I will call the "Relative Gains Frame," a typical question looked like a choice between "$10.00 today or $10.00 plus an additional 50% in 2 weeks." I will refer to 50% as the **relative gain** from waiting for 2 weeks. Finally, in Condition 3, which I will call the "Standard MEL Frame," a typical question looked like "$10.00 today or $15.00 in 2 weeks." In this condition, as in standard MEL tasks, no attempt was made to highlight the gains associated with waiting. My hypothesis was that participants in MEL tasks treat waiting as a loss that the countervailing monetary gains must overcome. By making these gains explicit, I attempted to induce greater odds of selecting the delayed goods that were available.

For each participant, I randomly generated 25 MEL questions that were designed to make the absolute and relative gains orthogonal across questions. To do this, I drew an absolute gain, a, uniformly at random from the list ($0.10, $1.00, $10.00, $100.00, $1000.00). This absolute gain was then combined with a relative gain, r, drawn uniformly at random from the list (1.01, 1.05, 1.10, 1.25, 1.50, 2.00, 3.00, 4.00), whose elements represent the multiple of the earlier, smaller good that would be presented as the larger, later good. Together a pair of absolute gain and relative gain values uniquely define a smaller and larger amount of money: I computed this unique pair of monetary amounts by calculating $x_1 = a/(r-1)$ and $x_2 = x_1 + a$.

To generate the times of receipt for these pairs of monetary values, I drew the front-end delay, t_1, uniformly at random from the list (0 months, 1 month, 2 months). An additional delay of 1, 2 or 3 months was added to the front-end delay, t_1, to determine

the later date, t_2. Collectively these randomly selected four numbers, (x_1, t_1, x_2, t_2), defined the earlier, smaller good and the later, larger goods used in each question. The reason for selecting absolute and relative gains that were orthogonal will become clear when I discuss the ITCH model of intertemporal choice in the next chapter.

Results: An ANOVA analysis found a highly significant main effect of condition in this task ($F(2) = 260$, $p < 2^{-16}$). The specific results for each condition separately are shown graphically in Figure 4.5 and numerically in Table 4.6.

Table 4.6: Framing decisions in terms of relative and absolute gains substantially increased the odds of selecting a delayed good

Condition	Frequency of Choosing Delayed Good
Absolute Gain Frame	38.7% [37.3%, 40.0%]
Relative Gain Frame	42.6% [41.2%, 44.0%]
Standard MEL Frame	22.9% [21.8%, 24.0%]

Discussion: These results suggested that at least part of the effect seen in my earlier manipulations of interest rates was simply the result of making the gains from waiting more salient to participants in a MEL task. But the still greater level of patience seen in the "Relative Gain Frame" over the "Absolute Gain Frame" also suggested that the use of percentage information further heightens levels of observed patience above and beyond the effect attributable to the heightened salience of monetary gains for waiting. Indeed, the results of a follow-up to the Relative and Absolute Gains study, which I call the Gains vs. Losses study and report in the next chapter, suggest that the Relative and Absolute Gains study underestimated how much additional patience was elicited by using relative instead of absolute gains. The Gains vs. Losses study provided the data I will use to calibrate the ITCH model that I will describe in the next chapter.

Taken as a whole, the results of the Relative and Absolute Gains study provided particularly strong evidence that merely shifting the participant's attention in a MEL

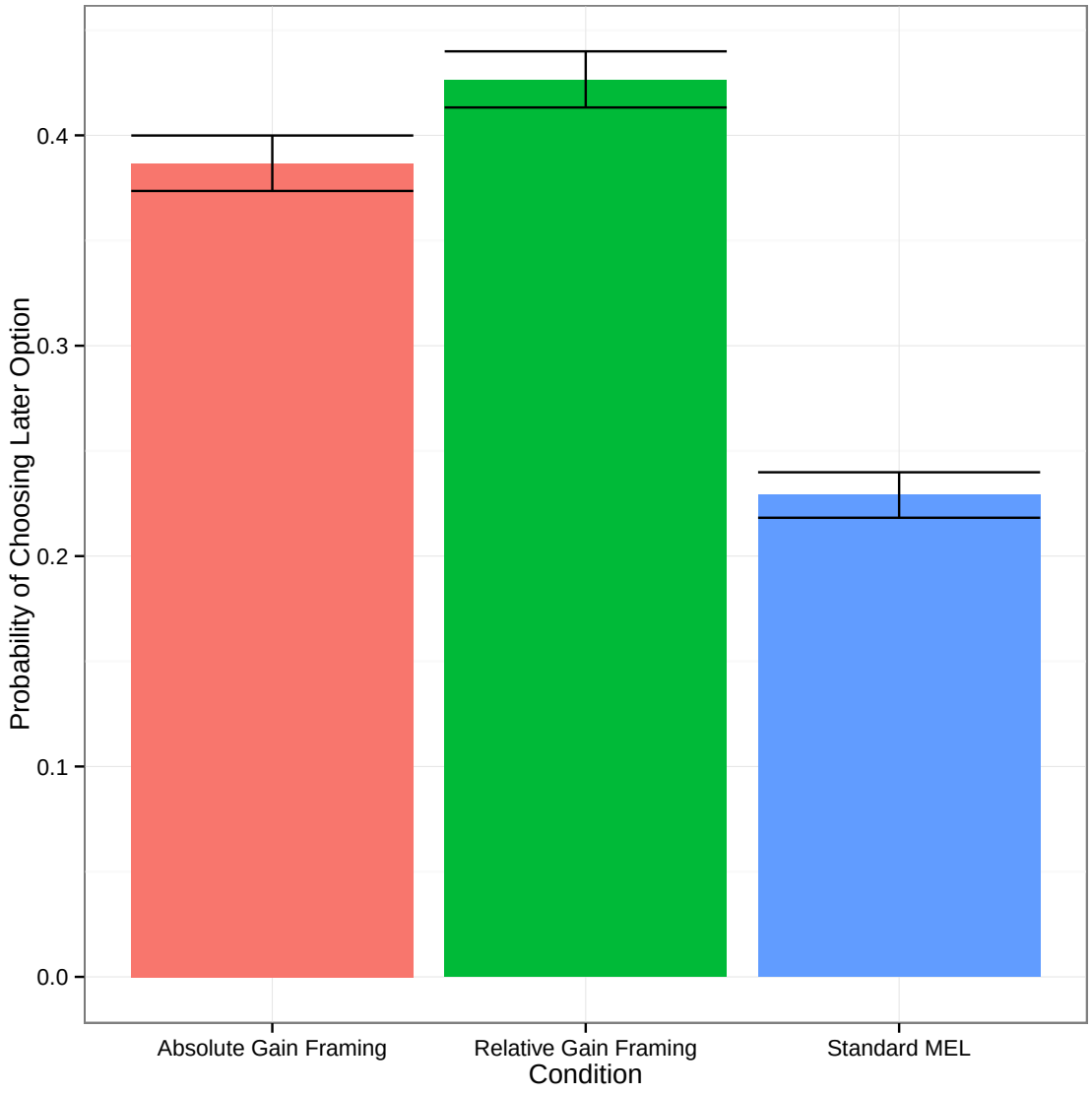


Figure 4.5: In the Relative and Absolute Gains Study, framing manipulations substantially increased patient responding. (Error bars reflect 95% confidence intervals based on the nonparametric bootstrap.)

task towards the specific numeric representation of the gains from waiting could increase patient responding. The next studies, which I call the Time Frame studies, extends this second point.

4.6 The First Time Frame Study

Introduction: Collectively, the results of the previous experiments suggested that MEL experiments produce estimates of the DU Theory-inspired discount function that are highly malleable to the specific verbal descriptions of time-dated goods. These results are consistent with the extensive literature on framing in psychology and behavioral economics in which choices are shifted by employing alternative verbal descriptions of a situation (Tversky, Kahneman, et al., 1981). A question remained: to what extent could minor variations of the standard MEL experiment – variations that do not introduce new types of language, but simply segment existing variation into classes – also exert large effects on behavior?

To answer this question, I developed the Time Frame paradigm. In the two Time Frame studies I will present, I did not introduce novel verbal descriptions of standard MEL questions. Instead, I identified existing variations from the literature and demonstrated that two different surveys with identical framing, but which employed distinct units of time, produced irreconciably different estimates of the participants' discount functions.

Specifically, the two Time Frames studies I will present demonstrated that participants' behavior in MEL tasks violates a minimal requirement for systematic discounting: decisions between time-dated goods should not be affected by the units of time (e.g. weeks, months) used to describe those goods, holding fixed the absolute time of reward delivery. For example, a reward delivered in one week should be equivalent in value to a reward delivered in seven days. In other words, the time units (days,

weeks, or months) used to describe a good should not matter, as long as the absolute time of its receipt is held constant. I refer to this assumption as the Axiom of Unit Invariance.

The first Time Frame study provided evidence for systematic and extreme violations of the Axiom of Unit Invariance. The second Time Frame study demonstrated that these violations persisted even when participants were given monetary incentives to respond in accordance with the Axiom of Unit Invariance.

Methods: The first Time Frame experiment explicitly tested whether participants make decisions that obey the Axiom of Unit Invariance. 750 participants were recruited through Amazon's Mechanical Turk online labor market. Each of these participants was randomly assigned to one of 3 conditions. These 3 conditions differed in the units of time used to describe the delivery of time-dated goods: Conditions 1-3 used units of days, weeks, and months respectively.

In all three conditions, participants saw a sequence of 14 binary hypothetical choices between an immediate reward and a delayed reward presented in an MPL format. Each choice asked the participant whether they would prefer to:

> Receive \$10 today
>
> OR
>
> Receive \$15 in T [unit of time, e.g., days]

Every binary choice was between a smaller, immediate reward and a larger, delayed reward to be received at $T > 0$ time units, where T ranged from 1 to 14 in increasing order. Each participant was assigned to a condition that used the same units of time in all 14 choices faced by that participant. Thus, each participant saw only one unit of time, with different participants seeing different units across conditions. Figure 4.6 shows an example of how these choices were presented in Condition 1, for which choices were presented using units of days. Conditions 2 and 3 were identical except that the word "day" was replaced with "week" or "month" in every question

(i.e. the numerals remained the same). After answering all 14 questions, participants were given an authorization code that was used to confirm the completion of the experimental task.

Results: The Axiom of Unit Invariance predicts that when two choices refer to the same points in absolute time, the units of time that describe those choices (e.g., days or weeks) should not affect the decision. I found substantial deviations from this prediction. I examined the three choices that referred to the same points in absolute time across conditions: this corresponded to questions about 7 days vs. 1 week, 14 days vs. 2 weeks and 4 weeks vs. 1 month. In each of these three cases, I rejected the null hypothesis that behavior obeyed the Axiom of Unit Invariance:

- 50% of participants chose "\$15 in 7 days" over "\$10 today," whereas 75% chose "\$15 in 1 week" over "\$10 today" ($t = -6.15$, $df = 513.9$, $p < 10^{-8}$).
- 29% of participants chose "\$15 in 14 days" over "\$10 today," whereas 59% chose "\$15 in 2 weeks" over "\$10 today" ($t = -7.01$, $df = 492.7$, $p < 10^{-11}$).
- 38% of participants chose "\$15 in 4 weeks" over "\$10 today," whereas 53% chose "\$15 in 1 month" over "\$10 today" ($t = -3.66$, $df = 520.5$, $p < 10^{-3}$).

After aggregating all of our participants' 14 choices, I used statistical interpolation techniques (Hastie and Tibshirani, 1990) to estimate the probability that each participant would choose a delayed offer of \$15 to be received in T days over an immediate offer of \$10. These estimates were computed separately for the three conditions and are shown in Figure 4.7. The Axiom of Unit Invariance predicts that these probabilities should be identical across all of these conditions, implying that the condition-specific curves in Figure 4.7 should be superimposed for the same values of absolute time. This superposition was clearly not observed: a likelihood ratio test that compared (i) a model in which only one choice curve existed across conditions

Please mark the box on the left in each row of this table if you would prefer the lefthand option over the righthand option. Alternatively, mark the box on the right if you would prefer the righthand option.

Suppose you had a choice between the following pairs of options. Which would you prefer?

Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 1 day
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 2 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 3 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 4 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 5 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 6 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 7 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 8 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 9 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 10 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 11 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 12 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 13 days
Pick one of these two:	☐ Receive $10 today	or	☐ Receive $15 in 14 days

Next Page

Figure 4.6: The MPL format used in the first Time Frame study

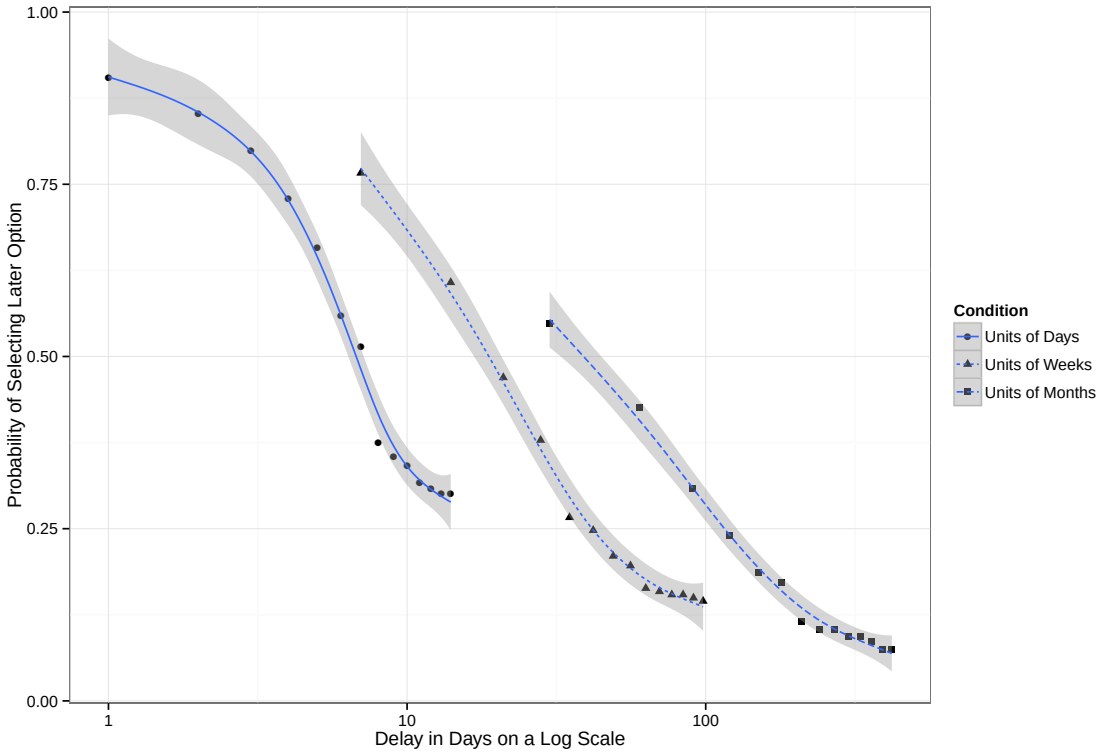


Figure 4.7: Choice probabilities indicate systematic violations of the Axiom of Unit Invariance. (Shaded regions indicate 95% theoretical confidence intervals.)

with (ii) a model in which the choice curves differed by a constant across conditions was highly significant (Chisq = 429.5, df = -2, $p < 10^{-14}$).

4.7 The Second Time Frame Study

Introduction/Methods: In order to confirm that the results of the first Time Frame study were not anomalies introduced by the hypothetical nature of the choices in that experiment, I conducted a second study in which participants were given monetary incentives instead of making hypothetical choices (Grether and Plott, 1979). For this second experiment, I recruited 500 participants from the same online labor market used in the first Time Frame study and randomly assigned each of these participants to one of two conditions: Days and Weeks. Except for the presence of monetary

incentives, the Days and Weeks conditions were identical to those in the first Time Frame study. In both the Days and Weeks conditions, I incentivized responding by telling participants that there was a 1-in-20 chance that one of their choices would be paid to them in the form of a bonus payment that would be received in the amount and on the date indicated in one of their choices, which would be chosen at random by a computer at the end of the experiment. Specifically, I showed participants the following instructions:

> Please take all of these decisions seriously, because one of them might count. Your name will be entered into a lottery and if you win (you have a 1 in 20 chance!), we will reward you by randomly picking one of the questions that you answered on this survey and giving you the option that you chose on that question. In other words, any one of your choices could be paid out for real.

After the task was completed, I paid participants using Mechanical Turk's bonus payment mechanism. This payment was made midday on the date indicated by the participant's choice.

Results: The results of the second Time Frame study were consistent with the results of the first Time Frame study:

- 47% of participants chose "\$15 in 7 days" over "\$10 today," whereas 67% chose "\$15 in 1 week" over "\$10 today" ($t = -4.09$, $df = 418.5$, $p < 10^{-4}$).
- 27% of participants chose "\$15 in 14 days" over "\$10 today," whereas 54% chose "\$15 in 2 weeks" over "\$10 today" ($t = -5.71$, $df = 402.5$, $p < 10^{-6}$).

The second Time Frame study therefore demonstrated that the violations of the Axiom of Unit Invariance seen in the first Time Frame study were not attributable to the hypothetical nature of the choices made in the first Time Frame study. Taken

together, the results of the first and second Time Frame studies call into question the Axiom of Unit Invariance and, as a consequence, the hypothesis that people exhibit any form of systematic discounting in making intertemporal decisions about money in MEL tasks.

4.8 The External Validity of MEL Tasks

All of the experiments reported thus far provide evidence that MEL tasks do not measure discounting of the sort described by DU Theory. The Time Frames experiments go further and raise doubts about whether MEL tasks measure any form of systematic discounting at all. The Axiom of Unit Invariance is far more fundamental to a notion of coherent time preferences than any of the assumptions of DU Theory that were questioned by previous research: robust violations of the Axiom of Unit Invariance are hard to reconcile with any theory based on the assumption that time has an absolute, unit-free meaning to people when they make decisions.

A question remains, though: what are MEL tasks measuring? The correlations between MEL and behaviors that seem to be linked to impatience are robust (Chabris, Laibson, Morris, Schuldt, and Taubinsky, 2008), albeit very small. Where do these correlations come from if they do not emerge from the kind of discounting described by DU Theory?

I will give one possible answer to the question of what MEL measures in the next chapter on the ITCH model, but I have another line of experimental work to discuss beforehand. In this additional line of experiments, I sought to determine the extent to which MEL tasks generate a measure of individual differences in impatience that is externally valid in the sense that the measure is predictive of field behaviors that *prima facie* seem to reflect a person's level of impatience.

To address this question, I conducted two experiments that I call the BIS/BAS

and R^2 studies. In these two studies, I asked participants to complete a brief MEL task as well as a battery of questions about their personality. These personality questions were derived from two sources: the BIS/BAS (Carver and White, 1994) and a list of personality questions used in a paper on the external validity of MEL that was written by Chabris et al. (2008), who used the questions in a study which demonstrated that measures of discounting elicited using MEL were predictive of several field behaviors that, prima facie, seem to reflect impatience. All of the questions I used in these studies are taken from this Chabris study and are listed at the end of this chapter. In the BIS/BAS study, I sought to test whether MEL is a better predictor of field behaviors than standard psychometric assessments of impatience such as the BIS/BAS. In the R^2 study, I sought to test whether the manipulations of MEL shown to increase patience in the previous sections of this chapter could also increase the correlations between estimates of discounting derived from MEL questions and several field behaviors that seem to reflect generalized impatience. These two studies are described in detail below and expand upon the original study by Chabris, which focused entirely on determining whether MEL tasks were predictive of field behaviors. We employed the questions used by Chabris for consistency with the previous literature, but I make no claims about the general validity of these questions as a proxy for impatience.

4.8.1 The BIS/BAS Study

Introduction: In this study, I treated the BIS/BAS questionnaire as a potential alternative to MEL's widespread use as a general purpose measure of impatience. Specifically, I compared the external validity of the BIS/BAS and a standard MEL task by computing the correlation between (a) the measures of impatience provided by both tasks and (b) the Chabris et al. questions, each of which measured a field behavior that was putatively related to impatience.

Methods: For the MEL task, I used a set of MEL questions in which the earlier, smaller good was one of the values in the list ($100, $200, $300, $500). For the $100 and $200 questions, the larger, later good was either $2, $4, $6, $8, $10, $12, $14, $16, $18, or $20 larger than the earlier, smaller good. In contrast, for the $300 and $500 questions, the larger, later good was either $5, $10, $15, $20, $25, $30, $35, $40, $45 or $50 larger than the earlier, smaller good. The timing of these goods always provided an immediate receipt for the earlier good and a delay of 1 year for the later good. In total, the full set of options was presented as 4 questions, each of which involved 10 rows of related delayed goods presented in an MPL design format.

Results: After collecting data from all participants, I constructed several predictors of impatience. The probabilities of selecting delayed goods in the MEL task were treated as the MEL-derived indices of patience for each participant. For the BIS/BAS, I constructed predictors by segmenting the BIS/BAS results into four components: the three components of the BAS (Drive, Fun-Seeking and Reward-Responsiveness) and the entire BIS as the fourth predictor. To determine the predictive power of each of these predictors, I ran a separate linear regression for each pair of (a) a predictor and (b) a field behavior of interest measured by one of the Chabris questions. I then averaged the R^2 values across all of these separate regressions to compute a general measure of predictive power. As seen in Figure 4.8 and Table 4.7, this average R^2 measure suggested that MEL questions as well as all of the sections of the BIS/BAS had very low predictive power as general-purpose measures of impatience. Moreover, the results suggested that MEL was not necessarily the best of the available predictors.

4.8.2 The R^2 Study

Introduction: In the R^2 study, I sought to determine whether any of the manipulations of MEL described earlier in this chapter would improve MEL's external validity

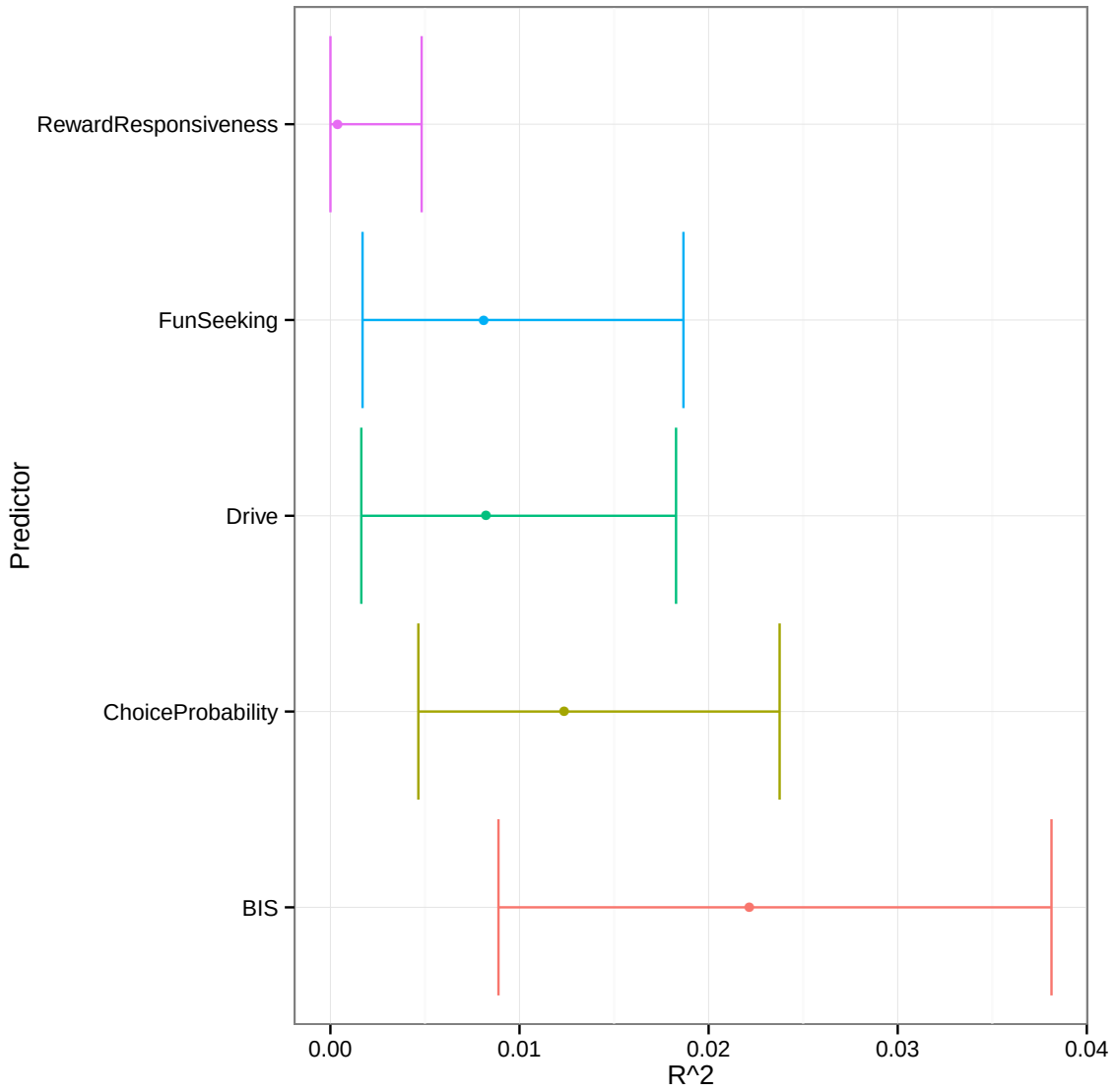


Figure 4.8: The average predictive power of the BIS/BAS components and MEL measured using R^2. (Error bars reflect 95% confidence intervals based on the non-parametric bootstrap.)

Table 4.7: The average predictive power of the BIS/BAS components and MEL

Predictor	Average R^2 with 95% Confidence Interval
MEL Measure of Patience	0.0124 [4.96e-03, 0.0236]
BAS 1: Drive	0.00824 [1.90e-03, 0.0180]
BAS 2: Fun-Seeking	0.00811 [1.69e-03, 0.0200]
BAS 3: Reward-Responsiveness	0.000393 [7.87e-07, 0.00479]
BIS	0.0222 [1.01e-02, 0.0398]

and produce higher R^2 values than I had found in the BIS/BAS study. Given that the BIS/BAS seemed to potentially outperform MEL and yet still performed quite badly as a predictor of field behaviors, I ran this second experiment in which I tested several variants of a MEL task and then compared the results for each of these distinct versions of MEL against the Chabris et al. questions. To collapse these field behaviors into a single measure of impatience instead of conducting many separate regressions, I used a weighted sum of the responses to all of these questions. In this weighted sum, the responses questions were either (a) weighted positively and assigned a weight of $+1$ if they were thought to reflect impatience or (b) weighted negatively and assigned a weight of -1 if they were thought to reflect patience. I treated this sum as a generic index of impatience that MEL tasks were asked to predict.

Methods: In total, I recruited 800 participants through Amazon's Mechanical Turk. These participants were assigned randomly to one of 8 conditions. These 8 conditions were defined by fully crossing 3 factors. The first factor, which I call the "Percentages" factor, determined whether the later, larger good in the MEL task was presented as an absolute dollar amount or as a percentage increase relative to the earlier, smaller good. The second factor, which I call the "Bank" factor, determined whether the MEL task was described in terms of deposits into a bank account rather than as direct money transfers. The third factor, which I call the "Magnitude" factor, determined whether the MEL task employed goods on the order of \$100 or on the order of \$10,000 dollars. All three factors were known from previous work to increase measured patience. Given that field behaviors suggest that people are more patient than would be predicted by standard MEL tasks, I hypothesized that increased patience in MEL tasks might produce a metric of patience that corresponded more closely to field behaviors that involve patience.

Thus, by testing several variants of the MEL experiment, I sought to determine if any of these 8 variants had a reliably higher correlation with field behaviors than

the standard MEL experient. As will be seen, no significant differences were found between conditions as indices of impatience, although each of the 3 factors we varied had a large and significant main effect on behavior in the MEL task.

I used the same set of MEL questions used in the BIS/BAS study in which the earlier, smaller good could be any of the values from the list ($100, $200, $300, $500). For the $100 and $200 questions, the larger, later good was either $2, $4, $6, $8, $10, $12, $14, $16, $18, or $20 larger. For the $300 and $500 questions, the larger, later good was either $5, $10, $15, $20, $25, $30, $35, $40, $45 or $50 larger. The timing of these goods was always an immediate receipt for the earlier good and a delay of 1 year for the later good. In total, these options were presented as 4 MPL format questions, each of which involved 10 rows of options. After each participant completed the MEL questions, I computed the probability that they selected the larger, later good and used this probability as a MEL-derived measure of patience. This MEL score was compared against the weighted sum of the Chabris et al. questions using linear regression analysis.

Results: Figures 4.9, 4.10 and 4.11 depict the main effects of the "Percentages" factor, "Bank" factor and "Magnitude" factor. An ANOVA analysis confirmed that each of these factors produced a significant main effect on behavior ("Percentages" factor: $F(1) = 27.91$, $p < 1.64^{-7}$; "Bank" factor: $F(1) = 10.45$, $p < 0.00128$; "Magnitude" factor: $F(1) = 18.85$, $p < 1.6^{-5}$).

Having determined that the factors I manipulated across variants of the MEL task were effective in changing behavior, I assessed whether any these variants was more or less predictive of field behaviors by computing a linear regression in which the Chabris et al. impatience scores were regressed against the MEL-derived measure of patience. An examination of the resulting R^2 values calculated for each of the 8 MEL variants found no reliable effects of any of the manipulations in increasing the external validity of MEL. Even in the best variant of MEL, the R^2 was only 0.14; in

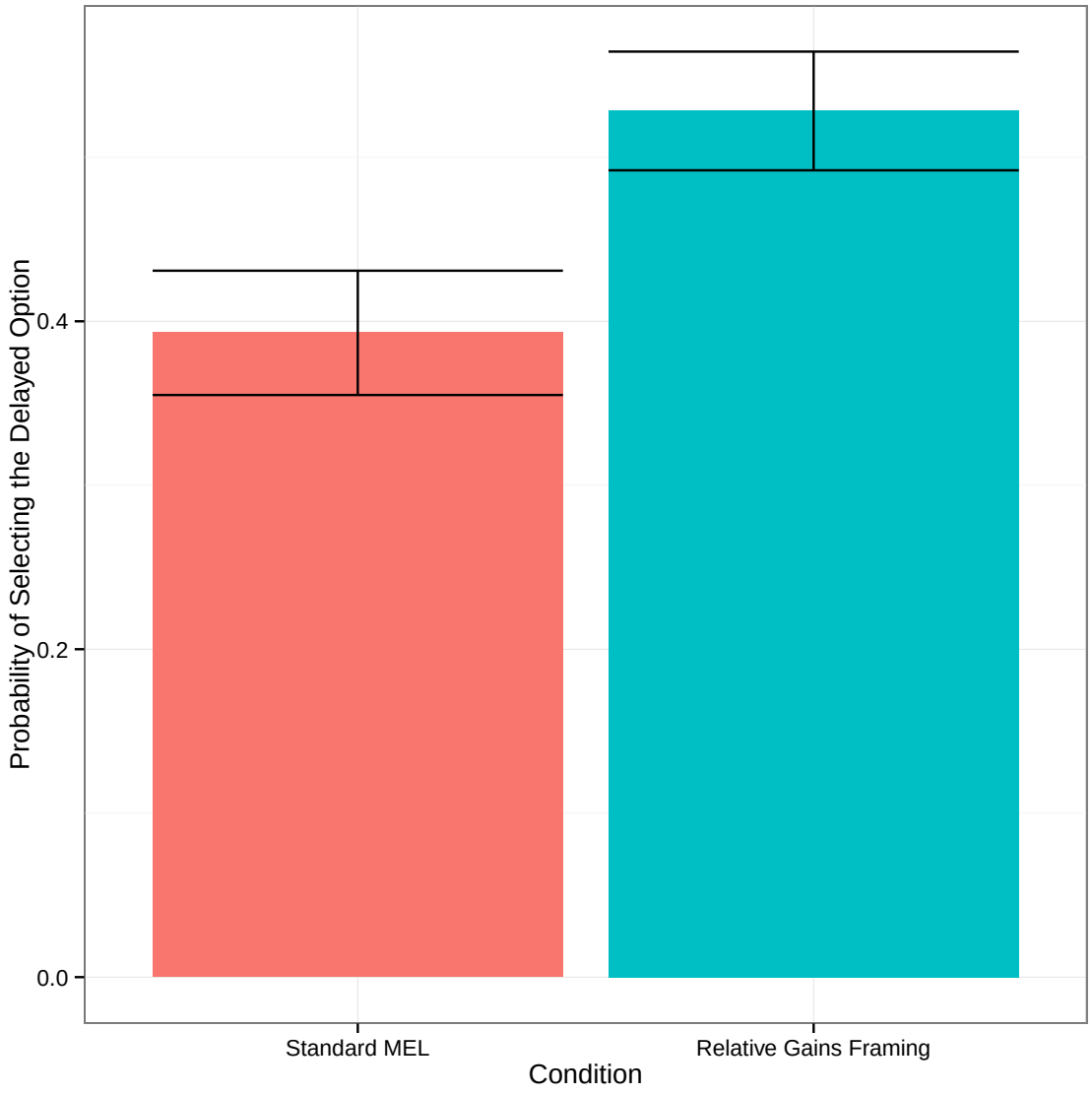


Figure 4.9: Main effect of the Percentages factor in the R^2 study. (Error bars reflect 95% confidence intervals based on the nonparametric bootstrap.)

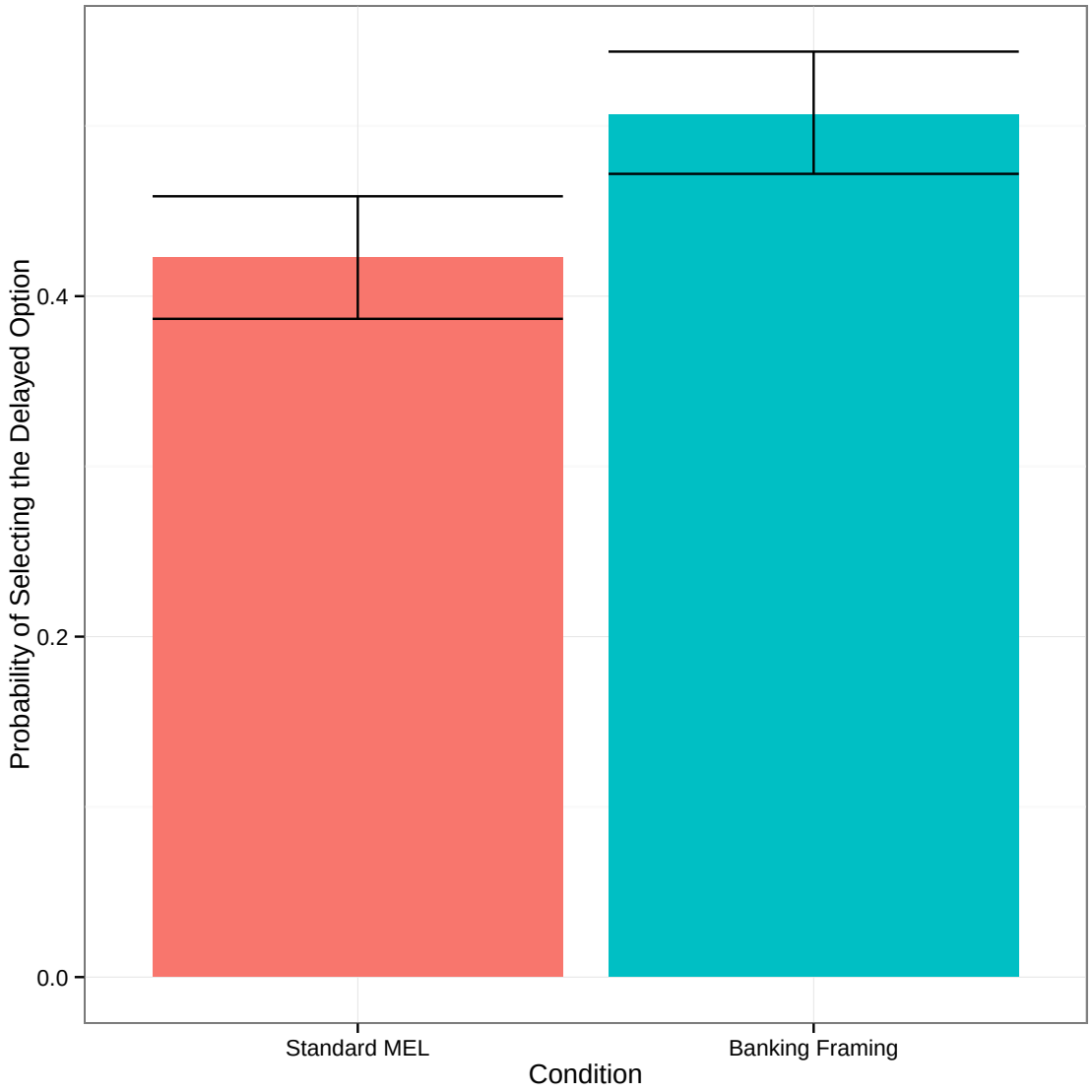


Figure 4.10: Main effect of the Banking factor in the R^2 study. (Error bars reflect 95% confidence intervals based on the nonparametric bootstrap.)

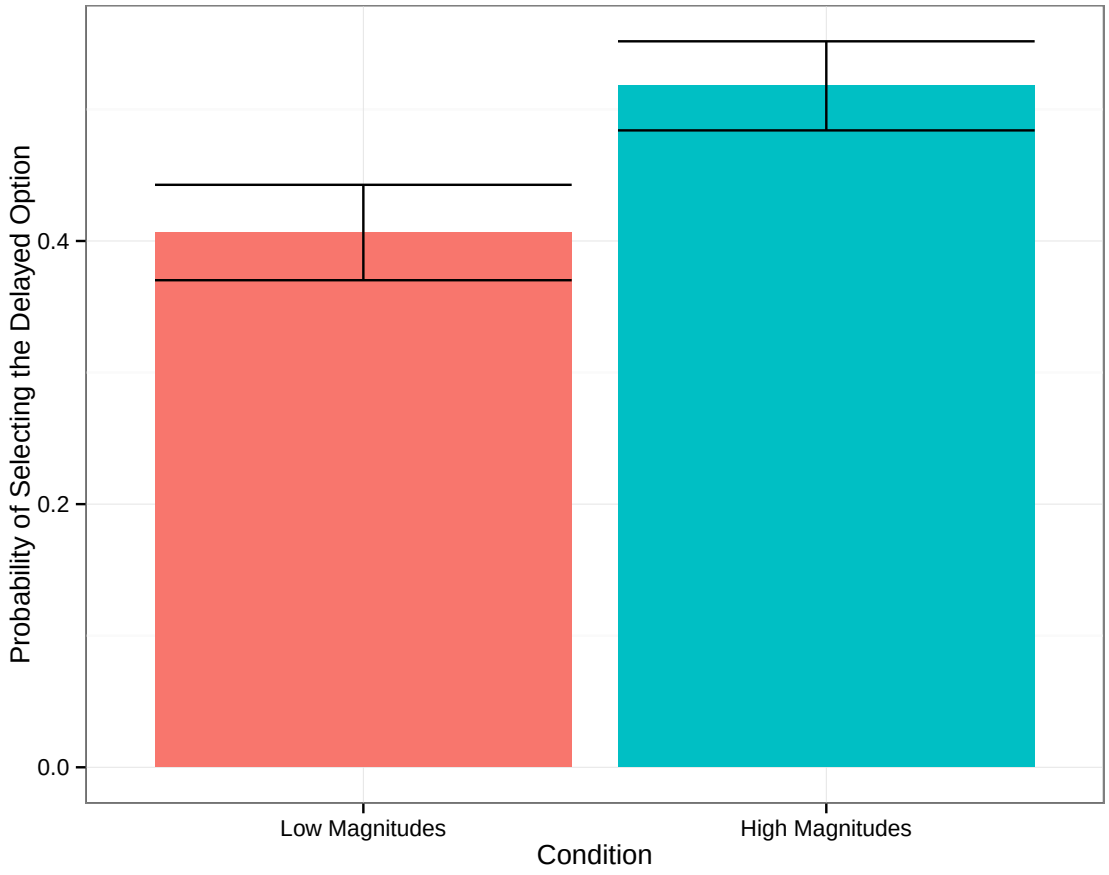


Figure 4.11: Main effect of the Magnitude factor in the R^2 study. (Error bars reflect 95% confidence intervals based on the nonparametric bootstrap.)

the worst variant, the R^2 was 0.0013. The full list of results is shown in Table 4.8.

Table 4.8: Correlations between MEL-derived measures of impatience and an general index of patience found in the R^2 Study

Percentages	Bank	High Magnitude	R^2
Standard MEL	Standard MEL	Low	0.10
Standard MEL	Standard MEL	High	0.094
Standard MEL	Banking Frame	Low	0.074
Standard MEL	Banking Frame	High	0.0018
Percentage Gains	Standard MEL	Low	0.0013
Percentage Gains	Standard MEL	High	0.14
Percentage Gains	Banking Frame	Low	0.089
Percentage Gains	Banking Frame	High	0.035

After pooling over all of these 8 variants of the MEL task, I found that a linear regression of the general impatience index scores on the probability of selecting the delayed good in the MEL task produced an adjusted R^2 of 0.048. This correlation is quite small, but is not likely to be spurious: in the fitted regression model the coefficient for task score was -3.4121 ($t = -5.906$, $p < 5.56^{-9}$.) Moreover, this value is comparable to the values found by Chabris et al. in their 2008 paper.

Discussion: In conclusion, the R^2 study failed to produce any variant of MEL that was clearly superior at eliciting a measure of impatience that predicted field behaviors. Along with the BIS/BAS study and work by other groups (e.g. the 2008 Chabris et al. study), this finding suggests that MEL tasks cannot be made substantially more externally valid without more substantive changes to the basic experimental paradigm.

List of Questions used in the BIS/BAS and R^2 Studies

The BIS/BAS questions (Carver and White, 1994) used in the BIS/Study are shown below.

1. "A person's family is the most important thing in life."

2. “Even if something bad is about to happen to me, I rarely experience fear or nervousness.”
3. “I go out of my way to get things I want.”
4. “When I’m doing well at something I love to keep at it.”
5. “I’m always willing to try something new if I think it will be fun.”
6. “How I dress is important to me.”
7. “When I get something I want, I feel excited and energized.”
8. “Criticism or scolding hurts me quite a bit.”
9. “When I want something I usually go all-out to get it.”
10. “I will often do things for no other reason than that they might be fun.”
11. “It’s hard for me to find the time to do things such as get a haircut.”
12. “If I see a chance to get something I want I move on it right away.”
13. “I feel pretty worried or upset when I think or know somebody is angry at me.”
14. “When I see an opportunity for something I like I get excited right away.”
15. “I often act on the spur of the moment.”
16. “If I think something unpleasant is going to happen I usually get pretty ‘worked up.”’
17. “I often wonder why people act the way they do.”
18. “When good things happen to me, it affects me strongly.”
19. “I feel worried when I think I have done poorly at something important.”

20. “I crave excitement and new sensations.”

21. “When I go after something I use a ‘no holds barred’ approach.”

22. “I have very few fears compared to my friends.”

23. “It would excite me to win a contest.”

24. “I worry about making mistakes.”

The Chabris questions (Chabris et al., 2008) that we used as a sample of field behaviors linked to impatience are listed below.

1. “What is your age?”

2. “What is your gender?”

3. “What is the highest level of education you’ve completed?”

4. “What is your height in feet and inches?”

5. “What is your weight in pounds?”

6. “How many hours per week are you physically active (for example, walking, working around the house, working out)?”

7. “How many of those hours represent exercise primarily intended to improve or maintain your health or fitness?”

8. “If you do any exercise primarily for health or fitness, how would you rate its intensity?”

9. “Do you smoke (including cigarettes, cigars, pipes, or anything else)?”

10. “If you smoke cigarettes, about how many packs do you smoke per week?”

11. "In a typical week, how often do you choose your food (the type and/or amount) with health and fitness concerns in mind?"

12. "How often do you visit your dentist for a check-up?"

13. "When your doctor gives you a prescription to fill at the drugstore (excluding birth control), do you follow it exactly (for example, by going to the drugstore, picking up the medication, taking all of the medication on schedule, and finishing the entire prescription)?"

14. "How often do you floss your teeth?"

15. "In a typical week, how often do you eat more than you think you should eat?"

16. "If you have any credit cards, over the past two years how many times were you charged a late fee for making a credit card payment after the deadline?"

17. "If you have any credit cards, over the past two years, how often have you paid your credit card bill in full, as opposed to paying less than the full amount? (Paying in full means carrying no debt to the next month's bill.)"

18. "Over the past three years, what percentage of your income have you saved? (Please include savings into retirement plans and any other form of savings that you do.)"

19. "On average, how many days per month do you gamble money, including visiting casinos, buying lottery tickets, betting on sports, playing poker, etc?"

20. "Compared to your friends who are close to you in age, how much wealth have you accumulated? (Wealth includes retirement savings, stocks, bonds, and mutual funds you own, money in bank accounts, the value of your home minus the mortgage, etc.)"

21. "Compared to the other members of your family in your generation – brothers, sisters, and cousins close to your age – how much wealth have you accumulated?"

22. "Are you currently attending school?"

23. "What is your current employment status?"

Chapter 5

ITCH: A New Model of Intertemporal Choice

DU Theory, as described in Chapter 4, underlies all of the popular models of intertemporal choice. These models assume that the discount function is a summary of time-tradeoff preferences and that these time preferences shape an individual's behavior in all domains. Under this premise, the discount function measured in MEL tasks should be the same discount function that guides all intertemporal choices, including retirement savings, educational investments, and health investments. The same discount function is also used in policy analysis e.g. to evaluate policies that mitigate global climate change (Stern, 2006).

The experiments described in Chapter 4 provided evidence that such domain generality cannot plausibly be expected from models of intertemporal choice. Indeed, these experiments suggested that a consistent discount function is not even employed across slightly different variants of the same MEL experiment. In particular, the two Time Frame studies I presented challenge the hypothesis that people exhibit any form of systematic discounting behavior in MEL experiments. To make sense of the results of MEL experiments, I have proposed an alternative model of how participants make

decisions in such tasks. I call this model the InterTemporal Choice Heuristics (ITCH) model.

5.1 Derivation of the ITCH Model

The ITCH model is distinguished from economic models of intertemporal choice in particular by its removal of the Axiom of Unit Invariance, for which robust violations were demonstrated in the two Time Frame studies described in Chapter 4. Instead of adopting a model of decision-making that is based on a discount function that devalues goods as they are delayed in absolute time, I have developed a model of how participants make choices in MEL tasks that is based on psychological principles. In this model, participants rely upon a weighted combination of several simple heuristics for choosing between the two options. These heuristics span the simplest types of comparisons that can be made between the two options (e.g., subtracting and/or dividing them) along each of the two relevant dimensions (i.e., time of delivery and value). The application of each heuristic to a particular pair of options generates a vector of features for that choice (i.e., the results of each simple computation), each of which favors one or the other choice. The model assumes that the final decision reflects a weighted combination of these features, which can be interpreted as the attentional focus on each heuristic.

The ITCH model therefore extends a large body of work in psychology on multi-attribute models of decision-making in which goods are represented in terms of vectors of features that are analyzed by a hypothetical process that ultimately produces a single decision. For example, it is closely related to the linear model of decision-making under risk developed by Lichtenstein and Slovic (Lichtenstein and Slovic, 1971). It is also similar to Tversky's Elimination by Aspects (EBA) model (Tversky, 1972) and shares properties with recent models like Hare and Rangel's model of

decisions over healthy and unhealthy foods (Hare, Camerer, and Rangel, 2009).

Having described the motivation for the ITCH model, I now derive the model in detail before examining how the model fits data from an experiment that I call the Gains vs. Losses study, which extends the Relative and Absolute Gains study from Chapter 4.

I begin by noting again that, in a MEL task, both the earlier, smaller good and the later, larger good are stimuli defined in terms of two attributes: a monetary value, x, and a time value, t. In the two-good MEL experiment I will adopt the notation $e = (x_1, t_1)$ and $l = (x_2, t_2)$ for the earlier and later goods respectively. I also adopt the strong assumption that the units of time and money are discarded when constructing these vectors so that only the numerals used to describe a good immediately affect the participant's decisions.

Given this notation, I formalize a probabilistic model of choice that specifies the probability of selecting each of several time-dated options. I do this by constructing a *scoring rule* that assigns a local value to each of the goods with the assumption that the goods are chosen in proportion to their scores. When I say that this value is local, I mean to emphasize that I am not constructing utilities in the sense of Utility Theory: I allow the possibility that the score for a good depends upon the other goods that are avaialble, which could lead to violations of the context-independence assumption implicit in Utility Theory. Such violations may seem undesirable from a normative perspective, but this assumption of local values is required in order to construct a model in which a changing reference point can alter behavior.

I now describe the construction of this scoring rule in detail. To do so, I first define the reference point against which each good will be compared. This reference point is designed to summarize the entire set of available options: in the two-option MEL task this means that the reference point depends upon both the earlier, smaller good and the later, larger good. For simplicity, I will assume that the reference point is the

centroid of the two options, which can be computed by taking the arithmetic mean of the two options. This produces the following definition of the reference point:

$$(x^*, t^*) = (\frac{[x_1 + x_2]}{2}, \frac{[t_1 + t_2]}{2}).$$

I then make the assumption that both the earlier, smaller good and the larger, later good are each compared against the reference point in both absolute and relative terms. This two-way comparison expands each of the two options into a four-dimensional vector of features:

$$(x_1, t_1) \rightarrow (x_1 - x^*, \frac{x_1}{x^*}, t_1 - t^*, \frac{t_1}{t^*}),$$

$$(x_2, t_2) \rightarrow (x_2 - x^*, \frac{x_2}{x^*}, t_2 - t^*, \frac{t_2}{t^*}).$$

I then assign scores to each good by differentially weighting the components of these four-dimensional vectors. These weights reflect the relative attention that is paid to each of the four possible comparisons that can be made against the reference point. This is done to allow the ITCH model to express the notion that framing manipulations may shift the dimensions on which attention is focused during the decision-making process. In addition to decision weights assigned to each of the four features, I add a constant term to the earlier and later option's scores that is meant to account for factors that are not described by the four features of the ITCH model. For example, these additional constant terms may reflect the non-normative preference for earlier options that was seen in Chapter 4 and was especially salient in the IRA study.

By combining the weighted combination of features and the constant terms, I

arrive at the final scores for each good:

$$s_1 = \gamma_1 + \beta_1(x_1 - x^*) + \beta_2\frac{x_1}{x^*} + \beta_3(t_1 - t^*) + \beta_4\frac{t_1}{t^*},$$

$$s_2 = \gamma_2 + \beta_1(x_2 - x^*) + \beta_2\frac{x_2}{x^*} + \beta_3(t_2 - t^*) + \beta_4\frac{t_2}{t^*}.$$

I then assume that each of these scores is transformed into a choice probability using the softmax function:

$$p_1 = \frac{e^{s_1}}{e^{s_1} + e^{s_2}},$$

$$p_2 = \frac{e^{s_2}}{e^{s_1} + e^{s_2}}.$$

In binary choice settings, these two probabilities are linked by the equation $p_2 = 1 - p_1$. As such, one only needs to define one of these two probabilities to define the ITCH model as a whole. I therefore construct a notational variant of the ITCH equations from above that is easier to reason about in binary choice settings.

To construct this simplified ITCH model, note first that,

$$p_2 = \frac{1}{1 + e^{-(s_2 - s_1)}}.$$

By expanding and simplifying the $s_2 - s_1$ term, one finds that

$$s_2 - s_1 = (\gamma_2 - \gamma_1) + \beta_1(x_2 - x^* - x_1 + x^*) + \beta_2(\frac{x_2}{x^*} - \frac{x_1}{x^*}) + \beta_3(t_2 - t^* - t_1 + t^*) + \beta_4(\frac{t_2}{t^*} - \frac{t_1}{t^*}).$$

If one then introduces a new variable, $\beta_0 = \gamma_2 - \gamma_1$, this expression simplifies to the equation,

$$s_2 - s_1 = \beta_0 + \beta_1(x_2 - x_1) + \beta_2\frac{x_2 - x_1}{x^*} + \beta_3(t_2 - t_1) + \beta_4\frac{t_2 - t_1}{t^*},$$

which then gives the final form of the ITCH model,

$$p_2 = L(\beta_0 + \beta_1(x_2 - x_1) + \beta_2\frac{x_2 - x_1}{x^*} + \beta_3(t_2 - t_1) + \beta_4\frac{t_2 - t_1}{t^*}),$$

where $L(z)$ is the inverse logistic function, $L(z) = (1 + \exp(-z))^{-1}$.

In this simplified form, it is clear that the reference points defined above only influence the ITCH when β_2 and β_4 are non-zero. We will see that the ITCH model, when fit to empirical data, reliably estimates that these terms are non-zero.

5.2 Extensions of the ITCH Model

The ITCH model could be extended to account for other findings in the intertemporal choice literature by employing more complex reference points and more complex scoring rules. In particular, one might consider extended ITCH models of the form:

$$s_1 = \gamma_1 + V_1(x_1 - x^*) + V_2(\frac{x_1}{x^*}) + V_3(t_1 - t^*) + V_4(\frac{t_1}{t^*})$$

$$s_2 = \gamma_2 + V_1(x_2 - x^*) + V_2(\frac{x_2}{x^*}) + V_3(t_2 - t^*) + V_4(\frac{t_2}{t^*})$$

If one assumes that (i) the reference point is closer to the first option presented than to the second option and that (ii) the value functions exhibit loss aversion like the convex-concave value function of Prospect Theory (Kahneman and Tversky, 1979), then an extension of the ITCH model predicts the behavioral effects that inspired Query Theory (Weber et al., 2007) as well as the delay/speedup asymmetry noted by Loewenstein (G. F. Loewenstein, 1988).

In what follows, I present the core form of the ITCH model rather than any such variant because the purely linear aggregation mechanism used in the standard ITCH model makes the model easier to understand and simultaneously facilitates the statistical estimation of the model parameters.

5.3 The Empirical Performance of ITCH Model

In this section, I show that the ITCH model can be fit to five different variants of a MEL task. The pattern of weights it assigns to the four main terms of the ITCH equation is consistent across participants and variants, suggesting that people use a stable combination of heuristics when evaluating MEL questions, irrespective of time units, magnitude or other manipulations of content and context. Moreover, the ITCH model is able to account for deviations from the Axiom of Unit Invariance because it is dependent upon the form of the options shown in a MEL task as well as their content, unlike economic models that depend only on the content of a choice. This dependence on form is a result of my earlier assumption that the ITCH model ignores the units of time and money used to define goods in a MEL task.

Before I fit the model to data, I note that the ITCH model treats the dimensions of time and money symmetrically. As a consequence of this symmetry, the ITCH model is able to account for many anomalous findings from MEL tasks that have been previously described in the literature, yet remained unexplained by previous models. For example, when β_3 and β_4 are both negative, the ITCH model predicts that the probability of choosing the later, larger good will increase as both goods are delayed by an additional identical amount of time, the phenomenon referred to as decreasing impatience in Chapter 4 and often taken as evidence of hyperbolic discounting. Similarly, when β_1 and β_2 are both positive (as they are in my estimates), the ITCH model predicts that participants will be more likely to choose the larger, later good if the monetary stakes are large. This prediction is seen in virtually all MEL tasks and was referred to as a violation of the magnitude invariance property defined in Chapter 4. Finally, I note that if β_0 is negative, β_1 is positive and all other terms are zero, the ITCH model reduces to the assertion that decisions in MEL are determined by a fixed cost that must be paid before a participant will be willing to wait for a delayed good. Exactly such a fixed cost model was put forward by

Benhabib and Bisin (Benhabib et al., 2010), who found that it both (a) outpredicts all of the classical models of the discount function and (b) accounts for the robust violations of the partition invariance described in Chapter 4. Thus the ITCH model strictly subsumes one highly successful model of behavior in MEL tasks.

In conclusion, the ITCH model, when parameterized appropriately, predicts many, if not all, of the DU Theory-incompatible findings that have accumulated in the literature on MEL. We will see that empirical fits of the ITCH models lead to model parameterizations that do predict all of these anomalies.

5.4 The Gains vs. Losses Study

Introduction: To calibrate the ITCH model's parameters and to compare it with three standard economic models (namely the exponential, hyperbolic, and quasi-hyperbolic models of discounting), I conducted an experiment that I will call the Gains vs. Losses study. The Gains vs. Losses study is an extension of the Relative and Absolute Gains study described in Chapter 4.

Methods: For the Gains vs. Losses Study, 1,000 participants were recruited from Amazon's Mechanical Turk. These participants were randomly assigned to one of five task variants. In each variant, the participants answered 25 hypothetical MEL choices. These binary choices were generated using the same orthogonaliation approach taken in the Relative Gains and Losses study: this orthogonaliation approach was essential to insuring that the parameters of the ITCH model were maximally identifiable.

In the Gains vs. Losses study, each of the five framing conditions involved a MEL task that differed in the language used to describe normatively equivalent goods:

- **Absolute Gains**: \$5 today vs. \$5 plus an additional \$5 in 4 weeks
- **Relative Gains**: \$5 today vs. \$5 plus an additional 100% in 4 weeks
- **Standard MEL Format**: \$5 today vs. \$10 in 4 weeks

- **Absolute Losses**: $10 in 4 weeks vs. $10 minus $5 today
- **Relative Losses**: $10 in 4 weeks vs. $10 minus 50% today

To compare models, I used a random-resampling based cross-validation approach in which each model was fit to one set of data and then tested on another set of data. Cross-validation provides an unbiased method for comparing the predictive power of models with different numbers of parameters, which is necessary in this case because the ITCH model has more parameters than the models with which I compared it (Kohavi et al., 1995). In all five conditions, the ITCH model outperformed all of the standard economic models of intertemporal choice in predicting participants' choices. See Tables 5.1–5.10 for detailed model comparison results. The last section of this chapter presents detailed information about the exact model calibration and assessment procedures that were used.

In addition to my analysis of each condition's data in isolation (which can be viewed as five separate experiments), I pooled data across all five MEL task variants to provide a global assessment of model performance. The results of this pooled analysis are shown in Figure 5.1.

Beyond outpredicting all of the classical models of intertemporal choice, the ITCH model produced more consistent fits. Although the magnitude of the parameters varied across fits, the relative values of the β parameters were stable across all five framing conditions. The consistency of the decision weights for the ITCH model across variants of the MEL can be seen visually in Figure 5.2 and numerically in Table 5.11. This consistency was assessed by computing the pairwise correlations between the fitted model parameters for each of the six data groupings and testing the average correlation against zero using a t-test: the result was highly significant ($t = 125.977$, $df = 35$, $p < 10^{-14}$).

The consistency of the estimated parameters of the ITCH model implies that participants assigned the same relative weight to the heuristics represented by each

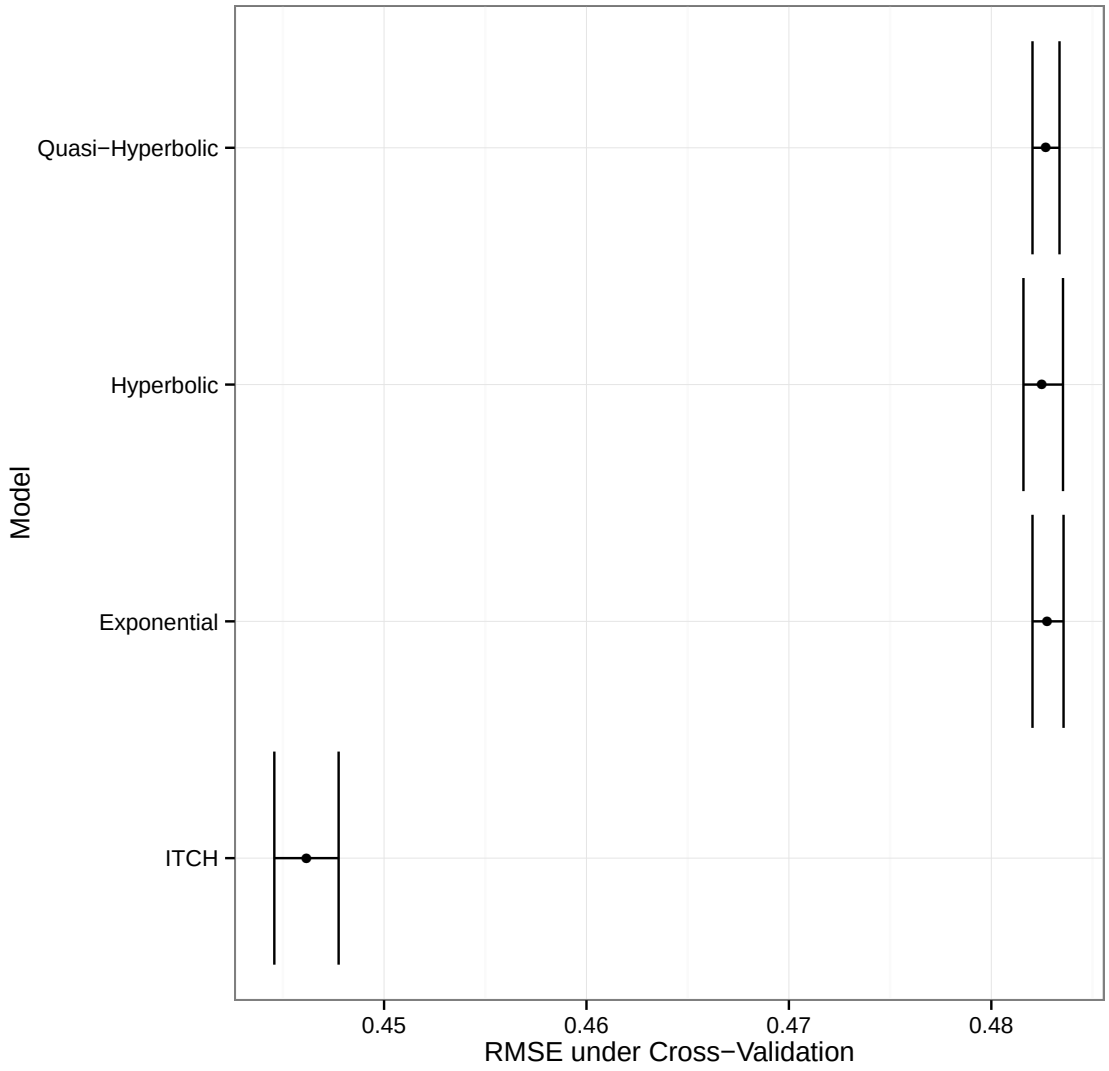


Figure 5.1: Model comparison results after pooling all data from the Gains vs. Losses study

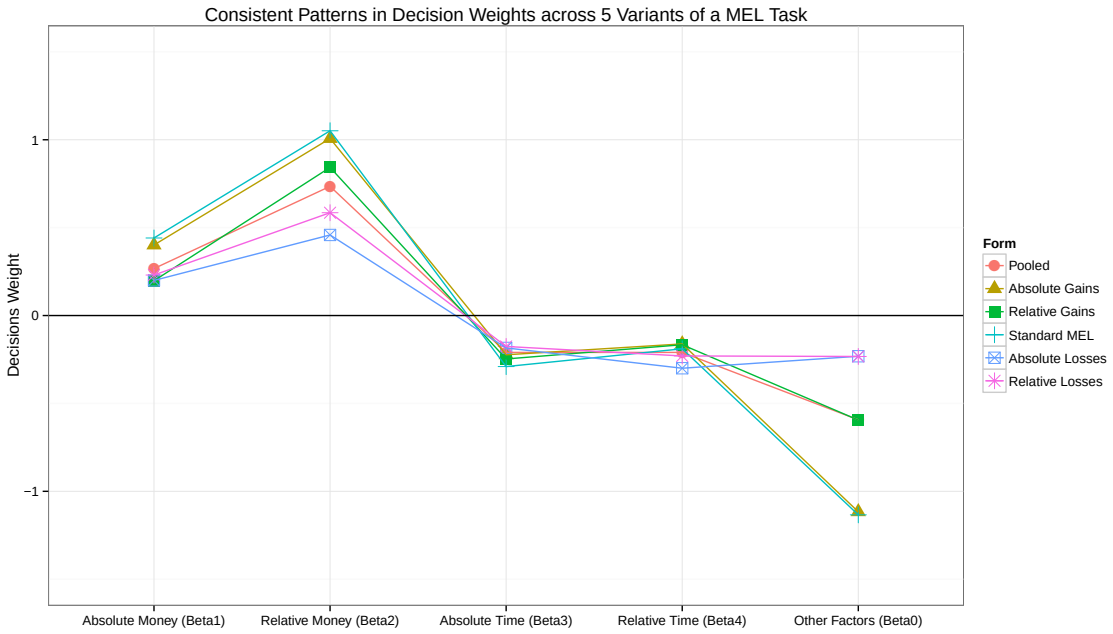


Figure 5.2: Coefficients

term of the ITCH model, irrespective of manipulations of task content or context.

5.5 Interpreting ITCH

MEL tasks are supposed to measure domain-general impatience parameters (i.e., a discount function), but the experiments reported in Chapter 4 show that MEL tasks generate robust inconsistencies. In my experiments, intertemporal choices are shown to be highly sensitive to many normatively irrelevant changes, including changes to the numeric format of the monetary gains derived from waiting and changes to the time units that are used to describe goods. For example, the first and second Time Frame studies demonstrated that participants were far more likely to choose an immediate reward when choosing between now vs. 7 days than when they were choosing between now vs. 1 week. This failure of the participants' choices to satisfy the Axiom of Unit Invariance implies that individuals do not systematically discount when confronted with MEL tasks.

To better understand participants' decision-making behavior, I developed an alternative model of behavior in MEL tasks (ITCH) based on basic psychological principles, including reference points and ratio comparisons. Future extensions of the ITCH model might incorporate additional psychological factors like loss aversion to explain other well-known effects in intertemporal choice, such as those described by Query Theory (Weber et al., 2007).

The results of the Gains vs. Losses study show that this alternative model is better able to predict choices in MEL tasks than standard economic models of intertemporal choice. Critically, the fit of the ITCH model demonstrated that the relative weights assigned to each of the heuristics in the model was stable across a wide range of conditions. Thus, while participants' decision making behavior in MEL tasks may not conform to the normative prescripts of standard economic models – or be meaningfully described by a unit invariant discount function –, the fit of the ITCH model suggests that decision-making in MEL tasks nevertheless exhibits a form of psychological stability that may generalize to other tasks.

As noted at the end of Chapter 4, an important remaining question about the interpretation of MEL tasks concerns their relationship to impatience. If responses to MEL tasks cannot be interpreted quantitatively in terms of economic models of intertemporal choice, why are the results from many MEL surveys consistently correlated, albeit weakly, with other measures of impatience (Chabris et al., 2008)? While some researchers have suggested that these correlations arise from omitted variables (Shamosh et al., 2008), my findings are also consistent with the claim that MEL tasks may reflect impatience filtered through the lens of the heuristics (or combinations of these) that people use to perform these tasks, such as those identified by the ITCH model. That is, while impatience cannot be equated with a simple unit invariant discount parameter estimated by a MEL task, it might be able to be described in terms of a systematic manner in which participants weigh time of delivery and value, as

described by the ITCH model. Along these lines, my findings do not negate those of studies that have used MEL tasks to qualitatively dissociate responses to immediate rewards from those to later rewards.

In summary, my findings raise two important points. First, they expose a critical limitation in a popular approach to empirical research on intertemporal choice and impatience – the use of MEL tasks – that may explain the inconsistencies and unreliability of findings in the literature using such tasks. Second, the ability of the ITCH model to predict behavior in MEL tasks suggests that for these tasks, as for many other decision-making tasks, people may apply simple heuristics that provide a quick and easy way to generate a response that is sufficient for their purposes. The ITCH model refines this approach to understanding human decision making, by modeling time-money tradeoffs in MEL tasks as resulting from a weighted combination of simple heuristics. This approach may prove useful in understanding behavior in a wider variety of tasks, and may reflect a decision-making strategy that economizes on cognitive costs (i.e., that simplifies the task) while attempting to maximize rewards.

5.6 Model Fitting and Comparison Procedures

After gathering data for all five conditions of the Gains vs. Losses study, I estimated the parameters of the ITCH model using standard maximum likelihood estimation techniques for Generalized Linear Models (McCullagh and Nelder, 1989). This is possible because the binary choice version of the ITCH model is a form of logistic regression in which specific customized features are employed: these features describe the putative psychological processes that are presumed to underlie the decision-process used by participants in an MEL experiment.

To compare models, I performed a standard cross-validation analysis in which I repeatedly split the data into two randomly generated parts: the first part consisted

of 75% of the data and was used to estimate the parameters of my model; the second part consistent of 25% of the data and was used to assess the model's fit to "held out" data (Kohavi et al., 1995). I assessed model fit using the Mean Absolute Deviation (MAD) between a model's prediction (which is a probability between 0 and 1) and the actual choice indicator (which is 0 or 1 depending on whether or not the later, larger option was chosen) because this metric is (a) immediately interpretable on the scale of probabilities and (b) robust to outliers. I have also assessed model fit using held-out log likelihood and Root Mean Squared Error (RMSE): these alternative measures produce the same qualitative results when the models are ranked in terms of predictive power.

I compared the ITCH model with seven other models of intertemporal choice: a baseline model that makes choices by flipping a biased coin, an exponential model of discounting (Samuelson, 1937), a hyperbolic model of discounting (Ainslie, 1975), a generalized hyperbolic model of discounting (G. Loewenstein and Prelec, 1992), a quasi-hyperbolic of discounting (Laibson, 1997) and a fixed cost model of discounting (Benhabib et al., 2010). I also estimated the "DRIFT" model of Read et al (Read et al., 2011). The DRIFT model was developed independently from the ITCH model, but has a similar functional form. It differs from the ITCH model in that the DRIFT model does not treat time and money symmetrically. The ITCH model outperforms all of these models.

Because the exponential, hyperbolic, generalized hyperbolic and quasi-hyperbolic models contain non-linearities, I estimate their parameters via maximum likelihood using a more general constrained optimization suite in R that implements the L-BFGS-B algorithm (Zhu, Byrd, Lu, and Nocedal, 1997).

The functional forms of each model are shown below. In each of these models, the

notation, $L(z)$, represents the inverse logistic function of z given by

$$L(z) = \frac{1}{1 + e^{-z}}.$$

The notation $I(x)$ represents the indicator variable that is 1 if x is true and 0 otherwise.

In the baseline model, I estimate a single constant probability of selecting the delayed good:

$$p_2 = \pi$$

In the exponential model, I estimate:

$$p_2 = L(x_2\delta^{t_2} - x_1\delta^{t_1})$$

In the hyperbolic model, I estimate:

$$p_2 = L(x_2(1 + \alpha t_2)^{-1} - x_1(1 + \alpha t_1)^{-1})$$

In the generalized hyperbolic model, I estimate:

$$p_2 = L(x_2(1 + \alpha t_2)^{-\frac{\beta}{\alpha}} - x_1(1 + \alpha t_1)^{-\frac{\beta}{\alpha}})$$

In the quasi-hyperbolic model, I estimate:

$$p_2 = L(x_2\beta^{I(t>0)}\delta^{t_2} - x_1\beta^{I(t>0)}\delta^{t_1})$$

In the fixed cost model, I estimate:

$$p_2 = L(x_2 - x_1 - c)$$

In the DRIFT model, I estimate:

$$p_2 = L(\beta_0 + \beta_1(x_2 - x_1) + \beta_2(\frac{x_2 - x_1}{x_1}) + \beta_3(\frac{x_2}{x_1}^{\frac{1}{t_2 - t_1}} - 1) + \beta_4(t_2 - t_1)$$

Because of the large range of monetary quantities over which participants make decisions, I find that the exponential, hyperbolic, quasi-hyperbolic and generalized hyperbolic models can be numerically unstable. To resolve this numeric instability, I compute the probability of an entire sequence of choices by bounding the per-choice probabilities estimated using the functional forms above at a value that is not 0 or 1. In my estimation, the exponential, hyperbolic, quasi-hyperbolic and generalized hyperbolic models perform better if they are bounded at 0.25 and 0.75, but I have also tested them bounded at 0.01 and 0.99, 0.05 and 0.95, and 0.1 and 0.9. In general, each bounded probability, p_B is computed as,

$$p_B = (1 - \epsilon)p + \frac{\epsilon}{2}.$$

The performance of all models for these four probability bounds are shown in Tables 5.1-5.8. In addition, I have estimated variants of the exponential, hyperbolic, quasi-hyperbolic and generalized hyperbolic models in which utilities are not complemented by constant variance logistic noise (as implied by the inverse logistic function), but are instead corrupted by scale-invariant, homothetic noise in which the probability of errors depends on the ratio of the utilities of the two goods, rather than upon their absolute difference. This scale-invariant, homothetic noise assumptions implies the alternative functional forms shown below.

In the scale-invariant, homothetic exponential model, I estimate:

$$p_2 = L(\log(x_2\delta^{t_2}) - \log(x_1\delta^{t_1}))$$

In the scale-invariant, homothetic hyperbolic model, I estimate:

$$p_2 = L(\log(x_2(1+\alpha t_2)^{-1}) - \log(x_1(1+\alpha t_1)^{-1}))$$

In the scale-invariant, homothetic generalized hyperbolic model, I estimate:

$$p_2 = L(\log(x_2(1+\alpha t_2)^{-\frac{\beta}{\alpha}}) - \log(x_1(1+\alpha t_1)^{-\frac{\beta}{\alpha}}))$$

In the scale-invariant, homothetic quasi-hyperbolic model, I estimate:

$$p_2 = L(\log(x_2\beta^{I(t>0)}\delta^{t_2}) - \log(x_1\beta^{I(t>0)}\delta^{t_1}))$$

The results of comparing the ITCH model against these scale-invariant, homothetic models are shown in Tables 5.9-5.10.

All of these analyses were repeated with a subset of the data that included only American participants from the Gains vs. Losses study: equivalent results were found in all conditions for this subset of the data.

Table 5.1: Mean absolute deviations for all models and conditions under cross-validation using data from the Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. Per-choice probabilities are bounded at 0.01 and 0.99. (Part 1 of 2)

Condition	Baseline	DRIFT	Exponential	Fixed Cost
1	0.4108 (0.0006)	0.3267 (0.0010)	0.4595 (0.0016)	0.4016 (0.0007)
2	0.4682 (0.0004)	0.4016 (0.0007)	0.4752 (0.0022)	0.4659 (0.0002)
3	0.4118 (0.0008)	0.3234 (0.0009)	0.4725 (0.0026)	0.4010 (0.0007)
4	0.4944 (0.0002)	0.4630 (0.0005)	0.4990 (0.0001)	0.4901 (0.0002)
5	0.4944 (0.0002)	0.4526 (0.0005)	0.4845 (0.0003)	0.4882 (0.0003)
Pooled	0.4673 (0.0002)	0.4088 (0.0003)	0.4922 (0.0002)	0.4607 (0.0002)

Table 5.2: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. Per-choice probabilities are bounded at 0.01 and 0.99. (Part 2 of 2)

Condition	Gen. Hyperbolic	Hyperbolic	ITCH	Quasi-Hyperbolic
1	0.4546 (0.0016)	0.4550 (0.0013)	0.3214 (0.0008)	0.4565 (0.0015)
2	0.4793 (0.0022)	0.4798 (0.0021)	0.3904 (0.0008)	0.4790 (0.0023)
3	0.4693 (0.0028)	0.4676 (0.0026)	0.3159 (0.0011)	0.4686 (0.0026)
4	0.4979 (0.0001)	0.4988 (0.0001)	0.4530 (0.0005)	0.4986 (0.0001)
5	0.4836 (0.0004)	0.4846 (0.0003)	0.4412 (0.0005)	0.4845 (0.0004)
Pooled	0.4920 (0.0001)	0.4922 (0.0001)	0.4001 (0.0004)	0.4922 (0.0001)

Table 5.3: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. Per-choice probabilities are bounded at 0.05 and 0.95. (Part 1 of 2)

Condition	Baseline	DRIFT	Exponential	Fixed Cost
1	0.4110 (0.0007)	0.3275 (0.0010)	0.4331 (0.0009)	0.4009 (0.0008)
2	0.4685 (0.0004)	0.4010 (0.0007)	0.4581 (0.0008)	0.4657 (0.0004)
3	0.4117 (0.0007)	0.3233 (0.0010)	0.4391 (0.0009)	0.4015 (0.0008)
4	0.4943 (0.0002)	0.4629 (0.0005)	0.4990 (0.0001)	0.4900 (0.0002)
5	0.4940 (0.0002)	0.4532 (0.0005)	0.4811 (0.0005)	0.4883 (0.0003)
Pooled	0.4672 (0.0002)	0.4081 (0.0003)	0.4737 (0.0004)	0.4609 (0.0002)

Table 5.4: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. Per-choice probabilities are bounded at 0.05 and 0.95. (Part 2 of 2)

Condition	Gen. Hyperbolic	Hyperbolic	ITCH	Quasi-Hyperbolic
1	0.4356 (0.0008)	0.4349 (0.0008)	0.3212 (0.0009)	0.4351 (0.0009)
2	0.4583 (0.0008)	0.4595 (0.0010)	0.3899 (0.0007)	0.4578 (0.0009)
3	0.4398 (0.0010)	0.4407 (0.0008)	0.3144 (0.0010)	0.4402 (0.0010)
4	0.4979 (0.0001)	0.4987 (0.0001)	0.4527 (0.0005)	0.4986 (0.0002)
5	0.4777 (0.0006)	0.4806 (0.0005)	0.4408 (0.0006)	0.4798 (0.0006)
Pooled	0.4728 (0.0002)	0.4729 (0.0003)	0.3996 (0.0003)	0.4735 (0.0003)

Table 5.5: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. Per-choice probabilities are bounded at 0.1 and 0.9. (Part 1 of 2)

Condition	Baseline	DRIFT	Exponential	Fixed Cost
1	0.4110 (0.0006)	0.3265 (0.0009)	0.4346 (0.0009)	0.4005 (0.0006)
2	0.4683 (0.0004)	0.4006 (0.0007)	0.4437 (0.0010)	0.4657 (0.0004)
3	0.4124 (0.0007)	0.3237 (0.0011)	0.4411 (0.0009)	0.4013 (0.0007)
4	0.4942 (0.0002)	0.4633 (0.0004)	0.4989 (0.0001)	0.4902 (0.0002)
5	0.4942 (0.0002)	0.4530 (0.0005)	0.4767 (0.0006)	0.4880 (0.0003)
Pooled	0.4672 (0.0002)	0.4081 (0.0003)	0.4658 (0.0003)	0.4609 (0.0002)

Table 5.6: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. Per-choice probabilities are bounded at 0.1 and 0.9. (Part 2 of 2)

Condition	Gen. Hyperbolic	Hyperbolic	ITCH	Quasi-Hyperbolic
1	0.4372 (0.0009)	0.4373 (0.0008)	0.3215 (0.0009)	0.4360 (0.0008)
2	0.4441 (0.0009)	0.4453 (0.0009)	0.3903 (0.0007)	0.4440 (0.0009)
3	0.4422 (0.0009)	0.4428 (0.0008)	0.3146 (0.0011)	0.4423 (0.0009)
4	0.4978 (0.0001)	0.4988 (0.0001)	0.4527 (0.0005)	0.4987 (0.0002)
5	0.4716 (0.0006)	0.4752 (0.0006)	0.4409 (0.0005)	0.4746 (0.0005)
Pooled	0.4608 (0.0005)	0.4641 (0.0006)	0.3994 (0.0004)	0.4610 (0.0005)

Table 5.7: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. Per-choice probabilities are bounded at 0.25 and 0.75. (Part 1 of 2)

Condition	Baseline	DRIFT	Exponential	Fixed Cost
1	0.4103 (0.0007)	0.3268 (0.0009)	0.4458 (0.0008)	0.4006 (0.0008)
2	0.4685 (0.0004)	0.4005 (0.0007)	0.4524 (0.0007)	0.4654 (0.0004)
3	0.4127 (0.0007)	0.3233 (0.0011)	0.4510 (0.0007)	0.4010 (0.0007)
4	0.4943 (0.0002)	0.4632 (0.0005)	0.4990 (0.0001)	0.4902 (0.0002)
5	0.4943 (0.0002)	0.4528 (0.0005)	0.4628 (0.0008)	0.4882 (0.0003)
Pooled	0.4673 (0.0002)	0.4085 (0.0003)	0.4591 (0.0003)	0.4609 (0.0002)

Table 5.8: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. Per-choice probabilities are bounded at 0.25 and 0.75. (Part 2 of 2)

Condition	Gen. Hyperbolic	Hyperbolic	ITCH	Quasi-Hyperbolic
1	0.4458 (0.0007)	0.4475 (0.0007)	0.3222 (0.0010)	0.4461 (0.0007)
2	0.4519 (0.0007)	0.4532 (0.0007)	0.3888 (0.0008)	0.4525 (0.0008)
3	0.4518 (0.0008)	0.4541 (0.0007)	0.3157 (0.0010)	0.4517 (0.0008)
4	0.4970 (0.0003)	0.4985 (0.0002)	0.4525 (0.0005)	0.4978 (0.0004)
5	0.4627 (0.0008)	0.4620 (0.0007)	0.4408 (0.0006)	0.4635 (0.0011)
Pooled	0.4601 (0.0004)	0.4597 (0.0003)	0.3998 (0.0003)	0.4597 (0.0003)

Table 5.9: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. For these tests, the noise distribution for the classical economic models was a homothetic noise distribution that was scale-invariant and, in my estimation, produced superior model fits than the traditional logistic noise used in discrete choice models. For comparison with the logistic noise results, I have also bounded per-choice probabilities at 0.01 and 0.99. (Part 1 of 2)

Condition	Baseline	DRIFT	Exponential	Fixed Cost
1	0.4108 (0.0006)	0.3267 (0.0010)	0.3522 (0.0008)	0.4016 (0.0007)
2	0.4682 (0.0004)	0.4016 (0.0007)	0.4030 (0.0007)	0.4659 (0.0002)
3	0.4118 (0.0008)	0.3234 (0.0009)	0.3472 (0.0009)	0.4010 (0.0007)
4	0.4944 (0.0002)	0.4630 (0.0005)	0.4663 (0.0004)	0.4901 (0.0002)
5	0.4944 (0.0002)	0.4526 (0.0005)	0.4534 (0.0005)	0.4882 (0.0003)
Pooled	0.4673 (0.0002)	0.4088 (0.0003)	0.4165 (0.0003)	0.4607 (0.0002)

Table 5.10: Mean absolute deviations for all models and conditions under cross-validation using data from Gains vs. Losses study. Values in parentheses indicate variability under cross-validation. For these tests, the noise distribution for the classical economic models was a homothetic noise distribution that was scale-invariant and, in my estimation, produced superior model fits than the traditional logistic noise used in discrete choice models. For comparison with the logistic noise results, I have also bounded per-choice probabilities at 0.01 and 0.99. (Part 2 of 2)

Condition	Gen. Hyperbolic	Hyperbolic	ITCH	Quasi-Hyperbolic
1	0.3470 (0.0010)	0.3602 (0.0009)	0.3214 (0.0008)	0.3485 (0.0010)
2	0.3979 (0.0008)	0.4019 (0.0006)	0.3904 (0.0008)	0.3984 (0.0008)
3	0.3398 (0.0011)	0.3573 (0.0009)	0.3159 (0.0011)	0.3421 (0.0009)
4	0.4589 (0.0006)	0.4585 (0.0005)	0.4530 (0.0005)	0.4582 (0.0005)
5	0.4474 (0.0006)	0.4479 (0.0006)	0.4412 (0.0005)	0.4482 (0.0005)
Pooled	0.4101 (0.0003)	0.4134 (0.0003)	0.4001 (0.0004)	0.4111 (0.0003)

Table 5.11: Decision weights for the ITCH model separated by condition. Values in parentheses indicate standard errors. To make the decision weights more readily comparable, I z-scored all of the input variables.

Condition	β_0	β_1	β_2	β_3	β_4
1	-1.11 (0.08)	0.40 (0.07)	1.01 (0.07)	-0.22 (0.08)	-0.16 (0.08)
2	-0.59 (0.07)	0.20 (0.06)	0.84 (0.06)	-0.25 (0.07)	-0.17 (0.07)
3	-1.13 (0.09)	0.44 (0.08)	1.05 (0.08)	-0.29 (0.08)	-0.19 (0.08)
4	-0.23 (0.06)	0.20 (0.06)	0.46 (0.06)	-0.18 (0.06)	-0.30 (0.06)
5	-0.23 (0.06)	0.23 (0.06)	0.58 (0.06)	-0.18 (0.06)	-0.23 (0.06)
Pooled	-0.59 (0.03)	0.27 (0.03)	0.73 (0.03)	-0.21 (0.03)	-0.21 (0.03)

Chapter 6

Theory of Exploration and Exploitation

6.1 The Exploration-Exploitation Dilemma

Natural selection has endowed all organisms, including humans, with an innate drive to seek out rewards and to avoid punishments (Lewin, 1935). To act on those drives, organisms must make use of the imperfect information that they have about their environments (Stiglitz and Weiss, 1981). Although an omniscient organism could maximize rewards by always doing what seems best, real organisms must make decisions that will be robust to potential inaccuracies in their beliefs.

In order to correct inaccurate beliefs, an organism must seek out new information. But the pursuit of information can conflict with the pursuit of reward. To see how, imagine that you have recently moved to a town with two restaurants: Alice's Eatery and Bob's Bistro. On your first night out, you decide to go to Alice's, where you have a meal that exceeds your expectations. On your next night out, will you return to Alice's or will you try Bob's instead? Because of the atypically good meal you had at Alice's, a second visit to Alice's is more likely to yield a rewarding meal than a first

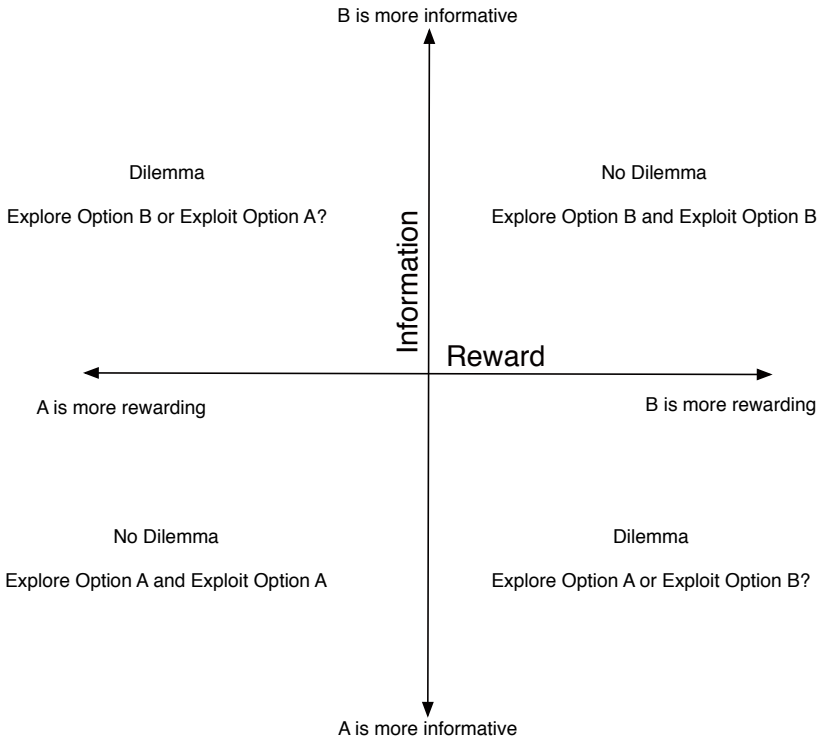


Figure 6.1: The Separate Dimensions of Reward and Information

trip to Bob's. But only a first trip to Bob's will give you the information needed to assess the accuracy of your newfound belief in Alice's superiority. The choice between returning to Alice's or trying Bob's for the first time therefore places your drive to maximize reward in direct conflict with your drive to gain information about your environment.

The choice between Alice's and Bob's is just one example of the general conflict between *exploring* the world for information and *exploiting* a known source of reward. More generally, you can imagine many different settings in which a choice between option A and option B involves consideration of the amount of reward and information that you can expect to gain from each option. This more general setting is shown

in Figure 6.1 in order to highlight the inherent multidimensionality of the problem faced by organisms that seek to satisfy the potentially conflicting goals of maximizing rewards and maximizing information.

In its most general form, the conflict between the competing aims of maximizing rewards and maximizing information has come to be known as the exploration-exploitation dilemma (Kaelbling, Littman, and Moore, 1996, Sutton and Barto, 1998). It affects any organism that can learn: fish in the sea, ants on land (S. C. Pratt and Sumpter, 2006) and birds in the sky (Krebs, Kacelnik, and Taylor, 1978). It even affects slime molds growing in a petri dish (Watkinson et al., 2005). As humans, we are affected by the exploration-exploitation dilemma when we seek out a spouse, choose a career, or decide where to go on vacation. And, as scientists, we face this dilemma when we plan our next experiment.

In the case of a choice between going out to eat at Alice's or Bob's, you might resolve the exploration-exploitation dilemma using one of two rules of thumb:

- *"I'll try anything once"*: You could choose between Alice's and Bob's by deciding to try Bob's at least once before forming a judgment about the relative quality of the two restaurants. More generally, you might resolve the conflict between rewards and information by first maximizing information and then later maximizing rewards. This priorization of information before reward is normatively defensible because information can be used to maximize rewards, but rewards cannot be used to maxizize information.

- *"I'll flip a coin"*: You could choose between Alice's and Bob's by using a coin to introduce randomness into your decision-making so that there are equal odds of maximizing reward and maximizing information. More generally, you could set up a randomized decision rule that, on average, maximizes both reward and information in any desired proportions. Randomization over repeated decisions is necessary because any specific decision, when rewards and information lead

to inconsistent actions, can only be resolved in a way that maximizes at most one of the two objectives.

Both of these strategies (i.e. information-prioritization and randomization) can be surprisingly effective in establishing a balance between exploration and exploitation. In a refined form, my model of human decision-making will employ a combination of these two strategies in order to resolve the exploration-exploitation dilemma. Simple heuristic rules like the two I have listed have value because, in general, achieving the greatest amount of long term reward in the face of the exploration-exploitation dilemma is an intractably hard computational problem. The problem is so difficult theoretically that it has been the subject of over eight decades of mathematical research (Thompson, 1933; Robbins, 1952; Bubeck and Cesa-Bianchi, 2012). In that time, many different computational strategies for dealing with the dilemma have been developed: some of which are normatively defensible (Jones and Gittins, 1972; Thompson, 1933; Sutton and Barto, 1998; Auer, Cesa-Bianchi, Freund, and Schapire, 2002; Auer, 2003) and some of which have been validated empirically, even though their theoretical justifications are still imperfectly understood (e.g. Thompson, 1933).

Inspired by the extensive theoretical literature on the exploration-exploitation dilemma, many researchers in the biological sciences have studied the mechanisms by which animals strike a balance between the maximization of reward and the maximization of information. For example, in a seminal paper Krebs and colleagues demonstrated that the foraging behavior of birds was nearly optimally calibrated in a simple two-alternative choice task that required the birds to make trade-offs between exploration and exploitation (Krebs et al., 1978).

In contrast to the long history of theoretical study of the exploration-exploitation dilemma and the extensive literature on the dilemma's role in shaping animal foraging behavior, research on how humans cope with the exploration-exploitation dilemma is in its infancy. In this paper, we summarize existing work on human mechanisms

for coping with the exploration-exploitation dilemma and propose a broad theoretical framework that situates these empirical findings in the context of the theoretical literature. Because the existing empirical literature until now has been insufficient to answer many of the questions raised by the theoretical literature, I also present the results of several novel studies that fill essential gaps in our understanding of how humans cope with the exploration-exploitation dilemma.

Theoretical methods for resolving the exploration-exploitation dilemma can be classified by their use of one of three primary strategies for coping with imperfect information. These strategies can be viewed as elaborations of the two rules of thumb that we described earlier: they resolve the exploration-exploitation dilemma by either (a) explicitly trying to maximize the information available to the decision-maker in addition to maximizing reward or (b) introducing randomness into the decision-making process so that information is collected incidentally during attempts to maximize reward. Both of these approaches allow decision-makers to achieve a compromise between the pursuit of reward and the pursuit of information.

The first strategy I discuss is **information-based exploration**. Information-based exploration biases decision-making towards the most informative options whenever the decision-maker decides not to exploit the most rewarding option. This information-biasing is achieved by assiging each option a numeric score that takes into account both (a) the expected reward of that option and (b) the expected increase in long-term reward that could be achieved by having more information about an option. I will refer to the expected increase in long-term reward that results from having more information as the **value of information**. When I refer to an approximation of the value of information, I will call this approximation an **ambiguity bonus**, under the assumption that this bonus can be added to the expected reward of each option in order to account for the additional value of information. If calibrated correctly, information-based exploration can produce optimal behavior. But,

in general, the ambiguity bonus required for producing optimal behavior cannot be computed using available information and must be approximated using heuristics.

The second strategy is **noise-based exploration**. In this strategy the decision-maker always selects the option that has the highest expected reward, but introduces noise into the estimation of expected rewards. The noise in the evaluation process promotes exploration by ensuring that the option with highest expected reward is not always taken, but is neverthless more likely to be chosen than any other option. Thus, this strategy can be viewed as assigning a reward-biased score to each option.

The third strategy I will describe is **random exploration**, in which a majority of choices are resolved by selecting the best available option and a minority of choices are resolved by randomly selecting from all of the available options. While random exploration is suboptimal in the long run because it has an excessively high probability of selecting the worst possible option, it is computationally cheap and can be highly robust to errors in the decision-maker's model of the world.

The structure of this section is as follows:

(1) I formalize the exploration-exploit dilemma using a standard mathematical framework. This allows me to articulate the conflict I have described in rigorous terms and to present a model of decision-making that is applicable to any situation in which the goals of maximizing reward and maximizing information can be in conflict with one another.

(2) I use my model to analyze data from two experiments in which a person must make a decision between two unchanging options that offer probabilistic rewards. I elaborate my model's solution to this problem and then compare it both qualitatively and quantitatively with results from existing studies. I present new studies that demonstrate why the novel components of my theory are necessary for interpreting observed behavior in settings that involve exploration-exploitation dilemmas.

(3) I extend my model to cover cases in which there are more options than could

ever be experimented with directly, so that humans have imperfect information because they have no *direct* information at all about some options. In order to cope with this case, I must consider the general problem of generalization. For tractability, I focus on problems in which generalizations are done in simple spaces that can be presented as two-dimensional images.

Chapter 7

MaxInfo: A New Model of Exploration and Exploitation

In this section I develop my framework for describing human decision-making in the face of the exploration-exploitation dilemma. I will refer to the process a person uses to establish a trade-off between reward and information as the **decision process**. Because the decision-maker's goal is to maximize future rewards using only information about past rewards, I can isolate the decision process from an earlier step in which past experiences are summarized to produce predictions about future rewards. I will refer to this initial step as the **inference process**. This division has a natural interpretation in terms of time: the inference process looks backwards to answer the question: "based on past experience, what do I know?", whereas the decision process looks forward to answer the question: "based on what I know now, what should I do in the future?". This division in time is illustrated in Figure 7.1 and allows me to partition my model into one component that relates to psychological research on learning and memory and a second component that relates to research on judgment and decision-making. Before considering these processes in detail I first describe the simplified decision problem, known as the Multi-Armed Bandit problem, that I will

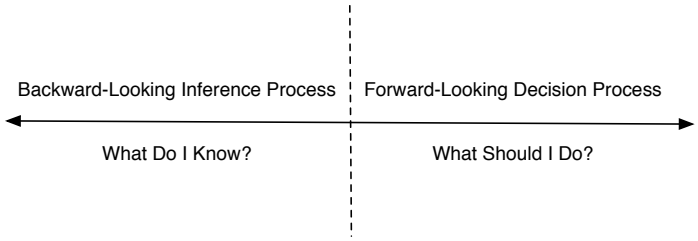


Figure 7.1: The Inference and Decision Processes

use to describe my framework.

7.1 The Multi-Armed Bandit (MAB) Problem

Rather than referring to a many-armed felon, **a Multi-Armed bandit** (MAB) is an abstraction and generalization of a type of slot machine, called a one-armed bandit, commonly found in a casino. A one-armed bandit is a gambling machine that pays out money probabilistically whenever a lever (or arm) is pulled. By analogy, a multi-armed bandit is an imaginary device with many arms. The assumptions of the MAB problem are that (1) each arm is associated with a different reward distribution, (2) only one arm can be pulled at a time and (3) the decision-maker wishes to maximize cumulative reward over a sequence of pulls. Thus, when faced with such a device, a decision-maker encounters an exploration-exploitation dilemma as they must decide which arm to pull on each trial, but have only imperfect information about the arm that will maximize reward if pulled repeatedly.

More formally I define K as the number of arms and T as the total number of trials, or time horizon, over which reward is accrued. I use t to denote the current trial number and thus on trial t there are $T - t$ trials remaining. I assume that the decision-maker's objective is to maximize the sum of rewards over all T trials.

I further assume that each time an arm, i, is played, the reward payout at time t, r_t, is sampled from a reward distribution associated with that arm, $p(r_t|\theta_t^i)$, where θ_t^i are the generative parameters associated with arm i at time t. For example, in the case of a Gaussian reward distribution, θ_t^i would refer to the mean and variance of the reward distribution for arm i at time t. In general, I use θ_t to denote the set of generative parameters for all K arms. For maximal generality, I assume that the generative parameters for any arm can change over time and can be correlated with the generative parameters of any of the other arms, although I will not discuss any distributions that change over time in this thesis.

A key element of the general MAB problem is that, even if the decision-maker has perfect information about the generative parameters, there can be inherent uncertainty in the actual rewards that they would receive on any given trial. This inherent uncertainty is known as the **risk** of arm i (J. W. Pratt, 1964), which was defined originally in the section on EU Theory in Chapter 2. I will quantify the risk of option i, $\mathcal{R}_i$, as the entropy of its reward distribution $p(r_t|\theta_t^i)$,

$$\mathcal{R}_i = \mathcal{H}\left(p(r|\theta_i)\right),$$

where $\mathcal{H}(p)$ denotes the entropy of distribution p. Other measures of uncertainty, such as variance, could also be used to quantify risk: I choose entropy because it provides a single metric that will also be used to measure another form of uncertainty that is relevant to the MAB problem. As will be seen, the experiments I have constructed were all designed to ensure that risk is held constant across all options to prevent confounding.

For concreteness, throughout this section, I will repeatedly employ an example of a simple MAB problem involving just two arms, A and B. This two-armed bandit pays out rewards sampled from a Gaussian distribution whose mean depends on the

choice of arm, but whose variance is the same for both arms: I refer to this constant variance as $\sigma^2_{\mathcal{R}}$. I also assume that the means of both arms are constant over time and given by θ^A for arm A and θ^B for arm B. Collectively, the generative parameters are $\theta_t = (\theta^A, \theta^B, \sigma_{\mathcal{R}})$. I will drop the time index t because I assume that these parameters do not change over time. Collectively these assumptions imply that the expected reward for playing option i is $\mu_i = \theta^i$ and that the risk associated with playing either arm A or B is

$$\mathcal{R}_i = \log \sigma_{\mathcal{R}} + \frac{1}{2} \log 2\pi + \frac{1}{2}.$$

7.2 The Inference Process

The inference process in the MAB problem involves computing what a decision-maker knows about each of the different arms based on the past outcomes associated with all arms. I assume that humans generalize from their past experiences using a Bayesian framework in which they maintain a probability distribution over the generative parameters for each arm that depends upon their history of choices and rewards. Although I employ a Bayesian framework, I note that any alternative statistical paradigm in which the estimated reward distributions associated with the bandit's arms are updated online could be employed as a model of the inference process.

I use c_t to denote the index of the arm that was chosen on trial t. I also use the shorthand $c_{1:t}$ as the set of choices from trial 1 up to and including trial t. Similarly I write $r_{1:t}$ for the set of past rewards. With this notation, the inferred probability over generative parameters is written as $p(\theta_t | c_{1:t}, r_{1:t})$.

This probability distribution is updated after each choice using a sequential Bayesian

update,

$$p(\theta_t|c_{1:t}, r_{1:t}) \propto p(r_t|\theta_t, c_t) \int d\theta_{t-1} p(\theta_t|\theta_{t-1}) p(\theta_{t-1}|c_{1:t-1}, r_{1:t-1}),$$

where $p(\theta_t|\theta_{t-1})$ describes the temporal dynamics of the generative parameters in the case in which the rewards associated with each arm can change over time. As noted, I will not need to deal with this fully general case for my specific experiments and can therefore simplify this equation to the form,

$$p(\theta|c_{1:t}, r_{1:t}) \propto p(r_t|\theta, c_t) p(\theta|c_{1:t-1}, r_{1:t-1}).$$

Note that, even if the reward distribution for any given arm has zero risk, my assumption that decision-makers maintain probabilistic beliefs about the reward distributions of the arms gives rise to a second source of uncertainty: namely, a form of uncertainty that reflects the imperfect information that the decision-maker has about the generative parameters, θ_i. This uncertainty is captured by the inferred distribution over the generative parameters $p(\theta_t|c_{1:t}, r_{1:t})$. The only case in which this additional sort of uncertainty is absent is when the decision-maker has perfect information about the parameters of the reward distributions and $p(\theta_t|c_{1:t}, r_{1:t})$ reduces to a delta function. Because this second form of uncertainty reflects the imperfection of decision-maker's knowledge of the world rather than a property of the world itself, it is commonly known as **ambiguity** (Knight, 2006, Ellsberg, 1961) in order to distinguish it from the form of uncertainty described as risk above.

The distinction between risk and ambiguity is essential to understanding the exploration-exploitation dilemma because the decision-maker's uncertainty about the reward distribution associated with any of the arms of a MAB can change based on new information. Specifically, as more samples from a particular arm are acquired, the ambiguity associated with that arm will, on average, decrease. Because the am-

biguity captures, in a quantitative way, the imperfection of the information that a decision-maker has, I will treat ambiguity as the measure of information that I assume decision-makers try to maximize. Although ambiguity reflects the absence of information, it makes no difference to my model whether I assume that decision-makers try to maximize information or that they try to minimize ambiguity. It is the assumption that decision-makers attempt to maximize information that leads to the name, MaxInfo, for my model of exploration-exploitation decisions.

In this paper, I will adopt an information-theoretic perspective and quantify the ambiguity, $\mathcal{A}_t$, at time t as the entropy of $p(\theta_t|c_{1:t}, r_{1:t})$, giving the formal definition,

$$\mathcal{A}_t = \mathcal{H}\left(p(\theta_t|c_{1:t}, r_{1:t})\right).$$

My model assumes that the decision-maker is able to use $p(\theta_t|c_{1:t}, r_{1:t})$ in order to compute the expected amount of reward, $\mu_{i,t}$, and the expected change in information, $I_{i,t} = \mathbb{E}[A_{t+1} - A_t|c_{t+1} = i]$, from playing option i. The expected reward on playing option i is given by,

$$\mu_{i,t} = \mathbb{E}\left(r_{t+1}|c_{t+1} = i\right),$$

where the expectation is taken over the reward outcome on trial $t+1$ conditional on arm i being chosen. That is

$$p(r_{t+1}|c_{1:t}, c_{t+1} = i, r_{1:t}) = \int d\theta_{t+1} \int d\theta_t p(r_{t+1}|\theta_{t+1}, c_{t+1} = i) p(\theta_{t+1}|\theta_t) p(\theta_t|c_{1:t}, r_{1:t})$$

The expected information is given by the expected change in ambiguity from playing option i,

$$I_{i,t} = \mathbb{E}\left(\mathcal{A}_{t+1}|c_{t+1} = i\right) - \mathcal{A}_t,$$

where, again, the expectation is over $p(r_{t+1}|c_{1:t}, c_{t+1} = i, r_{1:t})$. This measure captures the fundamental desideratum that the information gained from choosing option i

relates to both the total amount of ambiguity in the system and the extent to which choosing option i can resolve this systemic ambiguity.

In the most general case in which the arms' reward distributions (a) have arbitrarily complex marginal distributions, (b) are allowed to depend on the reward distributions of other arms and (c) allowed to change over time, the inference process required to estimate $p(\theta_t|c_{1:t}, r_{1:t})$ is intractable. As such, the measures of reward and information required by the decision process cannot be computed. Despite this general intractability, in many practical situations (such as those considered in this paper) the inference process can be reduced to relatively simple computations. For example, in the two-armed bandit example I have used throughout this section – in which (a) both arms have Gaussian distributions with known variance and (b) the generative parameters for the two arms are uncorrelated and constant over time – a complete mathematical analysis is simple. The independent and stationary Gaussians assumption allows the decision-maker's beliefs to be written in a factorized form:

$$p(\theta_t|c_{1:t}, r_{1:t}) = p(\theta_A|r^A_{1:t})p(\theta_B|r^B_{1:t})$$

where $r^A_{1:t}$ denotes only those rewards obtained from playing arm A. Assuming a Gaussian prior with mean μ_p and variance $\sigma^2_{\mathcal{R}}/v_p$ on θ_A and θ_B, I can further write the distribution over θ_i as a Gaussian with mean $m_{i,t}$ and variance $\sigma_{i,t}$, i.e.

$$p(\theta^i|r^i_{1:t}) = \frac{1}{\sqrt{2\pi}\sigma_{i,t}} \exp\left(-\frac{(\theta^i - m_{i,t})^2}{2\sigma^2_{i,t}}\right)$$

where the mean and variance are straightforwardly related to the observed rewards from that option, the risk variance and the number of times option i has been played. Using $v_{i,t}$ to denote the number of times that option i has been played, the inference

process for independent Gaussians reduces to the paired equations,

$$m_{i,t} = \frac{1}{v_p + v_{i,t}} \sum_{n=1}^{v_{i,t}} r_n^i,$$

and

$$\sigma_{i,t}^2 = \frac{\sigma_{\mathcal{R}}^2}{v_p + v_{i,t}},$$

where r_n^i denotes the nth reward obtained from playing option i. Thus the highly generic (and intractable) Bayesian update shown in equation 7.2 reduces to the efficiently computable problem of updating only two parameters, which can be done using simple closed-form delta rules. Furthermore, the expected reward and expected informational gain associated with option i also have closed-form expressions:

$$\begin{aligned} \mu_{i,t} &= m_{i,t} \\ I_{i,t} &= \frac{1}{2} \log \left(1 + \frac{1}{v_p + v_{i,t}} \right) \end{aligned}$$

Note that these two quantities are the only quantities required by the decision process that will be used to study the first two experiments I describe. I therefore now shift to describing the decision process. When the experiments I describe move beyond the simple settings in which this independent Gaussians inference model is adequate, I will present an alternative inference model that is more general – at the expense of greater conceptual complexity.

7.3 The Decision Process

Armed with the results of the inference process, I am now ready to describe the decision process. My goal is to construct a model that captures the key qualitative properties of any solution to an exploration-exploitation dilemma.

First, I expect that the decision process should reflect the drive to maximize

reward. Formally, the decision process should favor option i over option j if the expected reward from option i is greater than the expected reward from option j. Given the results of the inference process, this means that the decision process should be more likely to select option i than option j if $\mu_{i,t} > \mu_{j,t}$. I will call any decision that is consistent with this requirement an **exploitation-based** decision. Note that only a model of the inference process is required to determine whether any observed decision is exploitation-based or not.

Second, I expect that the decision process should reflect the drive to maximize information. Formally, the decision process should favor option i over option j if the expected informational gain from option i is greater than the expected informational gain from option j. Given the results of the inference process described earlier, this means that the decision process should be more likely to select option i than option j if $I_{i,t} > I_{j,t}$. I will call any decision that is consistent with this requirement an **exploration-based** decision. Again, I note that only a model of the inference process is required to determine whether an observed decision is exploration-based or not. Thus the simple classification of decisions into exploitation-based and exploration-based classes can be done with much weaker modeling assumptions than are required to describe my full model.

Finally, I expect that decision-makers must have a mechanism for adjucating between maximizing rewards and information. Because the exploration-exploitation dilemma arises whenever the option favored by **exploitation-based** decision-making is not the option favored by **exploration-based** decision-making, I know that some compromise must be established between these two competing objectives. Mathematicians that study optimization theory refer to problems in which there are multiple, irreconciable objectives as multi-objective optimization problems (Marler and Arora, 2004). I treat the exploration-exploitation dilemma as a multi-objective optimization problem and adopt one of the classic approaches to solving multi-objective optimiza-

tion problems: I assume that behavior attempts to maximize a weighted combination of both objectives. Specifically, I assume that the chosen option on trial t is selected to maximize a linear combination, which leads to the functional form of the full MaxInfo model in which the chosen option at time t is chosen to satisfy the requirement,

$$c_t = \arg\max_i \beta_\mu \mu_{i,t} + \beta_I I_{i,t}.$$

As stated, this model is fully deterministic and reflects the purest form of an *information-based* strategy for exploration as described in Chapter 6. There are many normative methods for resolving the exploration-exploitation dilemma that are deterministic in this fashion (e.g. Auer, 2003; Frazier, Powell, and Dayanik, 2008), but human behavior typically involves random fluctuations that are not well-captured by deterministic models – as noted in the section on Random Utility Theory in Chapter 2.

I therefore augment this simple model of exploration-exploitation by incorporating random noise terms, $n_{i,t}$, to give a new decision rule:

$$c_t = \arg\max_i \beta_\mu \mu_{i,t} + \beta_I I_{i,t} + n_{i,t}.$$

This decision noise may reflect variability in any one of the neural mechanisms underlying decision-making: for example, noise may corrupt signals from working or short-term memory stores that maintain the results of the inference process. Alternatively, noise may corrupt the processing of these signals in regions involved in the decision process. Noise may even occur at the motor output stage.

For my analyses, I need to make no assumptions about the precise location(s) in the brain at which noise occur(s). In my model, I require only that the noise terms, $n_{i,t}$, be sampled from a fixed noise distribution, $p(n)$. In order to make mathematical modeling of the decision process tractable, I assume that the noise terms are all

independent draws from a logistic distribution with fixed variance σ_n^2.

By averaging over many draws of these logistic noise terms, I can calculate the probability of choosing option i on trial t as,

$$p(c_{t+1} = i) = \frac{1}{Z} \exp\left(\frac{\beta_\mu \mu_{i,t} + \beta_I I_{i,t}}{\sigma_n}\right),$$

where the normalization constant Z is given by

$$Z = \sum_j \exp\left(\frac{\beta_\mu \mu_{j,t} + \beta_I I_{j,t}}{\sigma_n}\right).$$

I note that the incorporatation of randomness into my model of the decision process does not merely provide a superior model of observed human behavior: it may also provide a normatively superior method for coping with the exploration-exploitation dilemma if the inference process is likely to produce results that are biased in some way away from the true generative parameters, θ_t. When I estimate the value of these noise parameters in the next section, we will see that the noise terms that I originally added to Utility Theory in order to cope with model misfit may, in fact, reflect a normatively inspired strategy for coping with imperfect information – a concept not discussed at all when defining Neoclassical Utility Theory.

Indeed, the normative use of models that incorporate decision noise has a long precedent in theoretical analysis of the exploration-exploitation dilemma (Sutton and Barto, 1998). Many classical methods for resolving the exploration-exploitation dilemma do not incorporate any attempt to maximize information in order to promote exploration: rather, these methods promote exploratory decisions by adding random noise into a process that otherwise tries to maximize rewards alone.

For example, a special case of my model in which no weight is placed on information is the following,

$$c_t = \arg\max_i \beta_\mu \mu_{i,t} + n_{i,t}.$$

This special case of the MaxInfo model is exactly equivalent to the popular softmax strategy for resolving the exploration-exploitation dilemmas. The softmax strategy is used in most reinforcement learning models of human behavior (Daw, O'Doherty, Dayan, Seymour, and Dolan, 2006) and, as such, it is desirable that it be a special case of any new model of decision-making. More generally, this softmax model is a specific instance of the noise-based exploration strategy I discussed earlier. Because rewards are corrupted by noise before making a selection of actions, options that seem inferior based purely on the results of the inference process nevertheless have non-zero probability of being explored, which ensures that the decision-maker slowly overcomes the imperfect information they possess over time.

Alternative randomization strategies can also be employed: a popular normative method called the ϵ-greedy algorithm resolves the exploration-exploitation dilemma by maximizing reward with a fixed probability $1 - \epsilon$ and selecting completely at random from all available options with probability ϵ. This kind of exploration strategy cannot be expressed by my model, but a trivial extension of my model could incorporate such a noise term and might be able to capture "mistakes" made by experimental participants that are not well-explained by the logistic noise model I use. Differentiating this ϵ-greedy noise from the kind of logistic noise used by my model is a difficult problem that would require highly targeted experiments, but is possible in theory because the ϵ-greedy and softmax strategies have very different behaviors in their asymptotic regimes.

7.4 Inference and Decision Combined

In summary, the MaxInfo model works as follows:

- Decision-makers summarize their experiences using an inference process that estimates both the expected reward, $\mu_{i,t}$, associated with each option at time t

and the amount of information, $I_{i,j}$, that the decision-maker would gain from selecting each option.

- Decision-makers reconcile the incompatible goals of maximizing reward and information by combining both objectives into a single utility score for each option. These utilities are assigned to options by combining the expected change in cumulative reward, expected change in information and also random decision noise.

These two processes – the inference process and the decision process – collectively describe a conceptually complete model of how humans might cope with the exploration-exploitation dilemma. I now proceed to describe several experiments in which I have calibrated this model using experimental data, noting that the Max-Info model allows me to assess the way in which the trade-offs between exploration and exploitation are adjusted by subjects in response to structural changes in their environment.

Chapter 8

Experiments on Exploration and Exploitation

8.1 Designing Exploration-Exploitation Experiments

In order to study human strategies for resolving the exploration-exploitation dilemma, I have developed several experimental paradigms. The simplest of these paradigms is a variant of the two-arm MAB problem described in the previous chapter. In this experimental paradigm, participants make a sequence of choices between two arms over the course of T trials. Each arm generates random rewards sampled from an arm-specific probability distribution and the participant sees (and receives) only the reward of the arm that was chosen. I describe this task to participants using a cover story about two slot machines that are found inside a hypothetical casino. An illustration of this task is shown in Figure 8.1. I refer to this experimental design as the **two-bandit** task.

The two-bandit task is the simplest possible context in which an exploration-exploitation dilemma can be observed in the lab because there must be at least two arms for there to be substantive decision-making and there must be at least two trials

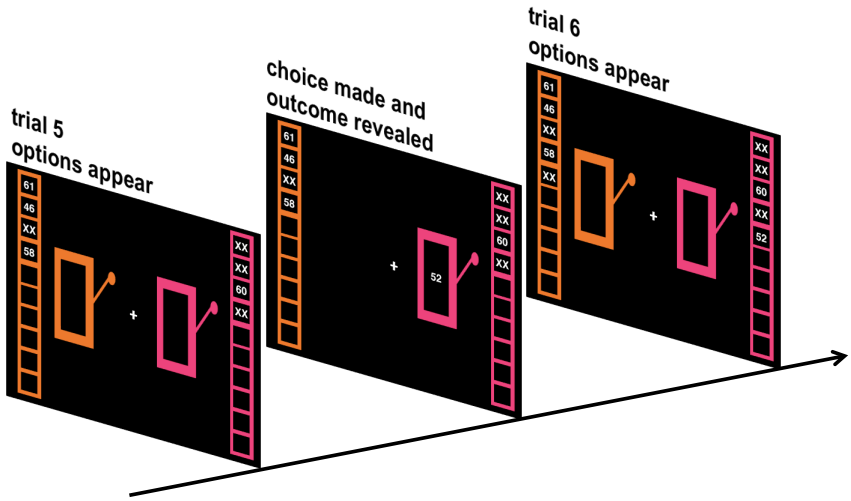


Figure 8.1: The two-bandit task as it appears to experimental participants

for there to be any incentive for participants to explore ambiguous options. Although the two-bandit task is a very simplified instance of the general exploration-exploitation dilemma, it nevertheless poses several difficulties that are not often encountered in other branches of experimental psychology. One problem of experimental design is particularly acute: a naive two-bandit task introduces a severe form of selection bias that is caused by the experimental participant being able to shape their own experiences. I will now describe this concern in detail before describing any specific experiments because resolving this confound is essential to correctly interpreting the results of any experiment on exploration and exploitation.

Selection Bias: Because the participant is allowed to decide which arms to pull during a two-bandit task, the information that the participant has about the arms is under their own control and is not an experimentally controlled variable. This introduces a subtle form of selection bias that can corrupt many analyses. A particularly acute problem resulting from this selection bias is that participants will tend to acquire more information about options that seem to offer higher reward: this implies that the amount of information that a participant has about any option

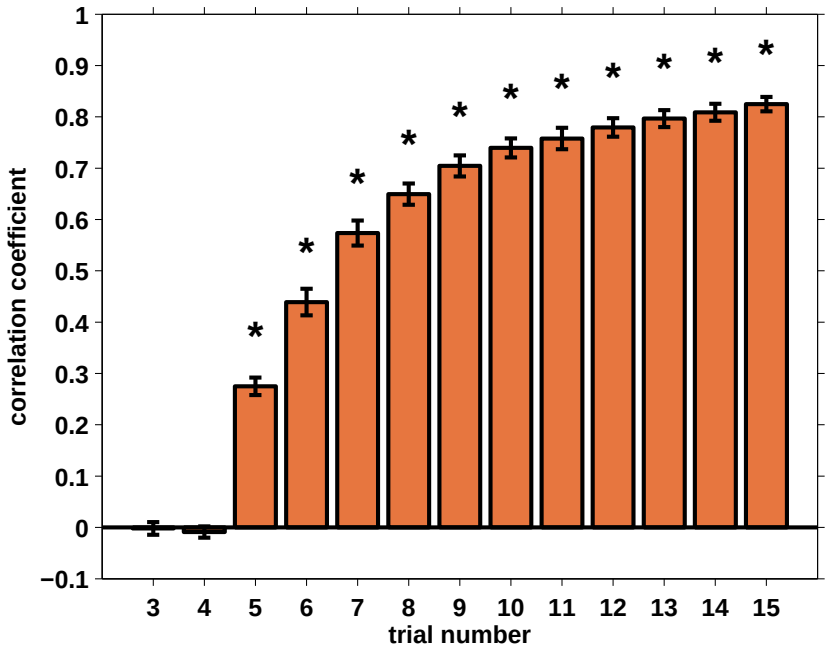
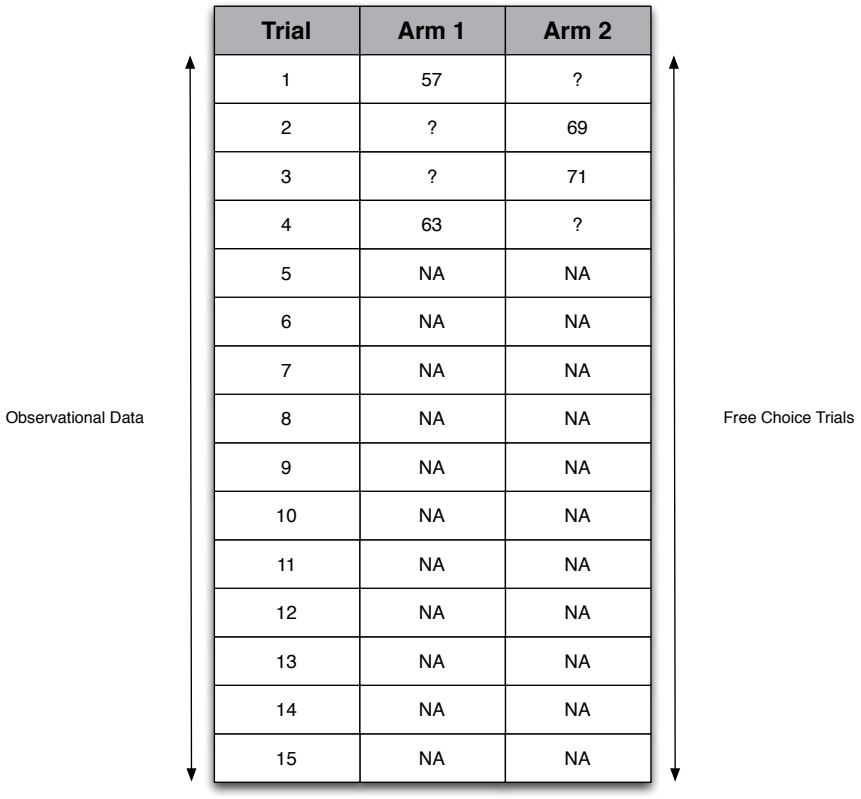


Figure 8.2: Correlation between Reward and Information

will become correlated with that option's expected reward over time. Trials towards the end of an exploration-exploitation experiment can therefore be almost completely uninformative about the strategies that participants use to resolve the exploration-exploitation dilemma, which is, after all, defined in terms of a trade-off between maximizing reward and information.

An example of this growing correlation between reward and information is shown in Figure 8.2 in which the correlation between the participant's knowledge of arm A becomes increasingly correlated with the participant's assessment of arm A's value relative to arm B. In the specific experimemt being shown below, this correlation only emerges starting at trial 5 because the participant is not allowed to make their own decisions on trials 1-4. I will elaborate on the concept of imposing a participant's choices in a moment.

The growing correlations seen in 8.2 can be interpreted as a failure of experimental randomization, upon which any sound statistical analysis depends (Rubin, 1980). To



Trial	Arm 1	Arm 2
1	57	?
2	?	69
3	?	71
4	63	?
5	NA	NA
6	NA	NA
7	NA	NA
8	NA	NA
9	NA	NA
10	NA	NA
11	NA	NA
12	NA	NA
13	NA	NA
14	NA	NA
15	NA	NA

Figure 8.3: Two-Bandit Task without Imposed Choice Trials

see how this correlation can be prevented, consider the two-bandit task shown in Figure 8.3. As can be seen in 8.3, on trial 5 the participant has chosen each of the two arms twice, receiving rewards of 57 and 63 from arm 1 and rewards of 69 and 71 from arm 2. During subsequent trials, we can therefore expect that participants will tend to choose arm 2 more often.

I resolve this problem by introducing a manipulation in which the participant is randomly assigned to making certain choices at the start of the exploratory period of decision-making. These randomly assigned choices must be made precisely as stated in order for the experiment to proceed. On each of these **imposed choice** trials, the participant is presented with two arms and then required to choose a specific one of these arms in order to move on. This imposition of choice has the effect of randomly assigning participants to a history of choices. As a result, all details of the history

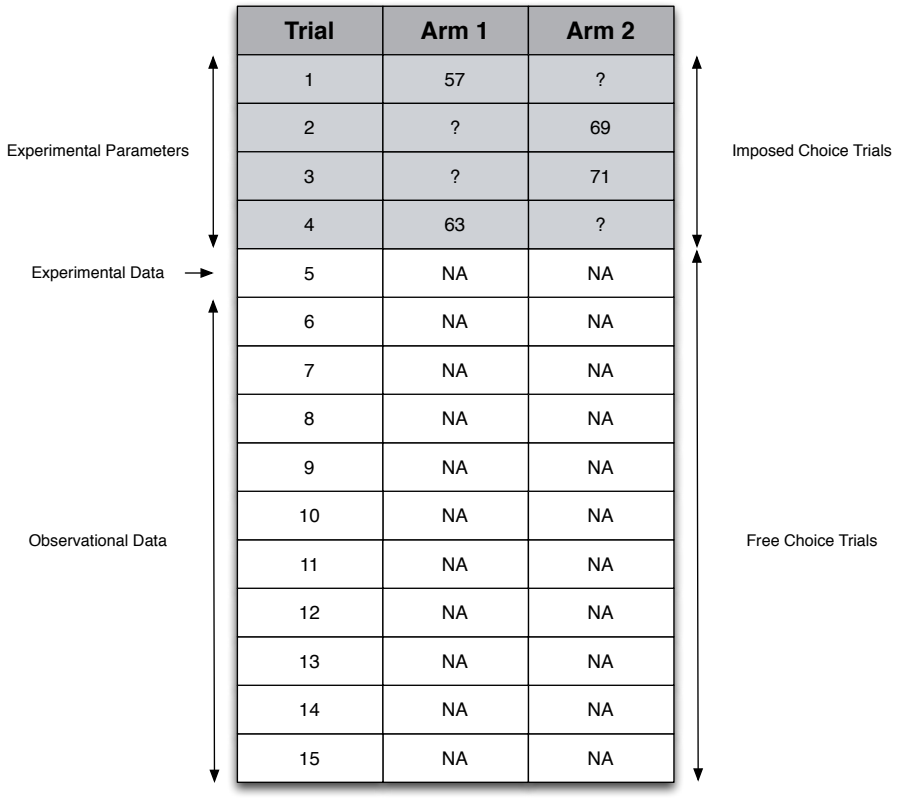


Figure 8.4: Two-Bandit Task with Initial Imposed Choice Trials

of choices can be controlled: in particular, this makes it possible to decorrelate the amount of information that a participant has about one arm from the expected rate of reward that is associated with that arm.

See Figure 8.4 for an example of a two-bandit experiment in which the first four trials are imposed choice trials. By modifying the structure of these four imposed choice trials, I can manipulate the information that the participant has about each arm independently from the difference in means between the arms. As such, the fifth trial, indicated with an arrow, is a true experimental trial in which the variables of interest are fully under the control of the experimenter. This allows me to explore the structure of human exploration and exploitation systematically in a way that is unbiased by the participant's own decisions.

In general, my approach is to employ a set of initial imposed choice trials inside

of a two-bandit experiment in which several additional variables are manipulated. The imposed choice approach has some weaknesses: because the initial choices are not freely chosen, but imposed, they may not have the same effects as naturalistic choices. I have not yet devised a solution to this concern. One could attempt to simply describe the outcomes of previous choices, but this runs the risk of running afoul of the Description-Experience divide that has been found in other studies of decision-making (Hertwig, Erev, et al., 2009). I therefore employ the imposed choice paradigm because it is more statistically sound than other procedures that have been used in the past.

I raise basic issues of experimental design before describing my work because these issues are central to understanding the results of two-bandit experiments. Many previous studies have not accounted for the selection bias that is introduced by decision-makers having control over their own experiences. As a result, these studies are often underpowered and may even suffer from systematic bias in their estimation of the relative role of reward and information in shaping decisions. By employing an experimental paradigm that removes this selection bias, I have attained that results are both (a) dissimilar from previous work in psychology and (b) more consistent with the predictions of normative models of exploration and exploitation.

In my studies of the two-bandit task, I have focused upon two classes of manipulations that occur in addition to the use of imposed choice trials: I have manipulated the horizon over which decisions are made and the incentive structures that decision-makers face. These manipulations are described in detail in the following two sections: both are designed to provide the experimenter with tools for shifting the relative balance of exploration and exploitation in shaping observed decisions in order to assess whether the MaxInfo model captures the dynamics of choice in changing environments.

8.2 The Effects of Manipulating Horizons

Introduction: As stated earlier, the number of trials remaining on which a decision-maker can choose among the K arms of a multiarmed bandit is called the **horizon**. Because longer horizons provide more opportunties to take advantage of the information gained by exploration, normative guidelines for behavior suggest that exploration will be increased if the horizon is extended. By identifying exploration with information-seeking behavior, my model makes it possible to directly measure exploration in a way that is not confounded with chance variability in responding. As such, my model provides a mechanism for testing the normatively-inspired prediction that extended horizons will increase exploratory behaviors. The hypothesis that longer horizons are required to promote exploratory behaviors also provides a mechanism for reconciling two literatures on the role of ambiguity in decision-making that, *prima facie*, seem to draw opposing conclusions. I review these literatures briefly below.

In economics, there is a long-line of research demonstrating ambiguity aversion – a term used to describe decisions from which the MaxInfo model would infer the presence of an ambiguity cost rather than an ambiguity bonus. The economic literature on ambiguity aversion begins with the seminal paper of Ellsberg (Ellsberg, 1961; Raiffa, 1961), in which a thought experiment was put forward that has subsequently come to be known as the Ellsberg Paradox. The Ellsberg Paradox involves decisions between three urns out of which a decision-maker will be allowed to draw exactly one ball. Three urns like those used in the original example are shown in Figure 8.5.

In the first urn, there are 10 blue balls. For now, I will assume that pulling out a blue ball will lead to a reward of $20. As such, the expected value of a draw from the first urn is $20. I will refer to the first urn as the **certain** urn.

In the second urn, there are 5 green balls and 5 red balls. The red balls are worth $0, whereas the green balls are worth $40. As such, the expected value of a draw from the second urn is also $20, but there is risk involved with each draw from the

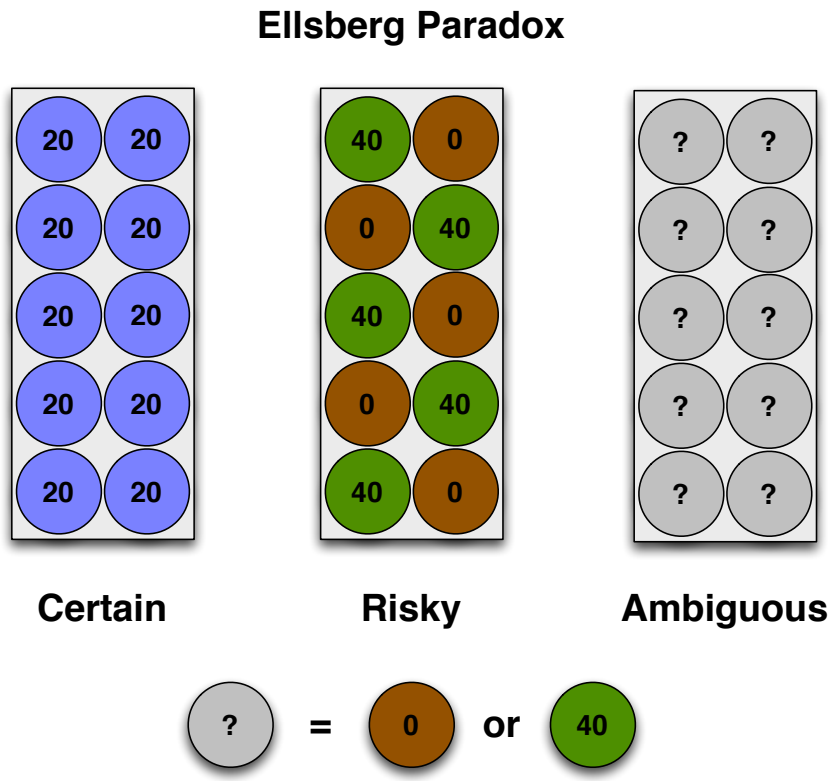


Figure 8.5: The Ellsberg Paradox

second urn because any given outcome is as likely to produce $0 as to produce $40. I will refer to the second urn as the **risky** urn. Although the MaxInfo model does not include any form of risk aversion, a large body of previous research has demonstrated that most people will select the certain urn over the risky urn – a phenomenon often referred to as **risk aversion** (J. W. Pratt, 1964).

Ellsberg's examples goes beyond risk aversion by introducing a third urn in which there are 10 balls of unknown color. The decision-maker is told that each of these 10 balls is either red or green, but the decision-maker does not know how many balls are red and how many are green. In Figure 8.5, I illustrate this by using grey balls to denote the fact that the decision-maker is uncertain not only which ball he or she would get from a random draw from the urn, but is also uncertain whether any given ball in the third urn is red or green. Because the decision-maker has imperfect information about the composition of the third urn, it is possible that the third urn contains only red balls. But it is also possible that the third urn contains only green balls. Furthermore, any midpoint between these two extremes is also possible. I therefore refer to the third urn as the **ambiguous** urn.

Many experiments have employed variants of Ellsberg's example: these experiments consistently find that most human beings react to this type of ambiguity by avoiding the third urn – a phenomenon often referred to as **ambiguity aversion** (Ellsberg, 1961). The results of experiments based on Ellsberg's example are typically taken as evidence that humans are, *in general*, ambiguity-averse. I summarize the extensive literature on ambiguity-aversion in economics because the assertion that humans are generally ambiguity-averse stands in sharp contrast to the literature in machine learning, which suggests that ambiguity-seeking behavior is essential for decision-makers that suffer from imperfect information. The canonical example of ambiguity-seeking in machine learning is the Upper Confidence Bounds (UCB) algorithm, in which arms are assigned a reward estimate that is the upper bound on a

confidence interval rather than its midpoint (Auer, 2003). This upper bound approach has the effect of giving priority to an uncertain arm over a certain arm if both arms have the same expected value. The empirical success of the UCB approach would lead one to conclude that an ambiguity bonus is essential for successful behaviors in uncertain environments.

I explain this discrepancy in terms of the horizons over which both fields traditionally study behavior. Economists, inspired by Ellsberg's example, have focused on one-shot decisions in which there is no value to acquiring information. In contrast, computer scientists have focused on repeated decisions in which the long (and sometimes infinite) horizon over which decisions take place makes the exploration of uncertain environments essential. I hypothesize that human decision-making exhibits ambiguity aversion over short horizons, but becomes ambiguity-seeking over long horizons.

To test this hypothesis, I manipulate horizons experimentally in a fully randomized design that allows me to directly measure the effect of extended horizons on human exploratory decision-making. This allows us to demonstrate that (1) humans can situate their behavior appropriately along a continuum from ambiguity-averse to ambiguity-seeking behavior and that (2) the strategies humans use to reason about exploration depend on horizons, which distinguishes human decision-making from many theoretical models of exploration in which the horizon is not involved in determining behavior. Policies for decision-making that do not depend upon horizons are typically called **anytime** policies: the standard ϵ-greedy, softmax and UCB-algorithms used in machine learning are all examples of anytime policies.

I thus adjudicate between ambiguity-averse and ambiguity-seeking dynamics in behavior at the same time that I test whether humans employ anytime policies. As will be seen in the results of my experiments, I find systematic increases in the amount of exploration that participants perform as horizons increase and use my

Trial	Arm 1	Arm 2
1	57	?
2	?	69
3	?	71
4	63	?
5	NA	NA

Trial	Arm 1	Arm 2
1	57	?
2	?	69
3	?	71
4	63	?
5	NA	NA
6	NA	NA
7	NA	NA
8	NA	NA
9	NA	NA
10	NA	NA

Trial	Arm 1	Arm 2
1	57	?
2	?	69
3	?	71
4	63	?
5	NA	NA
6	NA	NA
7	NA	NA
8	NA	NA
9	NA	NA
10	NA	NA
11	NA	NA
12	NA	NA
13	NA	NA
14	NA	NA
15	NA	NA

Figure 8.6: Two-Bandit Task with Changing Horizon

model to measure the precise extent of these changes. This provides evidence for our hypothesis that humans exhibit both ambiguity-averse and ambiguity-seeking behavior in appropriate contexts and that the discrepancies in the conclusions reached in economics and machine learning reflect dissimilarities in the two fields' domains of study.

Methods: To study the effects of manipulating horizons, I employed a two-bandit design in which the horizon varied between 5, 10 and 15 trials. A schematic of this design, which I call the Horizon task, is shown in Figure 8.6.

As in all two-bandit tasks, the Horizon task asked the participant to solve many brief MAB problems, each of which was independent from all of the others. I refer to each one of these brief MAB problems as a game. In the Horizon task, each participant played 150 games in a single hour-long experimental session. Across games, the horizons were equally likely to be 5, 10 or 15 trials. As such, the entire experiment lasted for 1500 trials.

At the start of each game, the means of the two arms were drawn from a list of

values ranging from 30-70 in increments of 5. The mean of arm 1 was drawn first and then the mean of arm 2 was drawn in order to insure that the difference between the two arms was -10, -5, 0, +5 or +10 with equal probability. The difference in the means of the two arms was counterbalanced across horizons. Thus, either of the two arms was as likely to be the better arm on average and the degree of difference between arms varied from a large absolute difference of 10 points to no difference at all.

Before the start of the experimental session, the participants were shown instructions that explained that their objective was to maximize the number of points they earned over the course of the entire experiment. Participants were further informed that they would be paid an additional wage for performing well in proportion to the number of points they earned during all of the games. In total, I recruited 38 undergraduate students from Princeton University, all of whom completed the required 1500 trials in exchange for monetary compensation.

In light of the issues of experimental design described in the previous section, I ensured that I could estimate the effects of reward and ambiguity in an unbiased way by imposing the participant's choices during the first four trials of each game. I then analyzed only the data from the fifth trial of each game, which is the only trial in which both (a) participants made their own decisions and (b) their experiences with the world were randomly assigned. Both of these steps were taken to insure that I can observe decision-making when the participant has seen 2 draws from both arms and also when the participant has seen 3 draws from one arm and only 1 draw from the other. By contrasting these two levels of relative ambiguity, I could directly test the effect of ambiguity in promoting exploratory decision-making in a way that is not confounded with the participant's beliefs about the relative superiority of either arm.

Results: To assess the presence of an ambiguity bonus, I first measured the probability that a participant made a decision consistent with exploitation by estimating

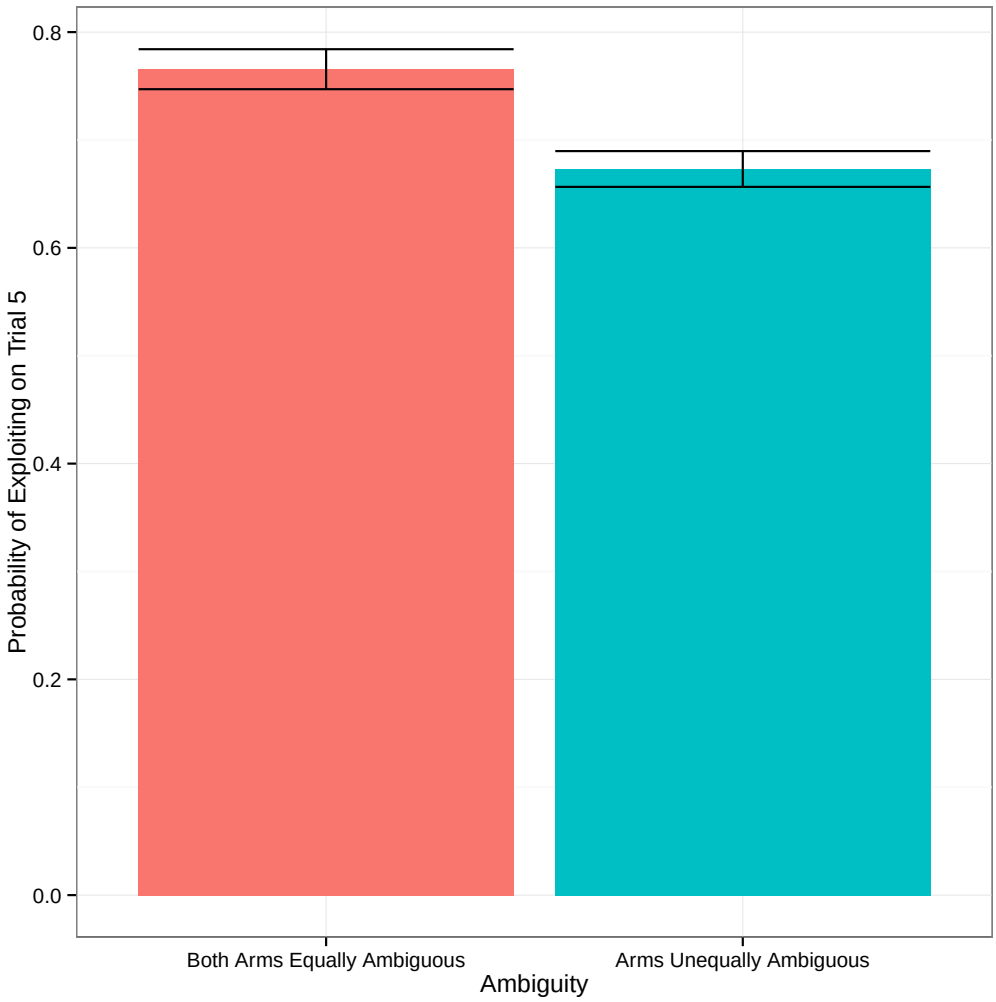


Figure 8.7: Two-Bandit Task with Changing Horizon

the frequency with which participants chose the option that seemed, based on previous experiences, to be the most rewarding. For example, if the observed mean for arm 1 was greater than the observed mean for arm 2 and arm 1 was chosen on trial 5, I classified this as an exploitation-based decision. This classification also applied to the symmetric case in which the observed mean for arm 2 was higher than the observed mean for arm 1 and arm 2 was chosen. The results of this model-free estimate of exploitation-based decision-making are shown graphically in Figure 8.7 and were highly significant under a t-test ($t(4494.442) = 7.1739$, $p < 8.486e^{-13}$).

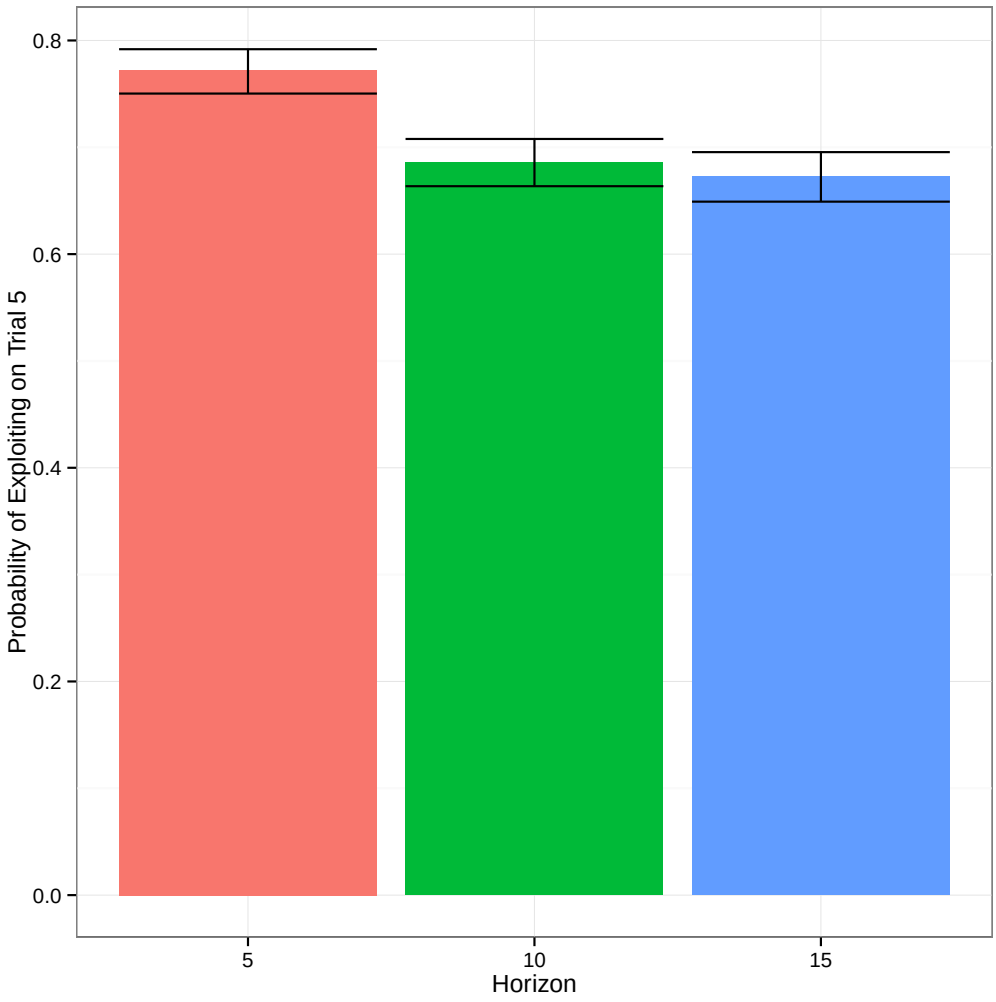


Figure 8.8: Two-Bandit Task with Changing Horizon

Using this same model-free assessment of exploitation, I also estimated the effects of horizons on shifting decision-making towards exploitation. The probability of making an **exploitation-based** decision for each of the three horizons (5, 10 and 15) is shown graphically in Figure 8.8 and was also highly significant when assessed for a linear trend using logistic regression ($z(4861) = -6.232$, $p < 4.62e^{-10}$).

Together, these results show that participants were (a) more likely to explore when there was information to be gained by exploration and (b) more likely to explore when the horizon was long enough that the information gained through exploration

could be put to use. Both of these are predicted by normative models of sequential decision-making, but have not been shown before in the empirical literature on human decision-making. Indeed, my finding that the presence of differential relative ambiguity promotes exploration stands in contrast to the results of Daw et al. (2006), who found no evidence of an ambiguity bonus. This discrepancy may reflect either (a) the distinct experimental approach they employed or (b) the distinct model assessment approach they took.

To make observations of heightened exploration more precise, I fit our formal model of exploration-exploitation dynamics to the pooled behavior of all participants. In Figure 8.9, I plot the parameters of our model in two-dimensions in order to emphasize the way in which decision-making moves from being strongly ambiguity-averse to being weakly ambiguity-seeking when the horizon grows larger. This change has two correlates in Figure 8.9: (1) the amount of weight placed on maximizing reward decreased and (2) the amount of weight placed on maximizing information increases as horizons increased. Visually, this is reflected as a movement of the model parameters that goes up and to the left as horizons grow longer. Numerically, these changes in parameter estimates are shown in Table 8.1.

Table 8.1: Model Parameters for Horizon Task

Horizon	Reward Weight (β_μ)	Info Weight (β_I)
5	1.80 ($sd = 0.11$, $p < 2^{-16}$)	-0.27 ($sd = 0.075$, $p < 0.0003$)
10	0.93 ($sd = 0.070$, $p < 2^{-16}$)	0.039 ($sd = 0.062$, $p < 0.5$)
15	0.71 ($sd = 0.066$, $p < 2^{-16}$)	0.13 ($sd = 0.061$, $p < 0.03$)

8.3 The Effects of Manipulating Incentives

Introduction: In the standard two-bandit task, there is a mixed incentive system: although the participant's ultimate incentive is to acquire the greatest amount of

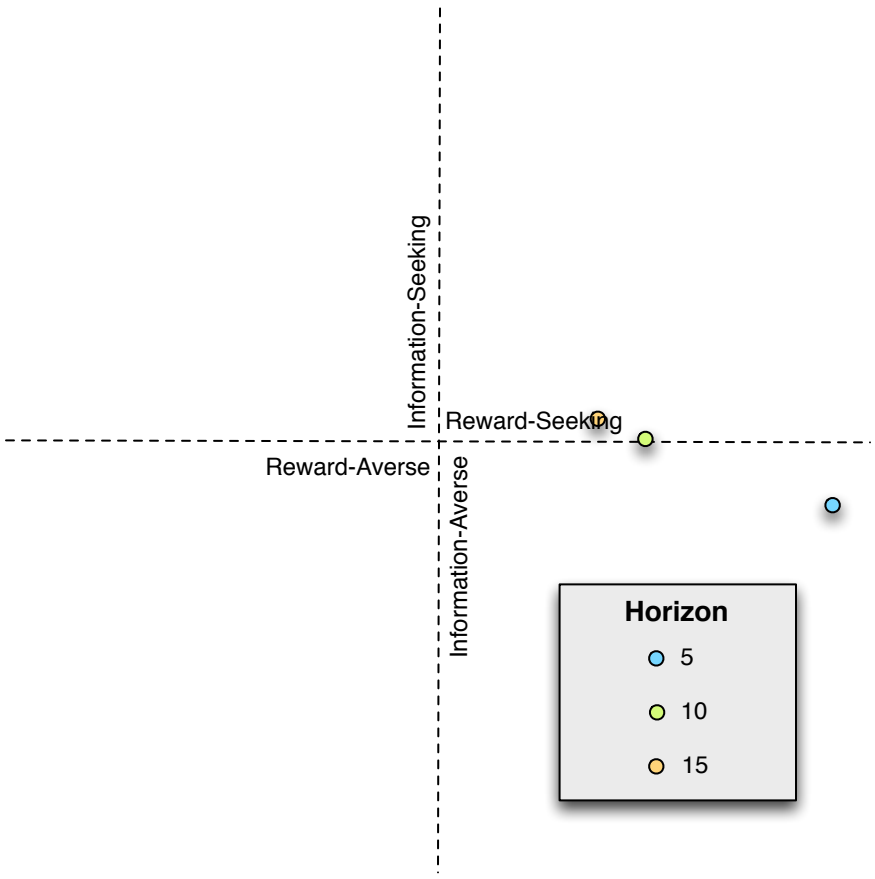


Figure 8.9: Two-Bandit Task with Changing Horizon

cumulative reward, this induces a proximate incentive to acquire information. For that reason, one can view a two-bandit task as one in which there are *direct* incentives to maximize both rewards and information. The MaxInfo model allows one to measure the relative importance of these incentives on determining decisions in any specific task.

But the relative importance of these two types of incentives can also be directly manipulated by the experimenter in order to observe the mechanisms of exploration and exploitation in isolation from one another. By separating the incentives for reward and information, I have designed an experimental paradigm in which it is possible to directly encourage participants to explore or exploit on a per-trial basis. I refer to this as the Reward vs. Information manipulation.

Methods: In the Reward vs. Information manipulation, participants are incentivized, on a single trial, for exactly one of three different combinations of reward and information:

1. **Incentivize Reward + Information**: After a series of imposed choice trials, a participant performs a standard two-bandit task in which every decision they make provides both reward and information. This incentive structure is equivalent to a standard bandit task.

2. **Incentivize Reward**: After a series of imposed choice trials, a participant makes one or more decisions in which they receive the rewards associated with their actions, but are provided with no informative feedback about these rewards. By removing all feedback, this incentive structure removes any incentive to maximize information and should produce pure exploitation. In the experiment I describe, I perform this manipulation on only one trial; all previous and subsequent trials occur as they would in a standard two-bandit task. An example of this single-trial modification approach is shown in Figure 8.10.

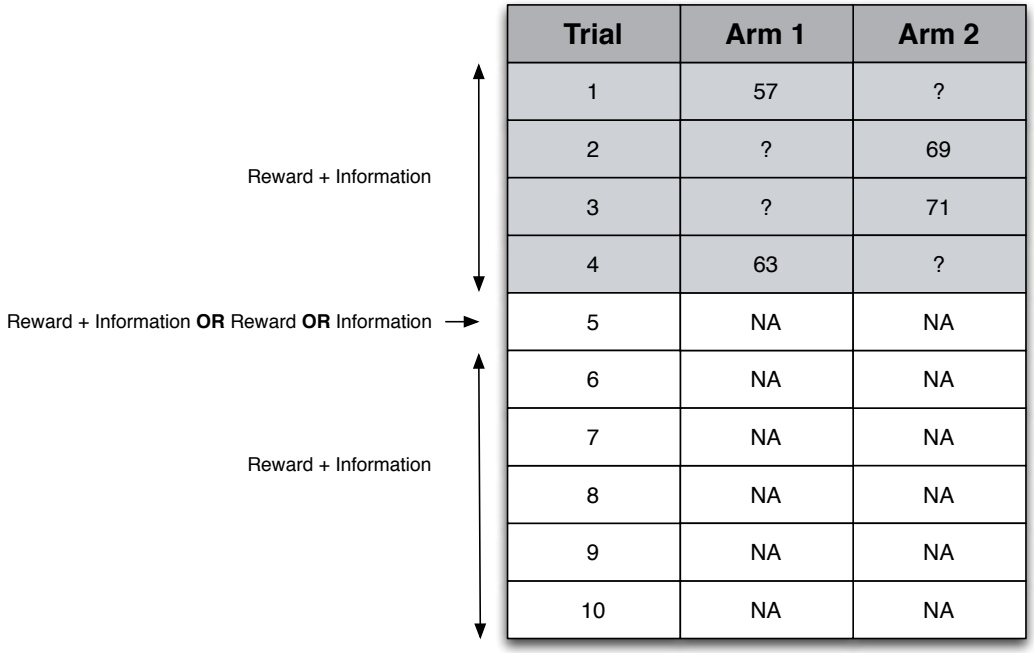
3. **Incentivize Information**: After a series of imposed choice trials, a participant makes one or more decisions in which they receive information about the rewards they would have received from their actions, but do not receive the actual rewards. By removing the rewards, this incentive structures removes any incentive to maximize reward and should produce pure exploration.

By randomly assigning the incentive structure on a per-trial basis, I am able to assess the flexibility of behavior and the capacity of participants to dynamically adapt their behavior to changing circumstances. Demonstrating that behavior is intentionally modified in response to incentives is essential for demonstrating that the modifications of behavior seen in the Horizon task described earlier are truly made in order to maximize information – as postulated by the MaxInfo model.

To assess the effects of incentive structures on behavior, I conducted a two-bandit experiment in which 31 Princeton undergraduate students were recruited to complete 168 brief MAB problems, which I again call games. Each of these games employed a randomly selected incentive structure from the three structures of Reward, Reward + Information and Information. Each game lasted for 10 trials, providing 1680 trials of data per participant.

At the start of each game, the means of the two arms were drawn from a list of values ranging from 30-70 in increments of 5. The mean of the arm 1 was drawn first and then the mean of arm 2 was drawn in order to insure that the difference between the two arms was -10, -5, 0, +5 or +10 with equal probability. The difference in the means of the two arms was counterbalanced across incentive structures. This is the same procedure as employed in the Horizon study: it was employed again in order to achieve the same effect of assessing the effects of reward over a parametric set of differences in mean reward for the two arms.

Before the start of the experimental session, the participants were shown instructions that explained that their objective was to maximize the number of points they



Trial	Arm 1	Arm 2
1	57	?
2	?	69
3	?	71
4	63	?
5	NA	NA
6	NA	NA
7	NA	NA
8	NA	NA
9	NA	NA
10	NA	NA

Figure 8.10: The effects of changing incentive structures in a two-bandit task on exploitation-based decisions

earned over the course of the entire experiment. Participants were further informed that they would be paid an additional wage for performing well. In total, I recruited 31 undergraduate students from Princeton University, all of whom completed the required 1680 trials in exchange for monetary compensation.

As in the Horizon task, I insured that I could estimate the effects of reward and ambiguity in an unbiased way by imposing the participant's choices during the first four trials of each game. I then analyzed only the data from the fifth trial of each game, which is the only trial in which the incentive structure could change. On the sixth through tenth trials, the incentive structure reverted to the standard Reward + Information structure used in all of the other experiments. The inclusion of these additional trials is essential because our previous results demonstrated that a 10-trial horizon is sufficient to encourage exploration when the standard Reward + Information incentive structure is employed.

In addition, I manipulated the relative ambiguity of both arms using the same imposed choice manipulation as in the Horizon task. As in the Horizon task, these imposed choices were designed to insure that I could observe decision-making when the participant had seen 2 draws from both arms and also when the participant had seen 3 draws from one arm and only 1 draw from the other.

Results: I then employed the same metric of exploitation used in the Horizon task to measure the effects of incentive changes in a model-free fashion. The results of this analysis are shown graphically in Figure 8.11 and were highly significant under an ANOVA ($F(2) = 73.5$, $p < 2e^{-16}$).

I followed up on our model-free analysis by estimating the parameters of our model in each of the three incentive structures. The results of this are shown graphically in Figure 8.12 and also presented in Table 8.2. As can be seen in Figure 8.12, when the incentive structure favors exploration, participants placed increasing weight on information when making decisions. They also continued to bias their exploration towards options that were likely to be highly rewarding because the information gained on trial 5 could be used to maximize rewards on trials 6-10.

Table 8.2: Model Parameters for Incentives Task

Incentives	Reward Weight (β_μ)	Info Weight (β_I)
Reward	0.094 ($sd = 0.0056$, $p < 2e^{-16}$)	-0.039 ($sd = 0.036$, $p < 0.28$)
Reward + Information	0.10 ($sd = 0.0058$, $p < 2e^{-16}$)	0.13 ($sd = 0.037$, $p < 0.0005$)
Information	0.032 ($sd = 0.0046$, $p < 2e^{-16}$)	0.35 ($sd = 0.034$, $p < 2e^{-16}$)

Conversely, when the incentive structures favored exploitation, participants placed increased weight on maximizing reward when making decisions. They even exhibited weak ambiguity aversion with a small preference for choosing a more certain option even when all options produce the same amount of reward in expectation, although this result was not statistically significant.

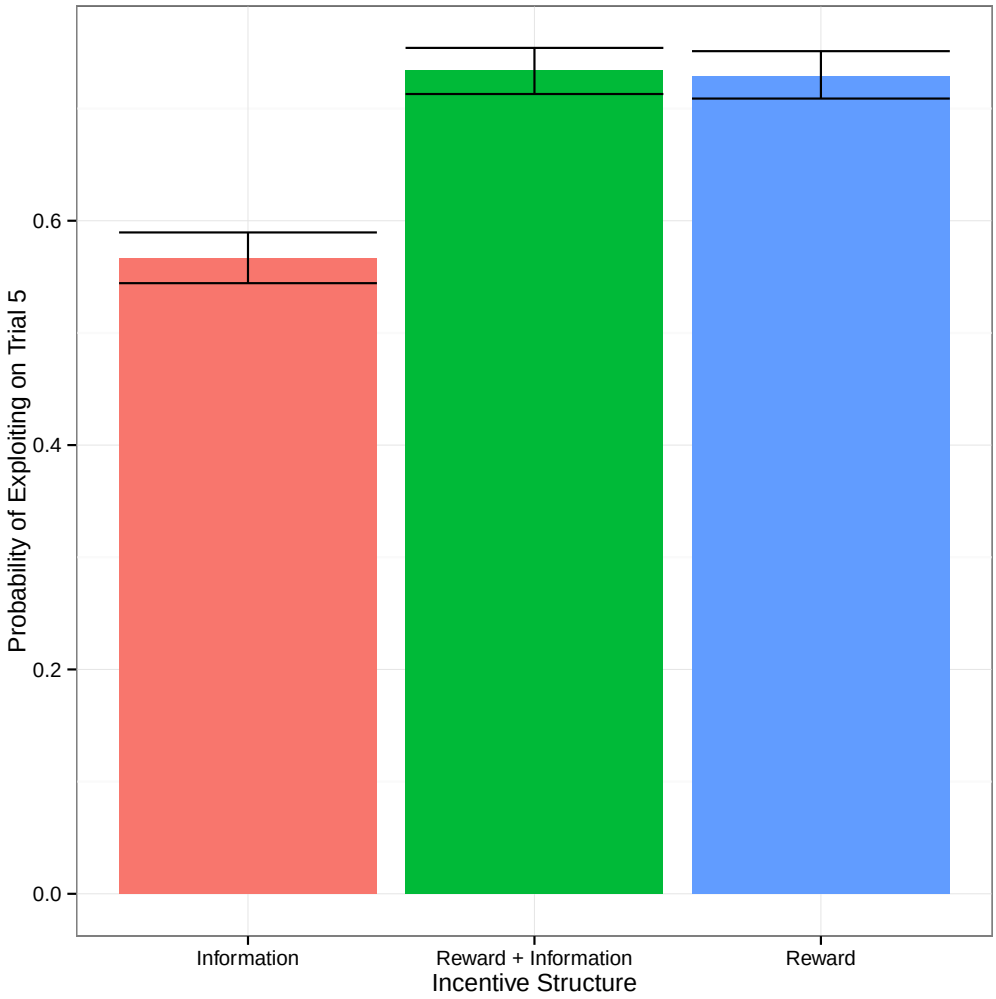


Figure 8.11: Two-Bandit Task with Changing Incentives

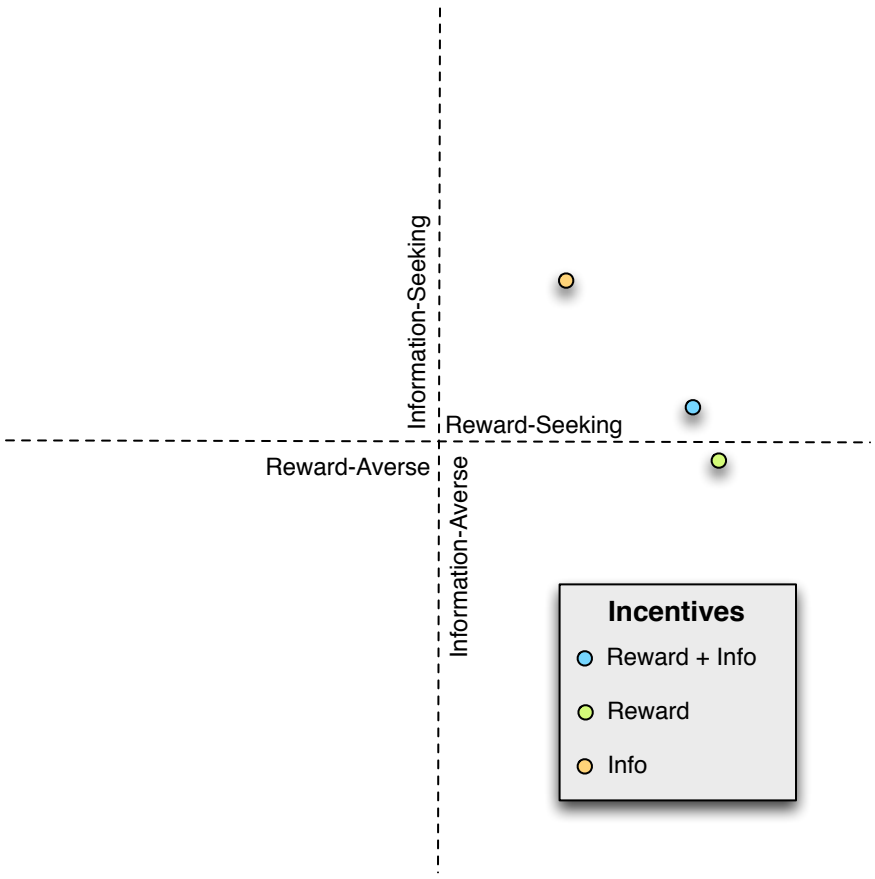


Figure 8.12: Results of the Incentive Task

The standard incentive structure for the MAB problem, in which both exploration and exploitation were advantageous, produced an intermediate type of behavior in which decisions favored highly rewarding options, but a small ambiguity bonus was also present that promoted exploration of unfamilar options.

8.4 The Design of Many-Bandit Tasks

Introduction: The simplicity of the two-bandit task makes it an ideal model system for studying the exploration-exploitation dilemma. But the simplicity of selecting between only two arms comes at a cost: the two-bandit task lacks essential features of ecologically valid environments. In many natural environments, humans must choose between a very large number of options. Coping with a large number of possible actions becomes particularly pressing when the number of actions is so large that the decision-maker cannot possibly have direct experiences with all of the actions available to them. In such cases, the decision-maker must generalize between options and make decisions based on purely hypothetical estimates of the reward rates associated with actions that have never been taken before. Because generalization is so important to decision-making in complex settings, the increase in ecological validity that comes from expanding a MAB problem to many arms comes with a cost of its own: researchers must make stronger assumptions about the inference process that participants employ to estimate the value of unknown arms.

For this reason, most of the previous work in psychology has introduced only a small number of arms: for example, Daw et al (2006) employed 4 arms and Stevyers et al. also employed 4 arms in their tasks. Nevertheless, it is clear that humans are able to explore effectively in spaces in which there are very many possible actions: in chess, for example, dozens of actions are available on each trial and yet even children are able to learn to play highly effectively (Thrun, 1995). As such, I have focused

my research on MAB problems in which (a) the number of arms is in the hundreds or thousands and (b) the arms have a statistical structure that links their expected rewards, which allows the participant to efficiently generalize from their experiences to unfamiliar arms.

Within the theoretical literature on the MAB problem, the canonical approach to generalization across arms is to consider the case of correlated arms (Klein and Rady, 2011). In the theory of correlated arms, one assumes that the reward distributions of the arms are not purely independent, but linked in some way. Typically, one assumes that the expected value for each arm is correlated with the expected value of other arms. Because the majority of theoretical research has focused on arms with correlated expected values, I have focused my efforts on this setting as well. The introduction of correlations between arms substantially complicates the inference process, so I must describe a new method for updating the decision-maker's beliefs in the MaxInfo model.

In my model and in the experiments I have used to test it, I assume that the expected values of all of the arms can be represented using an n-dimensional vector, ρ, which is drawn from a multivariate Gaussian with mean μ and covariance Σ:

$$\rho \sim \mathbf{N}(\mu, \Sigma).$$

In the tasks I will describe, I assume that the sub-components of μ are constant across arms: as such, any differences between the inferred expected values of the arms reflects only random deviations that affect only one specific draw from this multivariate Gaussian. In general, there is no arm that is reliably superior across tasks in the design I will employ. Thus each specific game that a participant will complete is defined entirely by determining the structure of the specific noisy draw that was made from this multivariate Gaussian. To simplify this process, the observations

that participants make (i.e. the rewards that are received when pulling on any given arm) are noiseless. To encourage exploration in this noiseless setting, I will force participants to select a new arm on each trial and prevent repeated draws from a single arm.

The multivariate Gaussian model described in the paragraph above defines the true, underlying generative model of the arms that I will ask experimental participants to explore. I complement this generative model by assuming that the decision-maker's beliefs are also represented using a multivariate Gaussian with distribution,

$$r_t \sim \mathbf{N}(m_t, S_t).$$

In this setting, m represents the inferred rewards associated with all of the arms and S represents the decision-maker's uncertainty about all of these rewards.

To simplify the inference process, I assume that the decision-maker knows the veridical values of μ and Σ. Based on this assumption, I then assume that the decision-maker's beliefs about r are initialized as:

$$m_0 = \mu$$

and

$$S_0 = \Sigma.$$

Note that m_t entirely defines the expected reward associated with each arm and the S_t entirely the amount of information that the decision-maker has about each arm. The MaxInfo model will therefore employ quantities derived from these two parameters in order to maximize a weighted combination of reward and information.

At time t, I use the notation u to index the arms that have not yet been observed and o to index the arms that have been observed. Under the noiseless observation

assumption made above, the decision-maker can use the values seen from the observed arms (i.e. the o arms) to update their predictions of the values at the unobserved arms (i..e the u arms) by smoothing over the values at observed locations based on the known correlational structure of the multivariate Gaussian that defines all of the arms. This reduces to the classical formula for the expectation of a multivariate Gaussian conditional on a partial observation of one draw:

$$m_{u,t} = m_{u,t-1} + \Sigma_{u,o}\Sigma_{o,o}^{-1}(m_{o,t-1} - \mu).$$

Under the noiseless observation assumption, $m_{o,t}$, never needs to be updated because observed values are known with perfect accuracy after a single observation is made for any specific arm.

Similar to the smoothing process used to update m_t, the decision-maker updates their measure of information using Σ as well. Let ϕ_t denote a "one-hot" indicator vector in which $\phi_t[i] = 1$ if arm i was observed on trial t and $\phi_t[j] = 0$ for all of the unobserved arms j such that $j \neq i$. Using this indicator vector notation for describing the position of the participant at time t, the decision-maker's uncertainty is then updated using the Sherman-Morrison rule (Sherman and Morrison, 1950):

$$S_t = S_{t-1} - \frac{S_{t-1}\phi_t\phi_t^T S_{t-1}}{\phi_t^T S_{t-1}\phi_t}.$$

Together, these updates of m and S represent the full results of the inference process in a many-armed bandit setting with arbitrary correlational structure relating the expected values of the arms. This kind of inference process is very broadly applicable and may serve as a useful approximation even in settings in which the arms are not defined by a multivariate Gaussian.

Based on the results of this updated inference process, the decision process assigns a score to each arm i using m_t and S_t as inputs. Specifically, the decision process

assumes that the expected reward for arm i is,

$$\mu_i = m_i,$$

and that the expected informational gain is,

$$I_i = \log(\det(S)) - \log(\det(S_{-i})),$$

where I use the notation S_{-i} to denote the expected covariance matrix that will arise after observing arm i. (NB: This expected covariance matrix can be computed by removing both the i^{th} row and column from S.) I use the log determinant of the matrix S as a general measure of the ambiguity that the decision-maker has: this simple log determinant rule is, in fact, equivalent to the entropic measure of ambiguity defined in Chapter 7.

8.5 The Grid Task Paradigm

Methods: In order to present many arms to a participant quickly, I employ an experimental paradigm in which participants see a rectangular grid of arms. Each location in this grid represents one arm. An example grid is shown in Figure 8.13 in which the mapping between locations in the grid and arms is shown explicitly, although experimental participants are never shown an explicit numbering system in the experimental tasks that I have studied. In my experiments, participants navigate the grid by clicking on any of the grid points with a mouse: this allows them to move arbitrarily far from their current position at any point in time.

The usefulness of this grid representation as an experimental technique is that the grid visualization allows participants to learn the correlational structure of arms more quickly because they can employ intuitions based on visuospatial reasoning to

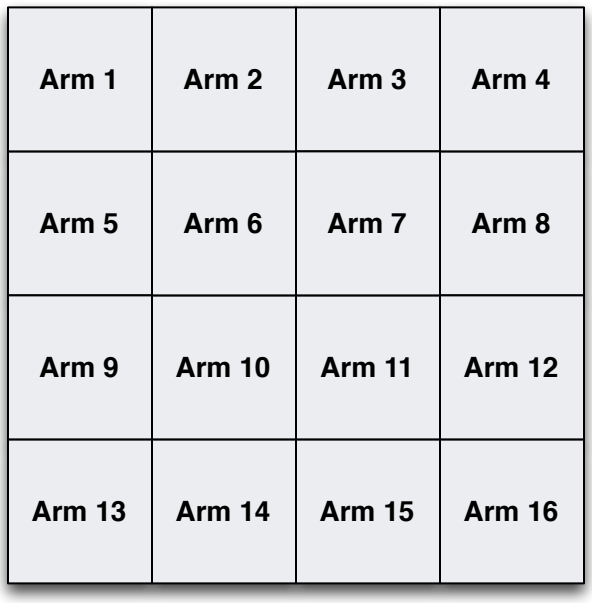


Figure 8.13: Example Grid Task

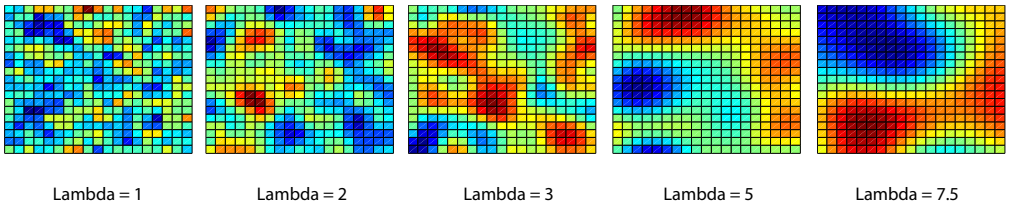


Figure 8.14: Example Grid Worlds with Increasing Covariance between Neighbors

make sense of the correlational structure that links the reward distributions of the arms. For example, in the grid shown in Figure 8.13, I could impose the requirement that the rewards from Arm 1 are more similar to the rewards from Arms 2 and 5 than to the rewards from Arms 12 and 15. This sort of distance-based reasoning comes naturally to participants, which makes it easier for the participants to learn about the structure of hundreds of arms at once (Pandey, Chakrabarti, and Agarwal, 2007). I will refer to experiments that employ spatial grids like this one as the *grid task* paradigm. In my work, I specifically impose spatial correlation structure by drawing the rewards for all arms from a multivariate Gaussian whose covariance matrix is defined as,

$$\Sigma_{i,j} = \exp(-\frac{D(i,j)}{\lambda}),$$

where $D(i,j)$ is the Euclidean distance between arms i and j in the grid I have shown. In this case, λ provides an experimentally tunable parameter that defines the length-scale of correlations: as λ grows larger, correlations become neighbors grow stronger and stronger, making the expected rewards on the grid appear smoother and smoother. An example of the increased smoothness associated with greater values of λ is shown in 8.14.

Because grid presentations of many arms make it easy for participants to reason

about large numbers of arms, the grid task is a natural extension of the two-bandit task. From an abstract perpsective, there is no reason that the two-bandit task cannot be replaced with a grid task that employs 1x2 or a 2x1 grid. Moreover, because the two-bandit task can be viewed as a special case of the grid task, all of the manipulations that I have employed in the two-bandit task could be extended naturally to the grid task.

The major difficulty with employing grid tasks is that the participant will not acquire any samples from many locations if the grid is large relative to the horizon. Because experimental participants do not have direct experiences with many locations, I must make the stronger assumptions about the generalization strategies that participants employ when estimating the utility of unobserved locations because my estimates of the values participants infer are not directly conditioned on observed values. This connects the grid task to the extensive literature on human generalization and pattern recognition abilities (Shepard et al., 1987; Tenenbaum, Griffiths, et al., 2001). In the analyses that follow, I will always assume that participants use the multivariate Gaussian inference process described earlier. As described earlier, this is equivalent to assuming that participants update m_t and S_t over time using the equations:

$$m_{u,t} = m_{u,t-1} + \Sigma_{u,o}\Sigma_{o,o}^{-1}(m_{o,t-1} - \mu)$$

$$S_t = S_{t-1} - \frac{S_{t-1}\phi_t\phi_t^T S_{t-1}}{\phi_t^T S_{t-1}\phi_t}.$$

This assumption, of course, depends upon an additional assumption that participants are aware of the mean value of the grid's arms and their correlational structure.

To ensure that participants could plausibly have such awareness, I presented participants with five example worlds that exemplified the correlational structure implied by a specific value of λ before asking participants to explore a grid. Each of these five example worlds was an independent draw from the same multivariate Gaussian that

was later used to generate the grids they the participants were asked to explore. I also described the structural properties of the grids to participants before the experiment begins by explicitly telling participants that the mean reward at any location was 120 and that the rewards ranged between 0 and 255 with a bell-shaped curve favoring rewards near 120.

One final complication in conducting a grid task like I have described is that observed choices can be excessively sparse relative to the size of the grid if the horizon is only 10 or 15 trials. In order to ensure that participants observe at least 10% of the grid, the horizon must be 40 trials. To acquire 40 trials of data for many grids, I removed the imposed choice manipulation used in all of the two-bandit tasks described earlier in this ection. Because I removed imposed choice trials form the paradigm, I will only present the results of statistical analyses that depend upon experimentally controlled variables. All of the modeling results I will describe should be be considered purely comparative because the parameter estimates of the MaxInfo model can be biased by the exclusion of a mechanism for decorrelating reward and information.

In particular, I will focus on two classes of comparative results: manipulations of the correlational structure determined by the parameter λ used to generate the values of the arms and manipulations of the task's incentive structure like those used in the Incentives task described earlier. I hypothesized that both of these manipulations would promote exploration, which is essential to demonstrating that participants employ a complex and adaptive model of their environment when attempting to maximize information.

8.6 The Effects of Manipulating Structure

Introduction: It is not possible for participants to explore all of the arms in a grid task unless the horizon is larger than the number of arms. In my tasks, the

horizon is generally an order of magnitude less than the number of arms. As such, the participant's ability to maximize reward is strongly dependent on their ability to infer the global structure of the grid from isolated examples. I therefore conducted an experiment in which I manipulated the structure of the underlying environment across grids in order to assess the ability of participants to adapt their behavior to complex correlational structure in the environment.

Methods: To assess the effects of correlational structures on behavior, I conducted a grid task experiment in which 23 Princeton undergraduate students were recruited to complete 75 brief MAB problems, which I again call games. Each of these games employed a randomly generated grid drawn from the multivariate normal with mean 120 and a distanced-based covariance matrix parameter with parameter λ. Before the start of the experimental session, the participants were shown instructions that explained that their objective was to maximize the number of points they earned over the course of the entire experiment.

Each participant completed 5 blocks of games. Within each block, the length scale parameter, λ, was held constant. In addition, at the start of a block of games, the participant was shown five example grids for that value of λ in order to inform the participant of the covariance matrix that my model of the inference process assumes they know. Across blocks, λ took on the values 1, 2, 3, 5, and 7.5. These values were counterbalanced across participants to minimize order effects. I will refer to changes in λ as the "Correlational Structure" manipulation.

Within a block, each participant completed 15 games, where each game corresponded to a single MAB problem with a horizon of 40 trials. In total, this provides 3,000 trials of data per participant during which the participant explored a 20x20 grid containing 400 arms. As such, the participant never observed the value of more than 10% of the locations available to explore.

During each game, the points earned at each location were flashed at the top of

the screen as a number and then remained on the screen as a heatmap color value in order to relieve the participant of any memory load when completing inference. After these 40 trials were completed, the participant completed a secondary task in which they were asked to estimate the value of a randomly selected location on the grid using the same heatmap that was employed during the main task. This secondary task was included in order to encourage greater exploration because participants were rewarded for the accuracy of their estimates. Specifically, participants were rewarded by computing the root mean squared error of their estimated values and subtracting this value from the maximum possible error of 255. I referred to the primary task, in which participants solved a MAB problem, as Phase I. I referred to the secondary task, in which participants estimated the value of locations in the grid, as Phase II. At the start of each game, the participant was told which phases were important for maximizing their score in that specific game.

In Condition 1, which I refer to as the "Reward" incentive structure, participants received points for Phase I only, which should encourage the least exploration. In Condition 2, which I refer as to the "Reward + Information" incentive structure, participants received points for both Phase I and Phase II. This was done by computing a 50/50 mix of points from the two phases after rescaling to make these points roughly comparable. This incentive structure should encourage an intermediate level of exploration. In Condition 3, which I refer to as the "Information" incentive structure, participants received points for Phase II only. This should encourage the most exploration. Collectively I will refer to these three conditions as the "Incentives" manipulation.

Although the "Incentives" manipulation is similar in spirit to the three incentive structures that I employed in the second two-bandit experiment described earlier in this section, the incentive structures used in this grid task were not equivalent to those used in the two-bandit task. Even in the "Reward" condition for this grid task,

Table 8.3: λ shifts information-seeking

Structure	Reward Weight (β_μ)	Info Weight (β_I)
$\lambda = 1$	0.0833	22.4
$\lambda = 2$	0.0900	7.42
$\lambda = 3$	0.0666	7.36
$\lambda = 5$	0.0924	3.18
$\lambda = 7.5$	0.0896	2.27

there was an incentive for participants to acquire information because they faced a generic MAB problem in which information is always valuable. In contrast, the "Information" condition in this grid task provides no incentive at all for traditional MAB problem behavior and should induce participants to systematically visit every location in the grid using a patterned search strategy. I therefore do not compare results from manipulations of incentives in this grid task to the results from the two-bandit task directly, but instead focus on the qualitative pattern of predictions for this manipulation insofar as this qualitative pattern mirrors the changes seen in the two-bandit task.

Results: Using maximum likelihood estimation via a Nelder-Mead computation (Nelder and Mead, 1965), I estimated the parameters of the MaxInfo model of the decision process by pooling the data from all participants and then splitting based on the λ parameter that controlled the degree of correlational structure in the world. The effects of manipulating λ are shown in Figure 8.15 and Table 8.3. Consistent with my predictions, the behavior across conditions shifted away from exploitation and towards exploration.

8.7 The Effects of Manipulating Incentives

Results: Because I manipulated incentive structures in the grid task as well as correlational structures in the grid task experiment outlined in the section above, I

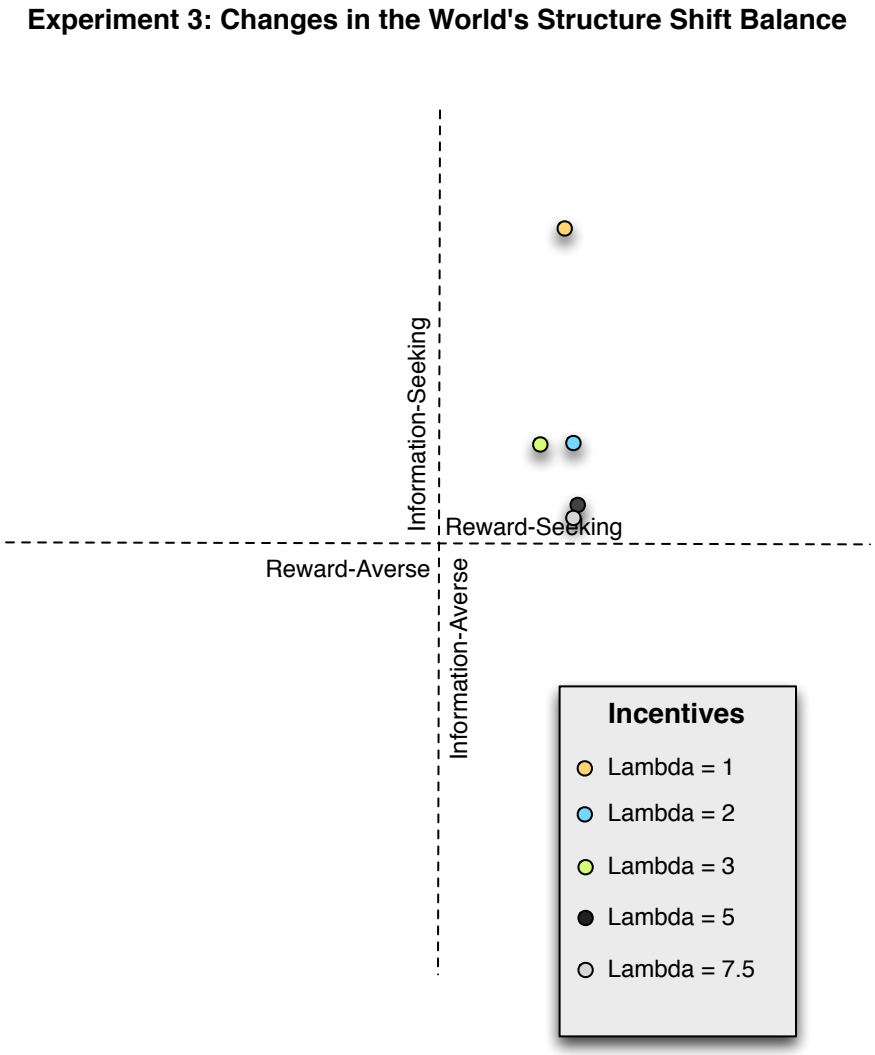


Figure 8.15: Manipulations of λ shift behavior

Table 8.4: Manipulations of incentives shift behavior

Incentives	Reward Weight (β_μ)	Info Weight (β_I)
Points	0.154	4.95
Points + Information	0.112	5.12
Information	0.0155	7.28

was also able to measure the changes in behavior that were caused by changes in these incentives after pooling over different values of λ. To assess these changes, I again fit the MaxInfo model to the pooled data from all of my participants after splitting this data set into three groups: one for each of the three incentive structure that was used. I predicted that the "Reward" incentive structure would place the least weight on information, that the "Reward + Information" incentive structure would place an intermediate level and the "Information" incentive structure would place the most weight on information. The results are shown in Figure 8.16 and Table 8.4. As can be seen by the movement of model parameters up and to the left in Table 8.4, my predictions were validated.

8.8 Interpreting the MaxInfo Model

I have argued that the structure of human behavior in settings in which an exploration-exploitation dilemma arises is best understood in terms of decision-makers' attempts to balance two competing objectives: maximizing reward and maximizing information. Previous research has argued that exploitation is primarily driven by adding noise into a purely reward-maximizing process that is corrupted by noise, but my work finds robust support across several experiments for the direct maximization of information as an additional factor promoting directed exploration that is more efficient than noise-based or random exploration and is therefore more likely to be useful in environments more complex than a two-bandit task.

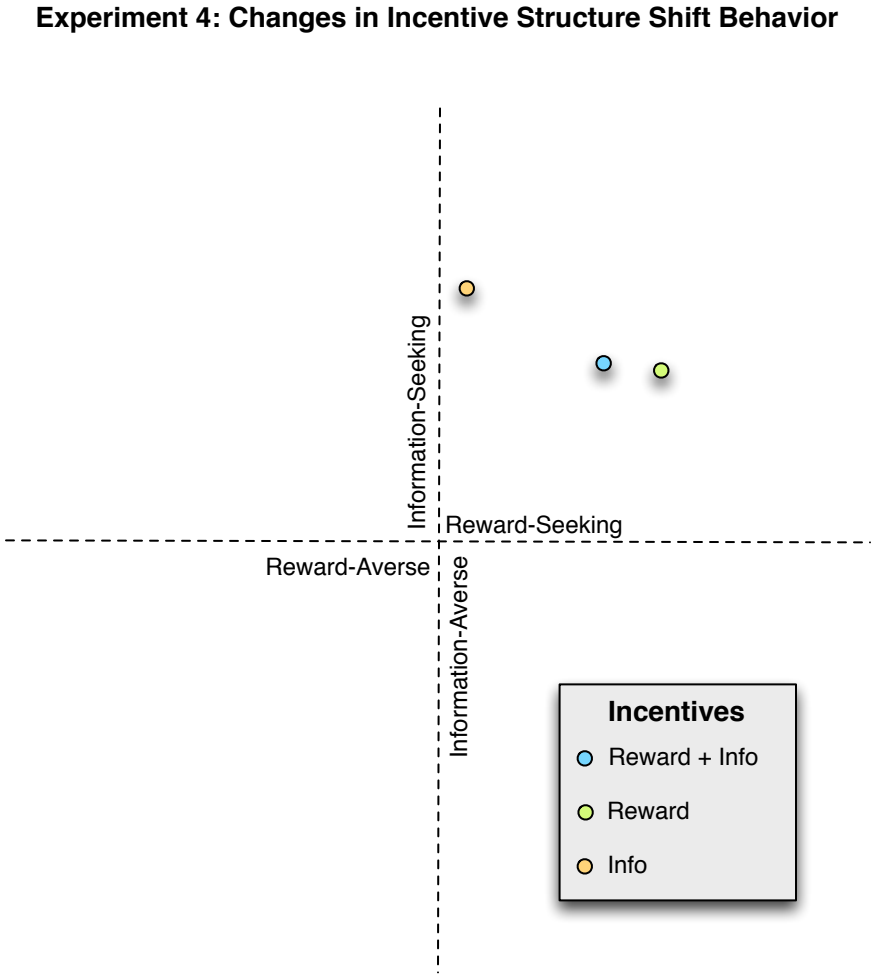


Figure 8.16: Manipulations of incentives shift behavior

This work required methodological innovations involving the introduction of imposed choice trials into the two-bandit task paradigm as well as the development of a formal model of decision-making that could directly measure shifts in behavior that correspond to changes from reward-maximization to information-maximization.

How do my results relate to the long theoretical literature on the exploration-exploitation dilemma? The first and notable result is that I find conclusive and reliable evidence across multiple experiments for information-based exploration as a mechanism for coping with the exploration-exploitation dilemma. I have not described traditional model comparisons analyses, but instead employed direct measurements and hypothesis tests that quantified the extent to which my experiments produced evidence for the role of information maximization in behavior.

In this regard, I find evidence for an exploration mechanism that is similar to state-of-the-art algorithms in machine learning, such as those that fall under the rubric of Upper Confidence Bound (or UCB) methods in which uncertain options are assigned an ambiguity bonus that drives decision-making towards maximizing information. But, unlike UCB algorithms, I also find evidence that longer horizons increase the strength of this drive towards maximizing information. Because of the dependence on horizons, my results are more consistent with the less well-known Knowledge Gradient algorithm for resolving the exploration-exploitation dilemma (Frazier et al., 2008).

Furthermore, the evidence I find for decision-makers valuing information per se is consistent with recent neuroscientific studies that suggest that the acquisition of information may be directly rewarding (Bromberg-Martin and Hikosaka, 2009) – studies that echo traditional results in economics about the value of resolving uncertainty earlier in time (Kreps and Porteus, 1978).

In addition, my work demonstrates that the extreme ambiguity aversion seen in the classical Ellsberg paradox is unique to the one-shot decision setting in which Ellsberg's example was defined. When decisions are made over longer horizons, I find

ambiguity-seeking behavior – although it is possible that humans still assign less value to information than would be assigned by an optimal decision-maker with infinite computational resources. The answer to this more subtle question strongly depends upon one's assumptions about the exact algorithms that people employ for resolving the exploration-exploitation dilemma and will require a fuller model of bounded rationality than I have employed. Producing a definitive answer to this question will therefore require further research.

Turning to the role of randomness of shaping decisions, previous work in psychology and neuroscience has found evidence for the softmax decision rule being superior to an ϵ-greedy rule (Daw et al., 2006), but this finding does not rule out the possibility that both noise-based exploration and random exploration are present in human decision-making at the same time. Future work will therefore need to target this question with focused experiments. Specifically, experiments will need to test behavior in the asymptotic regime of the softmax rule in which decisions head towards determinism: it is the behavior of these models when one option strongly dominates another option that distinguishes the softmax decision rule from the ϵ-greedy rule.

Additional questions remain about the best classification of human decision-making in terms of methods in the theoretical literature on MAB problems that I have not discussed. For example, the oldest known algorithm for solving the MAB problem is called Thompson sampling and involves drawing samples from the decision-maker's posterior distribution over the rewards that will be assigned from each of the arms (Thompson, 1933). This algorithm is surprisingly effective in practice, although a theoretical analysis was lacking until very recently (Kaufmann, Korda, and Munos, 2012). To what extent would such a sampling model of exploratory decision-making be consistent with other examples of sampling models in psychology, including the decision-by-sampling literature (Stewart, Chater, and Brown, 2006), studies of bimodal perceptions (Gershman, Vul, and Tenenbaum, 2009) and drift-diffusion models

(Ratcliff and McKoon, 2008)? My analysis does not offer answers to such questions.

Other properties of theoretical algorithms may also provide interesting qualitative summaries of human behavior. As I noted earlier, the MaxInfo model assumes (and my data demonstrate) that human decisions are influenced by the horizon over which exploration is possible. This implies that humans do not employ an anytime policy. As such, humans cannot be using any member of the UCB family of algorithms. As noted earlier, the Knowledge Gradient seems to the algorithm most similar to observed human behavior.

Many distinctions similar to the anytime policy categorization can be used to classify algorithms and situate human behavior with regard to broad taxa of strategies for resolving the MAB problem. For example, the Gittins index, the well-known standard of optimality in the MAB setting, is called an **index policy** because it is possible to assign every arm an index that is independent of the evaluations of the other arms.

Do humans employ an index policy? The success of MaxInfo model provides one tentative answer: because of the way in which I defined the value of information in the multivariate Gaussian case – namely, that the value of any arm is linked to the informational state of the other arms and cannot be computed in isolation – humans cannot be using an index policy. But future work will be needed to assess whether humans use an independent or joint assessment of the value of arms in MAB problems.

Moving beyond issues raised by the theoretical literature, future work could enrich the inference and decision processes that I have modeled by incorporating psychological phenomena that were not considered. One could imagine incorporating well-known phenomena like primacy effects (Deese and Kaufman, 1957), recency effects (Murdock Jr, 1962) and probability weighting (Kahneman and Tversky, 1979) into the inference process. In similar regard, how does my work relate to recent studies in the literature on decision-making from experience such as Hertweg and Erev's model

of underweighting of rare events (Hertwig, Barron, et al., 2004)?

In my studies, I carefully avoided considerations of risk-aversion, but future work will need to take risk preferences into account when modeling exploration-exploitation decisions. Recent work in machine learning (Sani, Lazaric, and Munos, n.d.) has begun to grapple with the issue of risk-aversion, but, to my knowledge, no relevant human studies have been published yet. Similarly, how do people cope with heavy-tailed probability distributions like those found in the stock market and other many real-world settings in which they confront imperfect information? Theoretical models for coping with these situations exist (Liu and Zhao, 2011), but empirical results on humans are still lacking.

From a different perspective, how do the dynamics of reward and information maximization differ between individuals? In many ways, the drive to exploration reflects a bias towards long-term gains. Is it impaired in clinical populations? And how does the capacity to explore relate to general capacities for planning like those studied by Simon and Newell (Newell, Simon, et al., 1972)?

In general, how does the inference process I have considered relate to broader issues in Reinforcement Learning (Dayan and Niv, 2008)? My model for the two-bandit task was essentially a model-free TD-learner, but the models I have used for the grid task are complex model-based learners that involve computations more sophisticated than the human literature on model-based learning has typically employed. To what extent are such sophisticated models of the inference process (i) a viable descriptive models of human learning and (b) biologically plausibly? The computational costs of estimating the entropy of a multivariate distribution are quite high. What approximations might humans use?

One final question concerns relating work on exploration-exploitation trade-offs in cumulative reward settings like the MAB problem to work in domains in which exploration is costly, but in which rewards occur only on a single final trial. This generic

problem, called the optimal stopping problem, lies at the heart of earlier psychological work (Rapoport and Tversky, 1970) on information-seeking and continues to be of interest to economists who study search (Gabaix, Laibson, Moloche, and Weinberg, 2006), but is less commonly studied in psychology than the MAB problem. Will a variant of the MaxInfo model work in those domains as well?

Many other questions remain about the mechanisms by which humans explore unfamiliar environments, but I believe the MaxInfo model represents the state-of-the-art in summarizing our field's current understanding of the mechanism humans employ when coping with the exploration-exploitation dilemma.

Chapter 9

Conclusion

I have discussed two extensive lines of research: one concerned with explicit trade-offs between earlier and later goods, the other concerned with implicit trade-offs between short-term and long-term reward. Stepping back from the specifics of the work I have described, what broader lessons do we, as psychologists, learn from this body work as a whole?

I believe we learn that:

- Asking humans to make decisions in unfamiliar situations like those employed in the MEL task is problematic. Such experiments are seductive because they embody in simplified form a question posed by normative theory, but they may end up measuring heuristics and properties of the psychophysics of numbers rather than the deeper dynamics of choice. The ITCH model articulates these concerns in a particularly precise form and demonstrably better describes the empirical data than the normative models that inspired the line of research that has employed MEL tasks. Although it emerges from a critical assessment of the literature on intertemporal choice, the ITCH model may nevertheless embody interesting properties of human decision-making. Future research should focus on those properties directly.

- At the same time that I have argued that the MEL paradigm is suspect, a good deal of research is still required to determine why MEL experiments have produced any useful results at all – which they have, time and time again. Why does any signal emerge from a paradigm that is ill-conceived? What does the existence of any signal at all in a problematic experimental paradigm tell us about the way we construct and validate our experiments as psychologists?

- In contrast, my work on implicit time-tradeoffs involving the value of information suggests that, when situations allow for trial-and-error learning over thousands of decisions, people are able to employ (or perhaps develop) strategies that closely resemble normative mechanisms for coping with imperfect information. Is this discrepancy yet another manifestation of the gap between description and experience that has recently emerged in studies of decision-making?

- Another interesting question that arises from studies of exploratory decision-making is the question of how valuable noise is for shaping human decision-making. This question is central to Adaptive Gain Theory (Aston-Jones and Cohen, 2005), which posits that noisy decision-making is the primary mechanism of human exploration and that it adapts in real-time to the environment. The MaxInfo extends this notion by incorporating information maximization as a driving force for decision-making, but the estimated parameters of the MaxInfo model provide clear evidence that noise continues to be present in the decision-making. Thus, a question remains: is noisy decision-making a vestigial mechanism for exploration that predates the ability to strategically maximize information? Or is it a robust alternative to model-based search that improves the fitness of decision-makers whose capacity to model the world is itself imperfect?

- Beyond the general lessons about exploratory decision-making that can be taken

from my work on the MaxInfo model, this line of work has produced several other interesting results. First, this work demonstrates that experiments that study sequential decision-making must be very carefully designed to prevent confounding. Second, this line of work demonstrates that humans generally have multiple, competing objectives when they make decisions – and that they can dynamically modulate their assessment of the relative importance of these objectives as needed based on changes in their environment. To what extent is the multi-objective optimization view of decision-making a unifying framework for the anomalies seen in decision-making, including those articulated in the ITCH model?

I conclude this thesis by noting that decision-making gains its rich complexity from observing situations in which people must make complex trade-offs involving multi-dimensional goods. The time-dated monetary goods employed in MEL and the reward/information goods that arise in the MAB problem are both exemplars of a broader problem in decision-making: how do humans cope with goods that are defined by a large array of attributes? Both the ITCH model and the MaxInfo model assert that they reason about such goods by computing a weighted mixture of all attributes. The experiments that assessed the ITCH model and the MaxInfo model also find that these weights are highly malleable to experimental manipulations that shift the focus of attention towards some attributes and away from others. How can this concept of selectively biased attention be used to define a broader theory of decision-making? The neural evidence for biased competition is quite powerful (Miller and Cohen, 2001). How can the theory of decisions be extended to embrace this perspective?

References

Ainslie, G. (1975). Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological bulletin*, *82*(4), 463.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383–405.

Andreoni, J., & Sprenger, C. (2010). Risk preferences are not time preferences. *Levine's Working Paper Archive No*, *926159300*.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci. 28*, 403–450.

Auer, P. (2003). Using confidence bounds for exploitation-exploration trade-offs. *The Journal of Machine Learning Research*, *3*, 397–422.

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (2002). The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, *32*(1), 48–77.

Ayer, A. J. (1966). *Logical positivism*. Free Press.

Baker, F., Johnson, M. W., & Bickel, W. K. (2003). Delay discounting in current and never-before cigarette smokers: similarities and differences across commodity, sign, and magnitude. *Journal of Abnormal Psychology*, *112*(3), 382.

Bar-Hillel, M., & Margalit, A. (1988). How vicious are cycles of intransitive choice? *Theory and Decision*, *24*(2), 119–145.

Becker, G. M., DeGroot, M. H., & Marschak, J. (2006). Measuring utility by a single-response sequential method. *Behavioral science*, *9*(3), 226–232.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience*, *10*(9), 1214–1221.

Benhabib, J., Bisin, A., & Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, *69*(2), 205–223.

Benzion, U., Rapoport, A., & Yagil, J. (1989). Discount rates inferred from decisions: an experimental study. *Management science*, *35*(3), 270–284.

Berns, G. S., Laibson, D. I., & Loewenstein, G. (2007). Intertemporal choice-toward an integrative framework. *Trends in cognitive sciences*, *11*(11), 482–488.

Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, *63*(1), 119.

Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the bis/bas scales. *Journal of personality and social psychology*, *67*(2), 319.

Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P., & Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty*, *37*(2), 237–269.

Chapman, G. B., & Elstein, A. S. (1995). Valuing the future temporal discounting of health and money. *Medical Decision Making*, *15*(4), 373–386.

Cubitt, R. P., & Read, D. (2007). Can intertemporal choice experiments elicit time preferences for consumption? *Experimental Economics*, *10*(4), 369–389.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology*, *18*(2), 185–196.

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of experimental psychology*, *54*(3), 180.

Edgeworth, F. Y. (1881). *Mathematical psychics: an essay on the application of mathematics to the moral sciences*. CK Paul.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 643–669.

Epper, T., Fehr-Duda, H., & Bruhin, A. (2011). Viewing the future through a warped lens: why uncertainty generates hyperbolic discounting. *Journal of Risk and Uncertainty*, 1–35.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*, 309–368.

Frazier, P. I., Powell, W. B., & Dayanik, S. (2008). A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, *47*(5), 2410–2439.

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

Fudenberg, D., & Levine, D. K. (2006). A dual-self model of impulse control. *The American Economic Review*, 1449–1476.

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly information acquisition: experimental analysis of a boundedly rational model. *The American Economic Review*, 1043–1068.

Gershman, S. J., Vul, E., & Tenenbaum, J. (2009). Perceptual multistability as markov chain monte carlo inference. *Advances in Neural Information Processing Systems*, *22*, 611–619.

Gilboa, I. (2010). *Rational choice*. MIT Press.

Glimcher, P. W. (2004). *Decisions, uncertainty, and the brain: the science of neuroeconomics*. MIT press.

Glimcher, P. W., Kable, J., & Louie, K. (2007). Neuroeconomic studies of impulsivity: now or just as soon as possible? *The American economic review*, *97*(2), 142–147.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, 623–638.

Gul, F., & Pesendorfer, W. (2003). Temptation and self-control. *Econometrica*, *69*(6), 1403–1435.

Gul, F., & Pesendorfer, W. (2006). Random expected utility. *Econometrica*, *74*(1), 121–146.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmpfc valuation system. *Science*, *324*(5927), 646–648.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall/CRC.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539.

Hertwig, R., Erev, I. (2009). The description-experience gap in risky choice. *Trends in cognitive sciences*, *13*(12), 517–523.

Hines, J. R., & Thaler, R. H. (1995). Anomalies: the flypaper effect. *The Journal of Economic Perspectives*, *9*(4), 217–226.

Houthakker, H. S. (1950). Revealed preference and the utility function. *Economica*, *17*(66), 159–174.

Hsee, C. K. (1996). The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247–257.

Ipeirotis, P. (2010). Demographics of mechanical turk.

Johnson, E. J., Payne, J. W., Bettman, J. R., & Schkade, D. A. (1989). *Monitoring information processing and decisions: the mouselab system*.

Jones, D. M., & Gittins, J. C. (1972). *A dynamic allocation index for the sequential design of experiments*. University of Cambridge, Department of Engineering.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: a survey. *arXiv preprint cs/9605103*.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263–291.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to bentham? explorations of experienced utility. *The Quarterly Journal of Economics*, *112*(2), 375–406.

Kaufmann, E., Korda, N., & Munos, R. (2012). Thompson sampling: an optimal finite time analysis. *arXiv preprint arXiv:1205.4217*.

Klein, N., & Rady, S. (2011). Negatively correlated bandits. *The Review of Economic Studies*, *78*(2), 693–732.

Knight, F. H. (2006). *Risk, uncertainty and profit*. Dover publications.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International joint conference on artificial intelligence* (Vol. 14, pp. 1137–1145). Lawrence Erlbaum Associates Ltd.

Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society*, 287–309.

Krebs, J. R., Kacelnik, A., & Taylor, P. (1978). Test of optimal sampling by foraging great tits. *Nature*, *275*(5675), 27–31.

Kreps, D. M. (1988). *Notes on the theory of choice*. Westview press Boulder.

Kreps, D. M., & Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: journal of the Econometric Society*, 185–200.

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

Lewin, K. (1935). A dynamic theory of personality.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of experimental psychology*, *89*(1), 46–55.

Liu, K., & Zhao, Q. (2011). Multi-armed bandit problems with heavy-tailed reward distributions. In *Communication, control, and computing (allerton), 2011 49th annual allerton conference on* (pp. 485–492). IEEE.

Loewenstein, G. (1994). The psychology of curiosity: a review and reinterpretation. *Psychological bulletin*, *116*(1), 75.

Loewenstein, G. (2003). *Time and decision: economic and psychological perspectives on intertemporal choice*. Russell Sage Foundation Publications.

Loewenstein, G. F. (1988). Frames of mind in intertemporal choice. *Management Science*, *34*(2), 200–214.

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: evidence and an interpretation. *The Quarterly Journal of Economics*, *107*(2), 573–597.

Loewenstein, G., & Thaler, R. H. (1989). Anomalies: intertemporal choice. *The journal of economic perspectives*, *3*(4), 181–193.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*(3), 215–233.

Luce, R. D., & Raiffa, H. (1989). *Games and decisions: introduction and critical survey*. Dover Publications.

Marler, R. T., & Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, *26*(6), 369–395.

Mason, W., & Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, *44*(1), 1–23.

Mazur, J. E. (1989). Theories of probabilistic reinforcement. *Journal of the Experimental Analysis of Behavior*, *51*(1), 87.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. Chapman & Hall/CRC.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*(1), 167–202.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1992). Delay of gratification. *Choice over time*, 147.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology; Journal of Experimental Psychology*, *64*(5), 482.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308–313.

Newell, A., Simon, H. A. (1972). *Human problem solving*. Prentice-Hall Englewood Cliffs, NJ.

O'Donoghue, T., & Rabin, M. (2001). Choice and procrastination. *The Quarterly Journal of Economics*, *116*(1), 121–160.

Pandey, S., Chakrabarti, D., & Agarwal, D. (2007). Multi-armed bandit problems with dependent arms. In *Proceedings of the 24th international conference on machine learning* (pp. 721–728). ACM.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica: Journal of the Econometric Society*, 122–136.

Pratt, S. C., & Sumpter, D. J. (2006). A tunable algorithm for collective decision-making. *Proceedings of the National Academy of Sciences*, *103*(43), 15906–15910.

Prelec, D. (2004). Decreasing impatience: a criterion for non-stationary time preference and hyperbolic discounting. *The Scandinavian Journal of Economics*, *106*(3), 511–532.

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the experimental analysis of behavior*, *55*(2), 233.

Raiffa, H. (1961). Risk, ambiguity, and the savage axioms: comment. *The Quarterly Journal of Economics*, *75*(4), 690–694.

Raiffa, H. (1968). Decision analysis: introductory lectures on choices under uncertainty.

Rapoport, A., & Tversky, A. (1970). Choice behavior in an optional stopping task. *Organizational Behavior and Human Performance*, *5*(2), 105–120.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of risk and uncertainty*, *23*(1), 5–32.

Read, D., Frederick, S., & Scholten, M. (2011). Outcome framing in intertemporal choice: the drift model. *Available at SSRN 1933099*.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*(5), 527–535.

Rubin, D. B. (1980). Randomization analysis of experimental data: the fisher randomization test comment. *Journal of the American Statistical Association*, *75*(371), 591–593.

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies*, *4*(2), 155–161.

Sani, A., Lazaric, A., & Munos, R. (n.d.). Risk–aversion in multi–armed bandits.

Savage, L. J. (1972). *The foundation of statistics*. Dover publications.

Shamosh, N. A., DeYoung, C. G., Green, A. E., Reis, D. L., Johnson, M. R., Conway, A. R., ... Gray, J. R. (2008). Individual differences in delay discounting relation to intelligence, working memory, and anterior prefrontal cortex. *Psychological Science*, *19*(9), 904–911.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Sherman, J., & Morrison, W. J. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, *21*(1), 124–127.

Skinner, B. F. (1963). Behaviorism at fifty. *Science*.

Stern, N. (2007). The economics of climate change: the stern report. *Cambridge, UK*.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, *53*(1), 1–26.

Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American economic review*, 393–410.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, *23*(3), 165–180.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge Univ Press.

Tabuchi, T., & Thisse, J.-F. (2002). Taste heterogeneity, labor mobility and economic geography. *Journal of Development Economics*, *69*(1), 155–177.

Tanaka, T., Camerer, C., & Nguyen, Q. (2006). Risk and time preferences: experimental and household survey data from vietnam. *Available at SSRN 877229*.

Tenenbaum, J. B., Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, *24*(4), 629–640.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, *8*(3), 201–207.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 285–294.

Thrun, S. (1995). Learning to play the game of chess. *Advances in Neural Information Processing Systems*, *7*.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological review*, *79*(4), 281–299.

Tversky, A., Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management science*, *39*(10), 1179–1189.

Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior (commemorative edition)*. Princeton university press.

Watkinson, S. C., Boddy, L., Burton, K., Darrah, P., Eastwood, D., Fricker, M. D., & Tlalka, M. (2005). New approaches to investigating the function of mycelial networks. *Mycologist*, *19*(1), 11–17.

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice a query-theory account. *Psychological Science*, *18*(6), 516–523.

Whittle, P. (1980). Multi-armed bandits and the gittins index. *Journal of the Royal Statistical Society. Series B (Methodological)*, 143–149.

Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: l-bfgs-b: fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, *23*(4), 550–560.