Essays in Revealed Preference Theory and Behavioral Economics

Thesis by

Taisuke Imai

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

Caltech

California Institute of Technology
Pasadena, California

2016
(Defended December 11, 2015)

To my family.

Acknowledgements

First and foremost, I would like to express my sincere thanks to my advisor, Colin Camerer, for his support, patience, and encouragement. I first met him in the hot summer of 2007 at the University of Tokyo. He gave a talk on then-emerging field of neuroeconomics. The day after his seminar, I had an opportunity to show him around several sightseeing places in Tokyo. During this tour, he enthusiastically talked about the history, challenges, and future directions of behavioral, experimental, and neuro-economics. I was astonished by his encyclopedic knowledge and was deeply inspired by his passion for pioneering new fields of researches. That was the moment I decided to come to Caltech to conduct researches in those fields. My research interest drifted during the course, but he always supported me and gave the best guidance at every stage of my doctoral work.

I benefited greatly from Federico Echenique and Kota Saito. Keeping up with their speed of discussion was sometimes challenging to me, but at the same time, working with them has always been enjoyable. Conversations with Ming Hsu, Pietro Ortoleva, Antonio Rangel, and Matthew Shum have also been immensely helpful. I would also like to thank my friends who gave me constructive feedbacks and thoughtful questions: Ryo Adachi, Rahul Bhui, Ben Bushong, Kyle Carlson, Matthew Chao, John Clithero, Cary Frydman, Keise Izuma, Andrea Kanady Bui, Yutaka Kayaba, Jackie Kimble, Ian Krajbich, Gidi Nave, Euncheol Shin, Alec Smith, Shinsuke Suzuki, Federico Tadei, and Romann Weber. Laurel Auchampaugh and Barbara Estrada have provided me with amazing administrative support.

I would also like to acknowledge the Nakajima Foundation for their financial support throughout my Ph.D. research.

Last but not least, I thank my wife Noriko who always believed in me. Without her love and support, I could not have achieved this long academic journey.

Abstract

Time, risk, and attention are all integral to economic decision making. The aim of this work is to understand those key components of decision making using a variety of approaches: providing axiomatic characterizations to investigate time discounting, generating measures of visual attention to infer consumers' intentions, and examining data from unique field settings.

Chapter 2, co-authored with Federico Echenique and Kota Saito, presents the first revealed-preference characterizations of exponentially-discounted utility model and its generalizations. My characterizations provide non-parametric revealed-preference tests. I apply the tests to data from a recent experiment, and find that the axiomatization delivers new insights on a dataset that had been analyzed by traditional parametric methods.

Chapter 3, co-authored with Min Jeong Kang and Colin Camerer, investigates whether "pre-choice" measures of visual attention improve in prediction of consumers' purchase intentions. We measure participants' visual attention using eye-tracking or mousetracking while they make hypothetical as well as real purchase decisions. I find that different patterns of visual attention are associated with hypothetical and real decisions. I then demonstrate that including information on visual attention improves prediction of purchase decisions when attention is measured with mousetracking.

Chapter 4 investigates individuals' attitudes towards risk in a high-stakes environment using data from a TV game show, *Jeopardy!*. I first quantify players' subjective beliefs about answering questions correctly. Using those beliefs in estimation, I find that the representative player is risk averse. I then find that trailing players tend to wager more than "folk" strategies that are known among the community of contestants and fans, and this tendency is related to their confidence. I also find gender differences: male players take more risk than female players, and even more so when they are competing against two other male players.

Chapter 5, co-authored with Colin Camerer, investigates the dynamics of the favorite-longshot bias (FLB) using data on horse race betting from an online ex-

change that allows bettors to trade "in-play." I find that probabilistic forecasts implied by market prices before start of the races are well-calibrated, but the degree of FLB increases significantly as the events approach toward the end.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation comprises four essays that address questions from several areas of microeconomics and behavioral economics. Time preferences, risk preferences, and attention are all integral to our economic decision making. The aim of this work is to understand those key components of decision making using a variety of approaches: providing axiomatic characterizations to analyze intertemporal-choice data, generating measures of visual attention to infer consumers' cognitive processes, and collecting and analyzing data from unique field settings.

Chapter 2, co-authored with Federico Echenique and Kota Saito, presents the first revealed-preference characterizations of the most common models of intertemporal choice: the model of exponentially discounted concave utility (EDU), and its generalizations including quasi-hyperbolic discounted utility (QHD) and time separable utility (TSU). This is the first axiomatization of these models taking consumption data as primitives. My characterizations provide non-parametric revealed-preference tests. I apply the tests to data from a recent experiment by Andreoni and Sprenger (2012a), and find that the axiomatization delivers new insights and perspectives on a dataset that had been analyzed by traditional parametric methods. I find that (i) 30% of the subjects' behavior is consistent with EDU, (ii) all subjects rationalized as QHD are also rationalized as EDU, and (iii) 52% of the subjects are rationalized by TSU. Those numbers may appear small, but I demonstrate that violations of EDU, QHD, and TSU are not due to small mistakes, using two measures of "distance" from rationality.

Chapter 3, co-authored with Min Jeong Kang and Colin Camerer, studies a well-known "hypothetical bias," in which consumers tend to overstate their intentions to purchase products compared to actual rates of purchases, using a novel dataset on consumers' visual attention. I investigate whether "pre-choice" measures of visual attention improve in prediction of actual purchase intentions.

In laboratory choice experiments, we measure participants' visual attention using eyetracking or mousetracking while they make hypothetical (whether they *would* buy a product at an offered price) as well as real purchase decisions (whether they *do* actually buy). I find different patterns of visual attention associated with hypothetical and real decisions. First, participants spend longer looking both at price and product image prior to make a real "Buy" decision than making a real "Don't buy" decision. Second, I find that during hypothetical choice, the more participants look at prices, and the longer they take to transition from looking to making a choice, the more likely they are to switch a hypothetical "Buy" to a real "Don't buy," when they are asked to make real decisions on the same product-price pair later in the experiment. I then demonstrate that including information regarding visual attention as well as product's price improves on using only price in predicting purchase decisions. This improvement is evident, although small in magnitude, using mousetracking data, but is not evident using eyetracking data.

Chapter 4 investigates individuals' attitudes towards risk in a high-stakes environment using data from a TV game show, *Jeopardy!*, as a natural experiment. Exploiting the rich nature of our dataset, I first quantify each player's subjective belief about answering correctly in the final round of the game. I then estimate the representative leading player's risk preferences in a subset of games in which no strategic thinking kicks in. Improved method enables me to observe evidence for risk aversion, unlike Metrick (1995), who finds that the representative player is risk neutral. I then compare players' observed wagering decisions with those suggested by the "folk" strategies known among the community of contestants and fans. In our dataset, trailing players tend to wager more than the folk strategies suggest, and this tendency is related to subjective beliefs. I also find gender differences in which male players appear to take more risk than female players, and even more so when they are competing against two other male players.

Chapter 5, co-authored with Colin Camerer, reports a new empirical evidence on the dynamics of the favorite-longshot bias (FLB). The bias is one of the most well-documented regularities in betting markets. Using data on horse race betting from an online exchange that allows bettors to trade "in-play," I examine the dynamics of FLB over time. I find a stark difference in the degree of FLB before and during the races. Probabilistic forecasts implied by market prices before the start are well-calibrated, which is in contrast with a wide evidence of FLB in parimutuel betting markets. However, the degree of FLB gets significantly larger, especially late in the races.

Chapter 2

Testable Implications of Models of Intertemporal Choice: Exponential Discounting and Its Generalizations

2.1 Introduction

Exponentially discounted utility is the standard model of intertemporal choice in economics. It is a ubiquitous model, used in all areas of economics. Our paper is the first revealed preference investigation of exponential discounting: we give a necessary and sufficient "revealed preference axiom" that a dataset must satisfy in order to be consistent with exponential discounting. The revealed preference axiom sheds light on the behavioral assumptions underlying the standard model of discounting. It also yields a nonparametric test of the theory, applicable in different empirical investigations of exponential discounting.

Consider an agent who chooses among intertemporal consumptions of a single good. One general theory is that the agent has a utility function $U(x_0, \ldots, x_T)$ for the consumption of x_t on each date t. The Generalized Axiom of Revealed Preference (GARP) tells us whether the choices are consistent with some general utility function U.

The empirical content of general utility maximization is well understood, but it is too broad (and GARP is too weak) to capture exponential discounting. The *exponentially discounted utility* (EDU) model assumes a specific form of U, namely

$$U(x_0, \ldots, x_T) = \sum_{t=0}^{T} \delta^t u(x_t).$$

In this study, we focus on concave EDU, in which u is a concave function. Concavity of u is widely used to capture a motive for consumption smoothing over time. The empirical content of concave EDU maximization is different from that of general utility maximization, and not well understood in the literature.

The first and most important question addressed in this study is: *What is the version of GARP that allows us to decide whether data are consistent with concave EDU?* The revealed preference axiom that characterizes concave EDU is obviously going to be stronger than GARP. Despite the ubiquity of EDU in economics, the literature on revealed preference has not (until now) provided an answer. Our main result is that a certain revealed preference axiom, termed the "Strong Axiom of Revealed Exponentially Discounted Utility" (SAR-EDU), describes the choice data that are consistent with concave EDU preferences.

SAR-EDU is a weak imposition on the data, in the sense that it constrains prices and quantities *in those situations in which unobservables do not matter.* The constraint on prices and quantities is simply that they are inversely related, or that "demand slopes down." Essentially, SAR-EDU requires one to consider situations in which unobservables "cancel out," and to check that prices and quantities are inversely related. This inverse relation is a basic implication of concave utility (that is, of the consumption smoothing motive).

In the paper, we study the empirical content of more general models of time discounting as well, including the quasi-hyperbolic discounting model (QHD; Laibson, 1997; Phelps and Pollak, 1968): $U(x_0, \ldots, x_T) = u(x_0) + \beta \sum_{t=1}^{T} \delta^t u(x_t)$, general time discounting (GTD): $U(x_0, \ldots, x_T) = \sum_{t=0}^{T} D(t) u(x_t)$, and time separable utility (TSU): $U(x_0, \ldots, x_T) = \sum_{t=0}^{T} u_t(x_t)$, where u and u_t are concave. In the following, we do not explicitly use the concave modifier when there is no risk of confusion. For example, we say EDU to mean concave EDU.

The contribution of this study is to characterize the empirical content of EDU and its generalizations. We provide the first revealed-preference axioms (axioms like GARP but stronger) characterizing EDU, QHD, GTD, and TSU. Our axioms shed new insights into the behavioral assumptions behind each of these models, and also constitute nonparametric tests. There are, of course, other axiomatizations of these models but they start from different primitives. The well known axiomatization of EDU by Koopmans (1960), for example, starts from complete preferences over infinite consumption streams.

To illustrate the usefulness of our results for empirical work, we carry out an application to data from a recent experiment conducted by Andreoni and Sprenger (2012a) (hereafter AS). AS propose the Convex Time Budget (CTB) design, in which subjects are asked to choose from an intertemporal budget set. [1] They find moderate support for the theory that agents are EDU maximizers.

The application of our methods to AS's data is, we believe, fruitful. We uncover features of individual subjects' behavior that are masked by traditional parametric econometric techniques. Our tests give a seemingly different conclusion from that obtained by AS. At first glance, we find scant evidence for EDU (or indeed QHD) whereas AS are cautiously supportive of EDU. Section 2.5 has more details, and reveals that the methodology of AS and our methodology are more concordant than what initially emerges.

It should be said that our methods rest on nonparametric revealed preference tests. As such, the tests are independent of functional form assumptions. The tests are also simple, and tightly connected to economic theory. The methodology used currently by experimentalists (such as AS) rests instead on parametrically estimating a given utility function by statistical methods. Our setup fits the experimental design of AS, and other CTB experiments, very well, but our results are also applicable more broadly, including to non-experimental field data.

Related literature. There are different behavioral axiomatizations of EDU in the literature, starting with Koopmans (1960), and followed by Fishburn and Rubinstein (1982) and Fishburn and Edwards (1997). All of them take preferences as primitive, or in some cases they take utility over consumption streams as the primitive. The idea is that the relevant behavior consists of all pairwise comparisons of consumption streams. From an empirical perspective, this assumes a infinite "dataset" of pairwise comparisons. Indeed, the stationarity axiom introduced by Koopmans (1960), and used by many other authors, requires infinite time. Our axiomatization of EDU is the first in an environment where agents choose from budget sets.

Other axiomatizations of EDU impose stationarity in different environments. In Fishburn and Rubinstein (1982), preferences are defined on one-time consumptions in continuous time. In Fishburn and Edwards (1997), preferences are defined on infinite consumption streams that differ in at most finitely many periods. The recent work of Dziewulski (2015) gives a characterization for binary comparisons

[1] Several recent experimental studies use the CTB design, both in the laboratory and in the field setting, including Andreoni et al. (2015), Ashton (2014), Augenblick et al. (2015), Balakrishnan et al. (2015), Barcellos and Carvalho (2014), Brocas et al. (2015), Carvalho et al. (2013), Carvalho et al. (forthcoming), Giné et al. (2013), Janssens et al. (2013), Kuhn et al. (2014), Liu et al. (2014), Lührmann et al. (2014), Sawada and Kuroishi (2015), and Shaw et al. (2014). Our methods are largely applicable to data from these experiments.

of one-time consumptions, a similar setup to Fishburn and Rubinstein (1982), but assuming finitely many data.

In continuous time setup, Weibull (1985) gives a general characterization of EDU, also taking preferences as primitives. He characterizes the general time discounting models and the monotone time discounting models as well. A more recent paper by Kopylov (2010) also provides a simple axiomatization of EDU model in a continuous time setup.

The QHD model is first proposed by Phelps and Pollak (1968), who did not propose an axiomatization. There are several more recent studies that present a behavioral characterization of QHD, but all take preferences and infinite time horizons as their primitives, and therefore differ from our results. See Hayashi (2003), Montiel Olea and Strzalecki (2014), and Galperti and Strulovici (2014) for axiomatizations.

Time separable utility (TSU) model is the most general model we axiomatize. In our application of our test to AS's data, however, we found that significant number of subjects are not TSU rational. This would suggest the importance of a non-time separable model. Gilboa (1989) has provided an elegant axiomatization of a non-time separable utility model. In the paper, by using Anscombe and Aumann's (1963) framework and studying preferences over finite sequences of lotteries, Gilboa (1989) axiomatizes a utility function that can capture a preference for (or an aversion to) variation of utility levels across periods.

In terms of data from (field) consumption surveys, Browning (1989) provides a revealed-preference axiom for EDU with no discounting (i.e., $\delta = 1$). Other papers on survey data do not provide an axiomatic characterization; they, instead, obtain Afriat inequalities for several models. Crawford (2010) investigates intertemporal consumption and discusses a particular violation of TSU, namely habit formation. Crawford (2010) presents Afriat inequalities for the model of habit formation, and uses Spanish consumption data to carry out the test (see also Crawford and Polisson, 2014). Adams et al. (2014) work with the Spanish dataset and test EDU within a model of collective decision making at the household level.

It is important to emphasize that the papers on survey data allow for the existence of many goods in each period, but they do not allow for more than one (intertemporal) purchase for each agent. This assumption makes sense because in consumption surveys one typically has a single observation per household. We have instead assumed that there is only one good (money) in each period, but we allow for more than one intertemporal purchase per agent. Allowing for multiple purchases is crucial in order to apply our tests to experimental data. This is because in experiments a subject is usually required to make many decisions and one choice is chosen randomly for the payment to the subject.

2.2 Exponentially Discounted Utility

Notational conventions. For vectors $x, y \in \mathbf{R}^n$, $x \leq y$ means that $x_i \leq y_i$ for all $i = 1, \ldots, n$, $x < y$ means that $x \leq y$ and $x \neq y$, and $x \ll y$ means that $x_i < y_i$ for all $i = 1, \ldots, n$. The set of all $x \in \mathbf{R}^n$ with $0 \leq x$ is denoted by $\mathbf{R}^n_+$ and the set of all $x \in \mathbf{R}^n$ with $0 \ll x$ is denoted by $\mathbf{R}^n_{++}$.

Let T be a strictly positive integer; T will be the (finite) duration of time, or *time horizon*. We abuse notation and use T to denote the set $\{0, 1, \ldots, T\}$. A sequence $(x_0, \ldots, x_T) = (x_t)_{t \in T} \in \mathbf{R}^T_+$ will be called a *consumption stream*. There is a single good in each period; the good can be thought of as money. Note that the cardinality of the set T is $T + 1$, but this never leads to confusion.

Remark 1. We can assume a more general time setup, $\{0, \tau_1, \ldots, \tau_T\}$, where $\tau_i < \tau_{i+1}$ for all $i < T - 1$. Even with this general time setup, our results hold without changes. The only requirement on the set of time periods is that it contains 0. Such flexibility in how one specifies time is necessary in the application of our results to experimental data of Andreoni and Sprenger (2012a). See Section 2.5.1 for details.

The model. The objects of choice in our model are consumption streams. We assume that an agent has a budget $I > 0$, faces prices $p \in \mathbf{R}^T_{++}$, and chooses an affordable consumption stream $(x_t)_{t \in T} \in \mathbf{R}^T_+$. Prices can be thought of as interest rates.

An exponentially discounted utility (EDU) is specified by a discount factor $\delta \in (0, 1]$ and a utility function over money $u : \mathbf{R}_+ \to \mathbf{R}$. An EDU maximizing agent solves the problem

$$\max_{x \in B(p,I)} \sum_{t \in T} \delta^t u(x_t) \tag{2.1}$$

when faced with prices $p \in \mathbf{R}^T_{++}$ and budget $I > 0$. The set $B(p, I) = \{y \in \mathbf{R}^T_+ : p \cdot y \leq I\}$ is the *budget set* defined by p and I.

The meaning of EDU as an assumption about an agent is that the agent's observed behavior is *as if* it were generated by the maximization of an EDU. To formalize this idea, we need to state what can be observed.

Definition 1. A *dataset* is a finite collection of pairs $(x, p) \in \mathbf{R}^T_+ \times \mathbf{R}^T_{++}$.

A dataset is our notion of observable behavior. The interpretation of a dataset $(x^k, p^k)_{k=1}^K$ is that it describes K observations of a consumption stream $x^k = (x_t^k)_{t\in T}$ at some given vector of prices $p^k = (p_t^k)_{t\in T}$, and budget $p^k \cdot x^k = \sum_{t\in T} p_t^k x_t^k$. We sometimes use K to denote the set $\{1, \dots, K\}$.

Let us clarify the meaning of a dataset by considering two examples. If we have field consumption data, collected through a consumption survey, then K is 1. There is one dataset for each agent, or household. This is the setup of Browning (1989), for example. On the other hand, if, in an experiment, one subject is asked to make a choice from 45 different budget sets, as in Andreoni and Sprenger (2012a), then K is 45. The experimenter would typically implement the choice from one budget set selected at random. It is important to note that our model allows, but does not require, that $K > 1$. Even if $K = 1$, our axiom for EDU is not satisfied trivially, and has testable implications.

Definition 2. A dataset $(x^k, p^k)_{k=1}^K$ is *exponential discounted utility rational (EDU rational)* if there is $\delta \in (0, 1]$ and a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$ such that, for all k,

$$y \in B(p^k, p^k \cdot x^k) \Longrightarrow \sum_{t\in T} \delta^t u(y_t) \leq \sum_{t\in T} \delta^t u(x_t^k).$$

As mentioned in the introduction, we restrict attention to concave utility. Our results will be silent about the non-concave case. So we are focusing on agents who seek to smooth out their consumption over time.

2.3 A Characterization of EDU Rational Data

EDU rational data will be characterized by a single "revealed preference axiom." We shall introduce the axiom by deriving the implications of EDU in specific instances. Here we assume, for ease of exposition, that u is differentiable, but our results do not depend on differentiability, and the statement of the theorem will not require the differentiability of u.

The first-order condition for maximization of EDU is for each $k \in K$ and $t \in T$,

$$\delta^t u'(x_t^k) = \lambda^k p_t^k. \tag{2.2}$$

The first-order conditions involve three unobservables: discount factor δ, marginal utilities $u'(x_t^k)$, and Lagrange multipliers λ^k. Quantities x_t^k and prices p_t^k are observable. Our approach proceeds by finding that certain implications of the model for the observables x^k and p^k must hold, regardless of the values of the unobservables.

We derive the axiom by considering increasingly general cases. First we consider the case of no discounting and one observation ($\delta = 1$ and $K = 1$). Then, we study the case of no discounting ($\delta = 1$ and $K \geq 1$). Finally, in Section 2.3.3 we discuss the general case (δ is unknown and $K \geq 1$) and present the axiom for EDU, SAR-EDU.

2.3.1 No Discounting and One Observation: $\delta = 1$ and $K = 1$

Suppose that $\delta = 1$ and $K = 1$. That is, we seek to impose EDU rationality in the special case when δ is known, equals 1, and our dataset has a single observation. Under these assumptions (omitting the k superindex, as $K = 1$) the first-order condition (2.2) becomes $u'(x_t) = \lambda p_t$ for each $t \in T$. For each pair $t, t' \in T$, we obtain

$$\frac{u'(x_t)}{u'(x_{t'})} = \frac{p_t}{p_{t'}}.$$

By concavity of u, for each pair $t, t' \in T$, we have

$$x_t > x_{t'} \implies \frac{p_t}{p_{t'}} \leq 1. \tag{2.3}$$

Thus we obtain a simple implication of EDU rationality in this special case: (2.3) means that *demand must slope down*. This "downward sloping demand axiom" coincides with the axiom obtained by Browning (1989) for the $\delta = K = 1$ case. [2]

Property (2.3) can be written in a different way. It is more complicated, and redundant for now, but will prove useful in the sequel:

Definition 3. A sequence of pairs $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^n$ has the *downward sloping demand property* if

$$x_{t_i}^{k_i} > x_{t'_i}^{k'_i} \text{ for all } i \text{ implies that } \prod_{i=1}^{n} \frac{p_{t_i}^{k_i}}{p_{t'_i}^{k'_i}} \leq 1.$$

The downward-sloping demand property is not only a necessary condition, but also a sufficient condition for EDU rationality in the case of $\delta = 1$ and $K = 1$.

[2] Browning (1989) is interested in the case of $K = 1$ because he uses survey consumption data.

2.3.2 No Discounting: $\delta = 1$

We now take one step towards our general result. Continue to assume that $\delta = 1$, but now allow that $K \geq 1$. The decision maker does not discount future consumptions, but the dataset may contain multiple observations. The first-order condition (2.2) becomes $u'(x_t^k) = \lambda^k p_t^k$ for each $t \in T$, and each $k \in K$.

If we try to proceed as in the previous section, we might consider pairs of observations with $x_t^k > x_{t'}^{k'}$:

$$\frac{u'(x_t^k)}{u'(x_{t'}^{k'})} = \frac{\lambda^k}{\lambda^{k'}} \frac{p_t^k}{p_{t'}^{k'}}.$$

By the concavity of u, we know that

$$x_t^k > x_{t'}^{k'} \Longrightarrow \frac{\lambda^k}{\lambda^{k'}} \frac{p_t^k}{p_{t'}^{k'}} \leq 1.$$

However, the ratio $\lambda^k/\lambda^{k'}$ does not allow us to conclude anything about the ratio of prices. We would like to conclude, along the lines of downward sloping demand, that $p_t^k/p_{t'}^{k'} \leq 1$. But the presence of $\lambda^k/\lambda^{k'}$ does not allow us to do that. Of course if we consider $x_t^k > x_{t'}^{k'}$ for the same observation ($k = k'$) then the conclusion of downward sloping demand continues to hold. When $K > 1$, downward sloping demand is still a restriction within each observation $k \in K$.

This suggests that we can obtain an implication of EDU (with $\delta = 1$) across observations as well. Consider a collection of pairs $(x_t^k, x_{t'}^{k'})$, chosen such that the λ variables will cancel out. For example, consider:

$$\frac{u'(x_{t_1}^k)}{u'(x_{t_2}^{k'})} \frac{u'(x_{t_3}^{k'})}{u'(x_{t_4}^k)} = \frac{\lambda^k}{\lambda^{k'}} \frac{\lambda^{k'}}{\lambda^k} \frac{p_{t_1}^k}{p_{t_2}^{k'}} \frac{p_{t_3}^{k'}}{p_{t_4}^k}.$$

Then the λ variables cancel out and we obtain that:

$$x_{t_1}^k > x_{t_2}^{k'} \text{ and } x_{t_3}^{k'} > x_{t_4}^k \Longrightarrow \frac{p_{t_1}^k}{p_{t_2}^{k'}} \frac{p_{t_3}^{k'}}{p_{t_4}^k} \leq 1,$$

that is, downward sloping demand.

The idea of canceling out the unknown λ^k's suggests the following definition.

Definition 4. A sequence of pairs $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$ is *balanced* if each k appears as k_i (on the left of the pair) the same number of times it appears as k_i' (on the right).

When $K = 1$ we know that a sequence must have the downward sloping demand property. Now with $K \geq 1$ this is only true of *balanced* sequences: *any balanced sequence has the downward sloping demand property.* This property is not only necessary, but also a sufficient condition for EDU rationality in the case when δ is known and $\delta = 1$.

2.3.3 General K and δ

We now turn to the case when K can be arbitrary and δ is unknown. Before, when $K = \delta = 1$, then λ was constant and δ was fixed. EDU rationality is characterized by downward sloping demand. When $K \geq 1$ we saw that we needed to impose downward sloping demand for balanced sequences. When δ is unknown we need to further restrict the sequences that are required to satisfy downward sloping demand. In fact, the relevant axiom turns out to be:

Strong Axiom of Revealed Exponentially Discounted Utility (SAR-EDU): *For any balanced sequence of pairs* $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^n$, *if* $\sum_{i=1}^n t_i \geq \sum_{i=1}^n t'_i$, *then the sequence has the downward sloping demand property.*

As in Sections 2.3.1 and 2.3.2, the key idea is to control the effects of the unknowns u, δ, and λ, by focusing on particular configurations of the data. It is easy to see how such restrictions are necessary. For example, consider two pairs

$$(x_{t_1}^{k_1}, x_{t_2}^{k_2}) \text{ and } (x_{t_3}^{k_2}, x_{t_4}^{k_1})$$

such that

$$t_1 + t_3 \geq t_2 + t_4.$$

By manipulating first-order conditions we obtain that:

$$\frac{u'(x_{t_1}^{k_1})}{u'(x_{t_2}^{k_2})} \frac{u'(x_{t_3}^{k_2})}{u'(x_{t_4}^{k_1})} = \left(\frac{\delta^{t_2}}{\delta^{t_1}} \frac{\lambda^{k_1} p_{t_1}^{k_1}}{\lambda^{k_2} p_{t_2}^{k_2}} \right) \left(\frac{\delta^{t_4}}{\delta^{t_3}} \frac{\lambda^{k_2} p_{t_3}^{k_2}}{\lambda^{k_1} p_{t_4}^{k_1}} \right) = \delta^{(t_2+t_4)-(t_1+t_3)} \frac{p_{t_1}^{k_1}}{p_{t_2}^{k_2}} \frac{p_{t_3}^{k_2}}{p_{t_4}^{k_1}}.$$

Notice that the pairs $(x_{t_1}^{k_1}, x_{t_2}^{k_2})$ and $(x_{t_3}^{k_2}, x_{t_4}^{k_1})$ constitute a balanced sequence of pairs, so that the Lagrange multipliers cancel out as in Section 2.3.2. Furthermore, the discount factor unambiguously increases the value on the left hand side, $\delta^{(t_2+t_4)-(t_1+t_3)} \geq 1$ for any $\delta \in (0, 1]$.

Now the concavity of u implies that when $x_{t_1}^{k_1} > x_{t_2}^{k_2}$ and $x_{t_3}^{k_2} > x_{t_4}^{k_1}$ then the product $\delta^{(t_2+t_4)-(t_1+t_3)}(p_{t_1}^{k_1}/p_{t_2}^{k_2})(p_{t_3}^{k_2}/p_{t_4}^{k_1})$ cannot exceed 1. Since $\delta^{(t_2+t_4)-(t_1+t_3)} \geq 1$ for any $\delta \in (0, 1]$, then $(p_{t_1}^{k_1}/p_{t_2}^{k_2})(p_{t_3}^{k_2}/p_{t_4}^{k_1})$ cannot exceed 1. Thus, we obtain an implication of EDU for prices, an observable entity. *No matter what the values of the unobservable δ and u, we find that the ratio of prices cannot be more than 1.*

The argument just made extends to arbitrary balanced sequences, and essentially gives the proof of necessity of SAR-EDU. [3] The argument simply amounts to verifying a rather basic consequence of EDU: the consequence of EDU for those situations in which unobservables either do not matter or have a known effect (the effect either resulting from u' being decreasing or from $\delta \in (0, 1]$). What is surprising is that such a basic consequence of the theory is sufficient as well as necessary.

Theorem 1. *A dataset is EDU rational if and only if it satisfies SAR-EDU.*

The proof is in Section 2.6. The proof that SAR-EDU is necessary is, as we have remarked, simple. The proof of sufficiency is more complicated, and follows ideas introduced in Echenique and Saito (2015).

Remark 2. It is not obvious from the syntax of SAR-EDU that one can verify whether a particular dataset satisfies SAR-EDU in finitely many steps. We can show that, not only is SAR-EDU decidable in finitely many steps, but there is in fact an efficient algorithm that decides whether a dataset satisfies SAR-EDU. The proof is very similar to Proposition 2 in Echenique and Saito (2015). So we omit the proof. SAR-EDU is on the same computational standing as GARP or the strong axiom of revealed preference. Another way to test SAR-EDU is based on linearized "Afriat inequalities," see Lemma 1 of Section 2.6.3. In fact, this is how we proceed in Section 2.5; see in particular the discussion at the end of that section.

2.4 More General Models

The ideas behind Theorem 1 can be used to analyze other models of intertemporal choice, including quasi-hyperbolic discounting (QHD), and more general models.

2.4.1 Quasi-Hyperbolic Discounted Utility

First we investigate QHD. The objective is the same as for EDU: we want to know when a dataset $(x^k, p^k)_{k=1}^K$ is consistent with QHD utility maximization, but the

[3] We have assumed differentiability of u in our informal derivation, but since u is concave, we can easily generalize the argument.

interpretation of a dataset is now more complicated. In the case of QHD, we assume that each x^k is a consumption stream that the agent *commits* to at date 0. The reason is that a QHD agent may be dynamically inconsistent, and revise their planned consumption. The commitment assumption happens to perfectly fit the application in Section 2.5 to the CTB experiment in Andreoni and Sprenger (2012a). The commitment assumption will, however, be violated by field data taken from consumption surveys. It is important to note that the assumption of commitment is not necessary to test the EDU model, which is dynamically consistent.

Definition 5. A dataset $(x^k, p^k)_{k=1}^K$ is *quasi-hyperbolic discounted utility rational (QHD rational)* if there is $\delta \in (0, 1]$, index for time-bias $\beta > 0$, and a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$ such that, for all k,

$$y \in B(p^k, p^k \cdot x^k) \Longrightarrow \sum_{t \in T} D(t) u(y_t) \leq \sum_{t \in T} D(t) u(x_t^k),$$

where $D(t) = 1$ if $t = 0$ and $D(t) = \beta\delta^t$ if $t > 0$. More specifically, if $\beta \leq 1$ in the above definition then the dataset $(x^k, p^k)_{k=1}^K$ is *present-biased quasi-hyperbolic discounted utility rational (PQHD-rational)*.

Strong Axiom of Revealed Quasi-Hyperbolic Discounted Utility (SAR-QHD): *For any balanced sequence of pairs* $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$, *if*

1. $\sum_{i=1}^n t_i \geq \sum_{i=1}^n t_i'$ *and*
2. $\#\{i : t_i > 0\} = \#\{i : t_i' > 0\}$,

then the sequence has the downward sloping demand property.

The condition that $\sum_{i=1}^n t_i \geq \sum_{i=1}^n t_i'$ plays the same role as it did in SAR-EDU, to control the effect of δ. In addition, we must now have $\#\{i : t_i > 0\} = \#\{i : t_i' > 0\}$ so as to cancel out β. If we instead focus on PQHD, then we know that $\beta \leq 1$ so the weaker requirement $\#\{i : t_i > 0\} \geq \#\{i : t_i' > 0\}$ controls the effect of β. [4] Formally, the axiom to characterize PQHD is as follows:

Strong Axiom of Revealed Present-Biased Quasi-Hyperbolic Discounted Utility (SAR-PQHD): *For any balanced sequence of pairs* $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$, *if*

1. $\sum_{i=1}^n t_i \geq \sum_{i=1}^n t_i'$ *and*

[4] It is easy to axiomatize *future-biased quasi-hyperbolic discounted utility (FQHD)*, which is a special model of QHD with $\beta \geq 1$. For FQHD, in turn, we need $\#\{i : t_i > 0\} \leq \#\{i : t_i' > 0\}$.

2. $\#\{i : t_i > 0\} \geq \#\{i : t'_i > 0\}$,

then the sequence has the downward sloping demand property.

To show the necessity of SAR-QHD, we proceed as in Section 2.3.3. The first order conditions for maximization of a QHD utility are:

$$D(t)u'(x_t^k) = \lambda^k p_t^k. \tag{2.4}$$

For example, consider a balanced sequence of pairs $(x_{t_1}^{k_1}, x_{t_2}^{k_2})$, $(x_{t_3}^{k_2}, x_{t_4}^{k_1})$ with the property that $t_1 + t_3 \geq t_2 + t_4$, and $\#\{i \in \{1,3\} : t_i > 0\} = \#\{i \in \{2,4\} : t_i > 0\}$; where $\#\{i \in \{1,3\} : t_i > 0\}$ and $\#\{i \in \{2,4\} : t_i > 0\}$ are the numbers of non-time-zero consumption in $\{x_{t_1}^{k_1}, x_{t_3}^{k_3}\}$ and $\{x_{t_2}^{k_2}, x_{t_4}^{k_4}\}$, respectively. By manipulating the first-order conditions we obtain that:

$$\begin{aligned}
\frac{u'(x_{t_1}^{k_1})}{u'(x_{t_2}^{k_2})}\frac{u'(x_{t_3}^{k_2})}{u'(x_{t_4}^{k_1})} &= \left(\frac{\beta^{1_{\{t_2>0\}}}}{\beta^{1_{\{t_1>0\}}}}\frac{\delta^{t_2}}{\delta^{t_1}}\frac{\lambda^{k_1}p_{t_1}^{k_1}}{\lambda^{k_2}p_{t_2}^{k_2}}\right)\left(\frac{\beta^{1_{\{t_4>0\}}}}{\beta^{1_{\{t_3>0\}}}}\frac{\delta^{t_4}}{\delta^{t_3}}\frac{\lambda^{k_2}p_{t_3}^{k_2}}{\lambda^{k_1}p_{t_4}^{k_1}}\right) \\
&= \beta^{\#\{i\in\{2,4\}:t_i>0\}-\#\{i\in\{1,3\}:t_i>0\}}\delta^{(t_2+t_4)-(t_1+t_3)}\frac{p_{t_1}^{k_1}}{p_{t_2}^{k_2}}\frac{p_{t_3}^{k_2}}{p_{t_4}^{k_1}} \\
&= \delta^{(t_2+t_4)-(t_1+t_3)}\frac{p_{t_1}^{k_1}}{p_{t_2}^{k_2}}\frac{p_{t_3}^{k_2}}{p_{t_4}^{k_1}}.
\end{aligned}$$

The balancedness of the sequence of pairs $(x_{t_1}^{k_1}, x_{t_2}^{k_2})$ and $(x_{t_3}^{k_2}, x_{t_4}^{k_1})$ implies that Lagrange multipliers cancel out. The assumption of $\#\{i \in \{1,3\} : t_i > 0\} = \#\{i \in \{2,4\} : t_i > 0\}$ implies that β cancels out. As in SAR-EDU, the discount factor unambiguously increases the value of the right hand side.

Finally, concavity of u implies that, when $x_{t_1}^{k_1} > x_{t_2}^{k_2}$ and $x_{t_3}^{k_2} > x_{s_4}^{k_1}$, we have that $(p_{t_1}^{k_1}/p_{t_2}^{k_2})(p_{t_3}^{k_2}/p_{t_4}^{k_1})$ cannot exceed 1. That is, the downward sloping demand property holds.

The next theorem summarizes our results on QHD.

Theorem 2. *A dataset is QHD-rational if and only if it satisfies SAR-QHD. Moreover, the dataset is PQHD-rational if and only if it satisfies SAR-PQHD.*

The proof of Theorem 2 is in Section 2.7.

One consequence of Theorems 1 and 2 is that, under certain circumstances, EDU and PQHD are *observationally equivalent*. These circumstances are very relevant for the discussion in Section 2.5 of AS's experiment: our next result, Proposition 1, shows that if an agent does not consume at the soonest date (i.e., $x_0^k = 0$

for all $k \in K$), then EDU and PQHD are observationally equivalent. In AS's experiment, 82.8% of the subjects (i.e., 24 out of 29 subjects) who satisfy SAR-EDU do not consume at the soonest date. This explains why, in AS's data, QHD has very limited scope beyond what can be explained by EDU.

Proposition 1. *Suppose that a dataset $(x^k, p^k)_{k=1}^K$ satisfies that $x_0^k = 0$ for all $k \in K$. Then $(x^k, p^k)_{k=1}^K$ is PQHD rational if and only if it is EDU rational.*

Proof. Of course, if the data is EDU rational then it is PQHD rational. Let us prove the converse. Choose a sequence $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$ such that (1) $x_{t_i}^{k_i} > x_{t_i'}^{k_i'}$ for all $i \in \{1, \ldots, n\}$, (2) $\sum_{i=1}^n t_i \geq \sum_{i=1}^n t_i'$, and (3) each k appears as k_i the same number of times as k_i'.

By (1), $x_{t_i}^{k_i} > 0$ for all $i \in \{1, \ldots, n\}$. Since $x_0^k = 0$ for all $k \in K$, we obtain $t_i > 0$ for all $i \in \{1, \ldots, n\}$. Therefore, $\#\{i \in \{1, \ldots, n\} : t_i > 0\} = \#\{i \in \{1, \ldots, n\}\} \geq \#\{i \in \{1, \ldots, n\} : t_i' > 0\}$. Therefore, the sequence satisfies all of the conditions in SAR-PQHD. Since the dataset is PQHD rational, Theorem 2 shows that

$$\prod_{i=1}^n \frac{p_{t_i}^{k_i}}{p_{t_i'}^{k_i}} \leq 1. \tag{2.5}$$

Therefore, Conditions (1), (2), and (3) imply (2.5), which is SAR-EDU. Therefore, by Theorem 1, the dataset must be EDU rational. □

2.4.2 More General Models of Time Discounting

Building on the ideas in the previous two theorems, we can characterize more general models of intertemporal choice. These models end up being useful in Section 2.5.2 when we classify subjects in AS's experiment.

Of course the most general model is utility maximization, without constraints on the form of the utility.

$$\max \ U(x_0, \ldots, x_T) \ \text{ s.t. } \ p \cdot x \leq I. \tag{2.6}$$

The relevant revealed preference axiom is GARP. In the following, we provide three special cases of (2.6), which are obtained by restricting U. We list the three models in order of generality. Let $\mathcal{C}$ be the set of all continuous, concave, and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$.

1. *Time-separable utility* (TSU): The class TSU of all U that can be written as

$$U(x_0, \ldots, x_T) = \sum_{t \in T} u_t(x_t),$$

where $u_t \in \mathcal{C}$ for all $t \in T$.

2. *General time discounting* (GTD): The class GTD of all U that can be written as

$$U(x_0, \ldots, x_T) = \sum_{t \in T} D(t) u(x_t),$$

for some $u \in \mathcal{C}$, and a function $D : T \to \mathbf{R}_{++}$.

3. *Monotone time discounting* (MTD): The class MTD of all U that can be written as

$$U(x_0, \ldots, x_T) = \sum_{t \in T} D(t) u(x_t),$$

for some $u \in \mathcal{C}$, and a function $D : T \to \mathbf{R}_{++}$ that is monotonically decreasing.

In the following definition, the set M of utility functions can be any of the classes defined above (i.e., TSU, GTD, MTD).

Definition 6. For M $\in$ {TSU, GTD, MTD}, a dataset $(x^k, p^k)_{k=1}^K$ is *M rational* if there is a utility function U in the class M of utilities such that for all k,

$$p^k \cdot y \leq p^k \cdot x^k \Longrightarrow U(y) \leq U(x^k).$$

It is easy to derive each axiom from first-order conditions as we did for EDU and QHD. The idea is to choose a sequence of pairs of observations so that we can cancel out the Lagrange multipliers and control or cancel out the effects of other unobservables. We omit the derivations.

Strong Axiom of Revealed Time Separable Utility (SAR-TSU): *For any balanced sequence of pairs* $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})_{i=1}^n$*, if each* $t_i = t'_i$ *for all* i*, then the sequence has the downward sloping demand property.*

Strong Axiom of Revealed General Time Discounted Utility (SAR-GTD): *For any balanced sequence of pairs* $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})_{i=1}^n$*, if each* t *appears as* t_i *(on the left of the pair) the same number of times it appears as* t'_i *(on the right), then the sequence has the downward sloping demand property.*

Strong Axiom of Revealed Monotone Time Discounted Utility (SAR-MTD): *For any balanced sequence of pairs* $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^{n}$*, if there is a permutation* π *of* $\{1, 2, \ldots, n\}$ *such that* $t_i \geq t'_{\pi(i)}$*, then the sequence has the downward sloping demand property.*

Each of these axioms imposes the downward sloping demand property of a balanced sequence under increasingly demanding conditions. For example, SAR-TSU imposes the downward sloping demand property of a subset of the sequences that are constrained by SAR-MTD; and SAR-MTD in turn constrains fewer sequences than SAR-EDU. How demanding an axiom is, in terms of imposing the downward sloping demand property, mirrors how demanding the theory is: EDU is a special case of MTD, which is a special case of TSU.

Theorem 3. *Let* $M \in \{TSU, GTD, MTD\}$*. A dataset is M-rational if and only if it satisfies SAR-M.*

The proof of Theorem 3 follows similar ideas to those used in the proofs of the other two results, and is relegated to Section A.1.

2.5 Empirical Application

2.5.1 Description of the Data

AS introduce an experimental method called the Convex Time Budget (CTB) design. In contrast with the "multiple price list method" (e.g., Andersen et al., 2008a), subjects in AS are asked to allocate 100 experimental tokens between "sooner" (time τ) and "later" (time $\tau + d$) accounts. Tokens allocated to each account have a value of a_τ and $a_{\tau+d}$, converting experimental currency unit into real monetary value for final payments. The gross interest rate over d days is thus given by $a_{\tau+d}/a_\tau$. There are three possible sooner dates $\tau \in \{0, 7, 35\}$, three possible delays $d \in \{35, 70, 98\}$ (the unit of period is one day), and five different pairs of conversion rates $(a_\tau, a_{\tau+d})$. Each subject thus completes 45 decisions. [5]

Each subject's decision in a trial is characterized by a tuple $(\tau, d, a_\tau, a_{\tau+d}, c_\tau)$: the first four elements $(\tau, d, a_\tau, a_{\tau+d})$ characterize the budget set she faces in this trial, and c_τ is the number of tokens she decides to allocate to the sooner payment.

[5] See Figure A.1 in Section A.2 for an illustration. For each pair of starting date and delay length (τ, d), the five budgets are nested. Looking at all 45 budget sets, except for eight cases in which $(a_\tau, a_{\tau+d}) = (0.2, 0.25)$, $a_{\tau+d}$ is fixed at 0.2 and a_τ ranges between 0.1 and 0.2. Participants' choices therefore always satisfy GARP by design.

In the experiment, subjects make a two-period choice. They choose $(x_\tau, x_{\tau+d})$ subject to $p_\tau x_\tau + x_{\tau+d} = I$. We need to formulate the problem as choosing $(x_0, \ldots x_T)$ subject to $\sum_{t\in T} p_t x_t = I$. We set prices to be $p_\tau = a_{\tau+d}/a_\tau$ and $p_{\tau+d} = 1$ (a normalization), and we define consumptions (monetary amounts) $x_\tau = c_\tau \cdot a_\tau$ and $x_{\tau+d} = (100 - c_\tau) \cdot a_{\tau+d}$.

We shall implicitly set the prices of periods that are not offered to be very high, so that agents choose zero consumption in those periods. For example, when subjects face a convex budget with $(\tau, d) = (35, 70)$, we treat prices p_t for $t \neq 35, 105$ as high. In any of the rationalizations we consider, marginal utilities at zero are finite. So by setting such prices high enough, the choices in such time periods do not affect whether a dataset is rationalizable. In this way, for each of the 97 subjects, we obtained a dataset with $K = 45$ and $T = \{t : t = \tau \text{ or } t = \tau + d \text{ for some } \tau \in \{0, 7, 35\} \text{ and } d \in \{35, 70, 98\}\}$.

Three features of the AS design make their experiment ideal for our exercise. First and most importantly, the experimental setup is precisely the situation our model tries to capture: subjects choose an intertemporal consumption from a budget set. As we briefly mention above, most previous experimental studies on intertemporal decision utilize an environment with discrete (in many cases, binary) choice sets. Strictly speaking, budgets in AS experiment are discrete as well, but we understand them to be a reasonable approximation to continuous choice (tokens are worth \$0.1 to \$0.25).

Secondly, the AS experiment has subjects *committing* to a consumption stream. Recall that to test for QHD and more general models (although not for EDU) we need to assume that agents commit to a consumption stream. In the AS design, the commitment assumption is satisfied.

Thirdly, AS put significant effort into equalizing the transaction costs of sooner and later payments, and minimizing the unwanted effects of uncertainty regarding future payments.

Before discussing our results, we summarize AS's main findings. AS estimate the per-period discount factor, present bias, and utility curvature assuming a QHD model with CRRA utility over money:

$$U(x_0, \ldots, x_T) = \frac{1}{\alpha} x_0^\alpha + \beta \sum_{t \in T \setminus \{0\}} \delta^t \frac{1}{\alpha} x_t^\alpha. \tag{2.7}$$

Their estimation uses pooled data from all subjects, fitting a common specification (2.7). AS find no evidence of present bias ($\widehat{\beta} = 1.007$, $SE = 0.006$; the

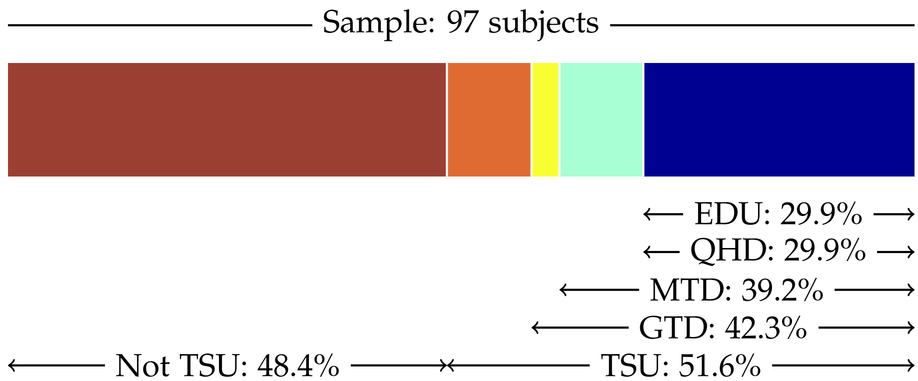


FIGURE 2.1: Classification of subjects in AS's experiment.

hypothesis of no present bias, $\beta = 1$, is not rejected; $F_{1,96} = 1.51$, $p = 0.22$).[6] AS also estimate (2.7) at the level of individual subjects and find that the estimated $\widehat{\beta}$'s are narrowly distributed around 1, with the median estimate being 1.0011.

2.5.2 Results

We test our axioms for each individual subject in AS's experiment. Note that we do not pool the choice data of different subjects. The tests are based on the linearized Afriat inequalities presented in Lemma 1. The models we examine are EDU, QHD, MTD, GTD, and TSU. In the sequel, we shall label a subject as "M rational" if her dataset passes the revealed preference test for model M and "M non-rational" otherwise. The models can be ordered by the tightness of the associated axioms. Essentially, we have that EDU $\subset$ PQHD $\subset$ MTD $\subset$ GTD $\subset$ TSU, and that EDU $\subset$ QHD $\subset$ GTD $\subset$ TSU, as QHD is not comparable to MTD (QHD allows $\beta > 1$). For this reason, when we find that a subject is EDU rational, she is of course also M rational for all other models M $\in \{\text{PQHD}, \text{QHD}, \text{MTD}, \text{GTD}, \text{TSU}\}$. We sometimes label a subject as "strictly M rational" for the most restrictive model M such that the agent is M rational. For example, a subject is strictly QHD rational if her dataset passes the QHD test but not the EDU test.

Figure 2.1 summarizes the results. We find that 29 subjects are EDU rational. QHD also rationalizes the same 29 subjects: there are *no* subjects who are strictly QHD rational. As we mentioned in Proposition 1, this is related to agents' peculiar

[6] AS estimate several model specifications (e.g., assuming CARA instead of CRRA, or incorporating additional parameters to capture background consumptions), and they also use different estimation methods (e.g., two-limit Tobit model to handle corner choices). In our comparison, we use their results from a nonlinear least squares estimation of quasi-hyperbolic discounting and CRRA utility function without background consumption.

pattern of choices. Proposition 1 shows that if an agent does not consume at the soonest date (i.e., $x_0^k = 0$ for all $k \in K$), then EDU and PQHD are observationally equivalent. In AS's experiment, more than 82.8% of the subjects who satisfy SAR-EDU (i.e., 25% of the total subjects) do not consume at the soonest date.

EDU and QHD are arguably the most important models of intertemporal choice used in economics, but it is interesting to go beyond these models and look at the more general utility functions described in Section 2.4.2. We find that nine additional subjects (9.3%) have utility functions in MTD, three additional subjects (3.1%) have utility functions in GTD, and nine more subjects (9.3%) become rational by allowing a general TSU. In all, 51.6% of subjects can be rationalized by one of the time-separable models.

In summary, while AS find moderate support for EDU, our conclusion is closer to a rejection of EDU. In fact, close to half of the subjects in the experiment do not even exhibit time separable preferences. In the next section, we look at why our methods and AS's give seemingly contradictory conclusions from the same data.

Similarities and differences with AS's findings. Our analysis is for individual subjects. But the main results in AS use pooled data from all subjects. If we instead focus on individual level estimates of the same parametric model as AS, the source of the differences becomes quite clear. We focus on the individual estimates from AS (see Tables A6 and A7 in the online appendix of Andreoni and Sprenger, 2012a). [7]

Figure 2.2 summarizes the comparison. Each bar in the figure corresponds to one subject. The vertical value of the bar is AS's estimated value $\widehat{\beta}$ for that subject. We categorized the subjects depending on their strict M-rationality. For example, the subjects who belong to the blue area pass our EDU test and the subjects who belong to the brown area do not pass any tests; they are not TSU rational.

There are two important facts one can glean from the figure. First, our test is consistent with AS's methodology and their estimates: the subjects who pass the EDU test have estimated $\widehat{\beta}$ very close to 1. So Figure 2.2 shows that our methodology and AS's methodology are, in fact, in agreement.

Secondly, those subjects who fail the EDU test but pass MTD, GTD, or TSU test tend to have $\widehat{\beta} \neq 1$. Moreover, those who do not pass any of the tests (i.e., TSU non-rational subjects) have estimated $\widehat{\beta}$ which are far from 1 in magni-

[7] We obtain parameters for 86 of the 97 subjects. The remaining 11 subjects were excluded from AS's analysis, since preference parameters were not estimable. We can run our tests on the 11 excluded subjects: seven of them pass the EDU and QHD tests.

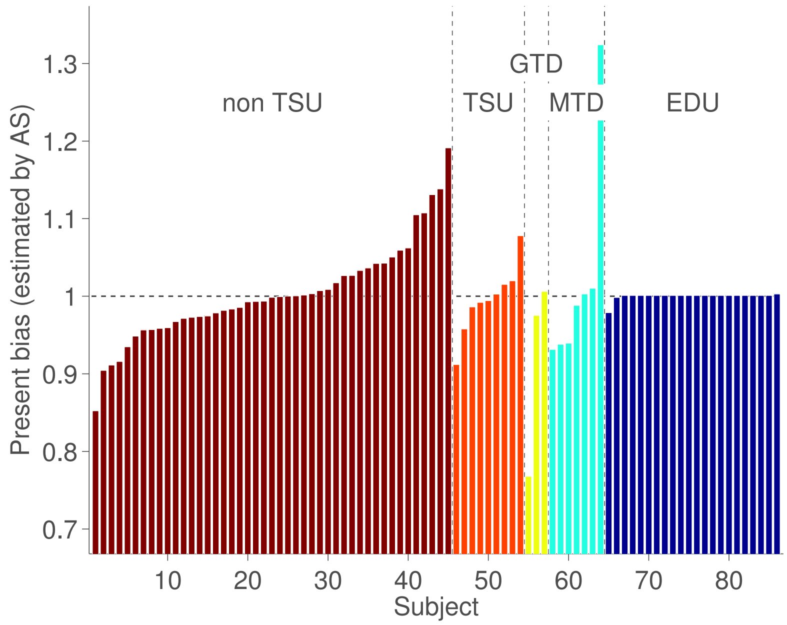


FIGURE 2.2: Estimated present-bias parameter for each category of subjects.

tude compared to the other groups of subjects, and are distributed symmetrically around 1. [8] Roughly speaking, for half of the TSU non-rational subjects, $\widehat{\beta} > 1$; for the other half, $\widehat{\beta} < 1$. Hence, the "average" subject looks, in some sense, as an EDU agent, even though the majority of subjects are inconsistent with EDU according to our test. It is therefore possible that AS's finding in favor of EDU in their aggregate preference estimation reflects the choice behavior of such an average subject.

Choice pattern of EDU rational and TSU non-rational subjects. Next we look into subjects' choice patterns, focusing on the two main groups that the subjects fall into: those that are EDU rational and those that fail the TSU test. We investigate three aspects. First, we study the fraction of choices at the corner of the budget set. Second, we check for violations of wealth monotonicity. Finally, we check for violations of WARP.

Corner vs. interior choice: For each subject, we calculate the proportion (out of 45 choices) of (i) interior allocations, (ii) corner allocations in which subjects spend all their budget on a later reward (called "all tokens later"), and (iii) corner allocations in which subjects spend all their budget on the earlier reward (called "all tokens sooner").

[8] See Section A.4 for further comparisons between AS's parametric model estimation and our nonparametric revealed preference tests.

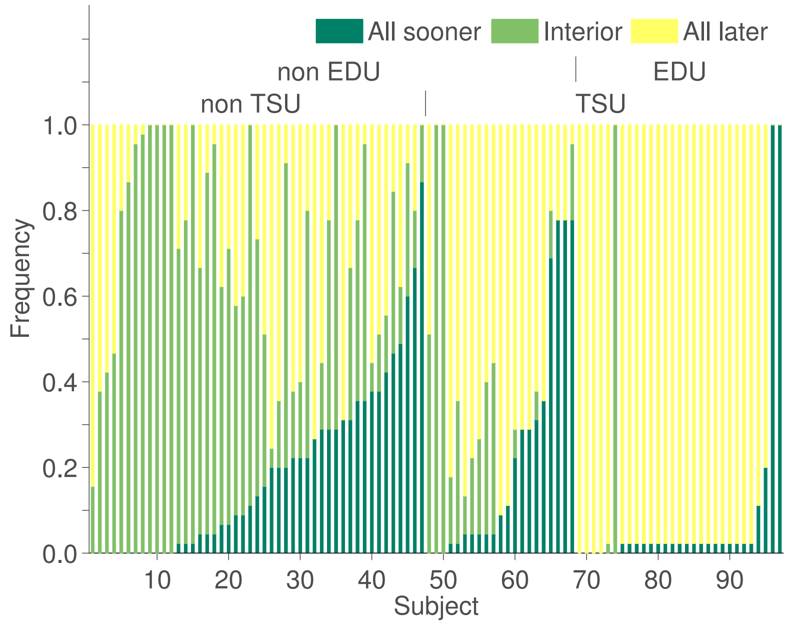


FIGURE 2.3: Individual choice patterns and class of rationality.

We observe that all but two subjects who pass our EDU test never made interior allocations during the experiment, and frequently chose to allocate all tokens to the later payments. [9] This point is made clear in Figure 2.3, which presents each subject's choice pattern, sorted by the results of our EDU and TSU tests. The fraction of interior allocations increases by moving from EDU rational subjects (only 6.9% of them, two subjects, made at least one interior allocation) to subjects who pass the TSU test but not the EDU test (66.7% of them made at least one interior allocation), and it increases further when we look at subjects who fail the TSU test (in fact, 48.9% of them chose interior allocations in at least half of the trials).

The high incidence of corner solutions for EDU rational agents should be considered in light of Proposition 1. EDU and QHD are observationally equivalent when a subject never chooses date 0 consumption, and this happens for 82.8% (24 out of 29) of EDU rational subjects. So, for the vast majority of the subjects that pass SAR-EDU, the theory would have no power to distinguish between QHD and EDU.

Wealth monotonicity: In AS's experiments, eight out of the nine time frames contain a wealth shift. We check wealth monotonicity, or normality of demand, using choices in those time frames. Monotonicity requires that c_τ and $c_{\tau+d}$ should

[9] AS already remark on the incidence of of corner choices, and comment on how this phenomenon may suggest that the curvature of utility is small and close to that of a linear function.

be weakly increasing in wealth, holding the price rate constant in the eight time frames. Obviously, all of the EDU rational subjects satisfy monotonicity. On the other hand, most of TSU non-rational subjects (43 out of 47 subjects) violate monotonicity.

Eliminating the wealth-shift observations from the data does not, however, suffice to make most subjects TSU rational. The choices of the TSU non-rational agents are inconsistent with TSU for more complicated reasons than simply a violation of normal demand.

WARP: In AS's experimental design, budget lines never cross at an interior point of the budget. However, there are budgets that cross at the corner of consuming all later (when "all later" corresponds to the same date; see Figure A.1 in the appendix). In particular, eight out of the nine time frames contain four budgets that share the same "all later" allocation at \$20. In the remaining time frame $(\tau, d) = (7, 70)$, all five budgets share the same "all later" allocation at \$20. So we can test WARP by using such choices. We found that seven out of 97 subjects violated WARP. None of these seven pass the TSU test (of course).

Distance measure 1: minimum price perturbation. We find that many subjects' choices in the AS experiment are inconsistent with EDU, QHD, and even TSU. Is a subject inconsistent with these models because she made a few mistakes, or is her behavior severely inconsistent with the model? The answer to these questions is that the violations we have detected are severe, and not due to small mistakes.

We quantify the distance to rationality using two different approaches. The first approach adds noise to the data, and measures how much noise has to be added in order to reconcile the theory with the data. This allows us to set up a statistical test, and calculate the probability of having as much noise as we need to explain the data if the model is right.

Borrowing ideas from the minimal perturbation based test studied in Varian (1985), the money pump test in Echenique et al. (2011), and the upper bound test of Fleissig and Whitney (2005), we propose a measure of distance to rationality that involves a perturbation to the model or to the data. Assuming that the perturbation is random we can calculate the probability of observing the kind of perturbation needed to rationalize the data. As a result, we obtain a statistical test: a p-value for the hypothesis that a subject is EDU rational.

The perturbation can take two equivalent forms. We can think of perturbing prices (so-called measurement error in prices, this is the approach taken in Echenique et al., 2011) or we can think of perturbing marginal utilities. In the first interpretaion, the environment faced by the subject does not fit exactly the experimental design. In the second interpretation, the subject has a random utility, and his utility function is allowed to be different in each observation. The two interpretations are mathematically equivalent; we shall stick to the first for concreteness.

Let $D_{\text{true}} = (q^k, x^k)_{k=1}^K$ denote a "true" dataset and $D_{\text{obs}} = (p^k, x^k)_{k=1}^K$ denote an "observed" dataset. The true and observed datasets are connected by the relationship $q_t^k = p_t^k \varepsilon_t^k$ for all $t = 0, \ldots, T$ and $k = 1, \ldots K$ where $\varepsilon_t^k > 0$ is a random variable.

Let H_0 and H_1 denote the null hypothesis that the true dataset D_{true} is EDU rational and the alternative hypothesis that D_{true} is not EDU rational. Consider a test statistic, which is the solution to the following optimization problem given a dataset $D_{\text{obs}} = (p^k, x^k)_{k=1}^K$:

$$\Phi^* \left((p^k, x^k)_{k=1}^K \right) = \min \left\{ \sum_{k=1}^K \sum_{t=0}^T \frac{1}{K(T+1)} \left| \log \varepsilon_t^k \right| \, \middle| \, H_0 \text{ is true} \right\}. \tag{2.8}$$

Then, we can construct a test as follows:

$$\begin{cases} \text{reject } H_0 & \text{if } \int_{\Phi^*\left((p^k, x^k)_{k=1}^K\right)}^{\infty} f_{\widehat{\Phi}}(z) dz < \alpha, \\ \text{accept } H_0 & \text{otherwise,} \end{cases}$$

where α is the size of the test and $f_{\widehat{\Phi}}$ is the pdf of the distribution of $\widehat{\Phi} = \sum_{k,t} |\log \varepsilon_t^k| / (K(T+1))$. [10] We explain details in Section A.5.

For each of the 68 EDU non-rational subjects (similarly for QHD and TSU), we calculate the test statistic Φ^*. The empirical CDFs of calculated Φ^*_{EDU} along with Φ^*_{QHD} and Φ^*_{TSU} are shown in the left panel of Figure 2.4.

In order to perform a statistical hypothesis test, we need to know the distribution of $\widehat{\Phi}$ and its critical value C_α. First we randomly draw 10,000 sets of $K \times (T+1)$ noise terms ε from a log-normal distribution. The shape of the log-normal distribution, F_ε^σ, is determined such that $\log \varepsilon \sim N(0, \sigma^2)$ with $\sigma \in \{0.025, 0.1, 0.2, 0.5\}$. [11] Second, we calculate $\widehat{\Phi} = \sum_{k,t} |\log \varepsilon_t^k| / (K(T+1))$. Since

[10] The size of a test is the probability of falsely rejecting the null hypothesis. That is, it is the probability of making a Type I error: $\alpha = P(\text{test rejects } H_0 | H_0 \text{ is true})$.

[11] Mean zero assumption of $\log \varepsilon$ is reasonable given the relationship between true and observed dataset, $q_t^k = p_t^k \varepsilon_t^k$. There is no theoretical guidance as to the choice of standard deviation σ. We instead rely on economic intuition. When $\log \varepsilon \sim N(0, 0.025)$, the distribution of ε has 25-percentile at 0.983 and 75-percentile at 1.017. This implies that a subject whose perceptual random

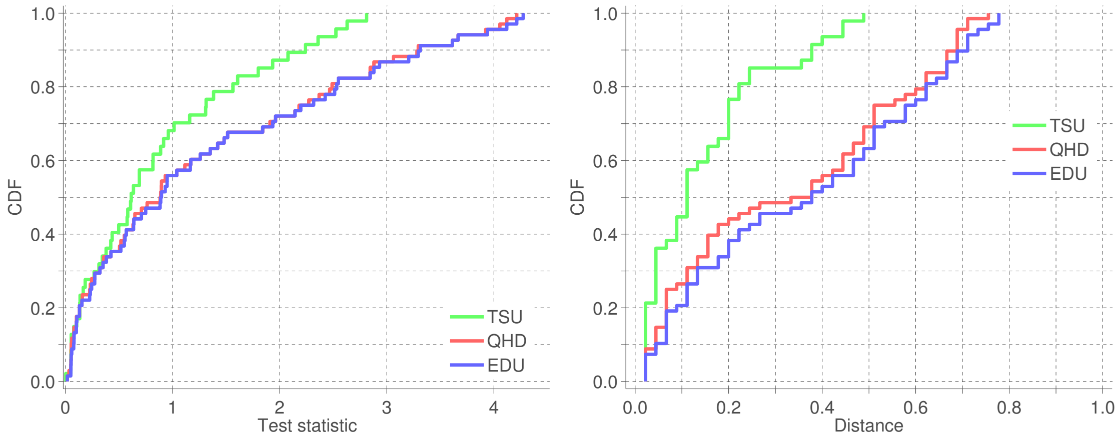


FIGURE 2.4: Empirical CDFs of distance measures based on minimum price perturbation (left) and maximal subset (right).

TABLE 2.1: H_0 rejection rates.

Standard deviation (σ) of F^σ_ε	0.0250	0.1000	0.2000	0.5000
25-percentile of F^σ_ε	0.9830	0.9301	0.8568	0.6299
75-percentile of F^σ_ε	1.0167	1.0644	1.1218	1.2365
Critical value $C^\sigma_{0.05}$	0.0226	0.0905	0.1819	0.4664
EDU	0.9853	0.8676	0.7794	0.6471
QHD	0.9853	0.8529	0.7647	0.6471
TSU	0.9706	0.7794	0.6324	0.5000

we have 10,000 of such $\widehat{\Phi}$, we have empirical distribution which well captures the true shape of the distribution. We then find critical values for 5% significance level, $C^\sigma_{0.05}$ for each σ. Table 2.1 summarizes the results of the hypothesis test under several different values of σ. The main lesson from this exercise is that even under a relatively large standard deviation of $\sigma = 0.5$ for $\log \varepsilon$, we reject the null hypothesis that the true dataset $(q^k, x^k)_{k=1}^K$ is EDU rational for about 65% of EDU non-rational subjects.

Distance measure 2: maximal subset. The second approach to quantify the distance to rationality relies on dropping observations until the data can be rationalized. If one has to drop many observations to rationalize the data, then the data

shocks are drawn from this particular distribution inflates or deflates prices at most 1.7% with probability 50%. When $\log \varepsilon \sim N(0, 0.5)$, the distribution of ε has 25-percentile at 0.630 and 75-percentile at 1.237, which indicates that much larger price distortions are more likely.

is far from being rational. [12] In particular, we take the following steps. For each EDU non-rational (similarly for QHD and TSU) subject's dataset (i) we randomly drop one observation from the dataset, and (ii) we implement the EDU test. If the dataset is EDU rational, we stop here. Otherwise, we drop another observation randomly and test for EDU rationality again, and (iii) we repeat this procedure until the subset becomes EDU rational.

Ideally, one would check all possible subsets of data, but such a calculation is obviously computationally infeasible. Our approach of sequentially choosing (at random) one observation to drop is a rough approximation to the ideal measure. In particular, the conclusion can depend on the particular sequence chosen. To address this problem we iterate the process $30,000$ times for each EDU non-rational subject. [13] Let n_m be the number of observations required to be dropped from the original dataset to make the subdata EDU rational, in the m-th iteration. We define the distance of the dataset from EDU rationality by $d'_{\text{EDU}} = \min\{n_1, \ldots, n_{30000}\}/45$. By definition, the measure is between 0 and 1, and the smaller d'_{EDU} is the closer the dataset to be EDU rational. We also note that the measure is an upper bound on the distance we want to capture, due to the random nature and path-dependence of our approach. [14]

The right panel of Figure 2.4 shows the empirical CDFs of d'_{EDU} along with d'_{QHD} and d'_{TSU}. Note that the sample size is different for each line: d'_{EDU} and d'_{QHD} are calculated for the 68 EDU and QHD non-rational subjects, while d'_{TSU} is calculated for the 47 TSU non-rational subjects. We find that the median d'_{TSU} is 0.111, implying that half of the 47 TSU non-rational subjects become TSU rational by dropping at most 11% of the observations. For EDU and QHD, on the other hand, more observations need to be dropped to rationalize the data: median d'_{EDU} and d'_{QHD} are 0.378 and 0.356, respectively. This shows that subjects' violation of EDU and QHD are not due to small mistakes. [15]

[12] This approach is motivated by Houtman and Maks (1985), who measure the distance to rationality by finding the largest subset of observations that is consistent with GARP.

[13] We first performed 10,000 iterations and then prepared two additional sets, of 10,000 iterations each, as a way to check robustness of our approach. One might worry that this sampling approach may be far from the optimal exhaustive search over all subsets, but we increased the sample size very significantly without detecting important changes. We refer to Section A.6 for more details.

[14] We should observe $d'_{\text{EDU}} \geq d'_{\text{QHD}} \geq d'_{\text{TSU}}$ as a logical consequence (if the subset of data, after dropping n observations, is EDU rational, then the same subset is QHD rational, and so on). In reality, however, due to sample variations in the stochastic algorithm we use to compute distances, we observe several instances in which $d'_{\text{EDU}} \geq d'_{\text{QHD}}$ is violated. We correct for this by simply replacing d'_{QHD} with d'_{EDU} whenever such a violation is observed.

[15] We also find that the distributions of d'_{EDU} and d'_{QHD} are almost indistinguishable (the null hypothesis of equal distribution is not rejected by the two-sample Kolmogorov-Smirnov test).

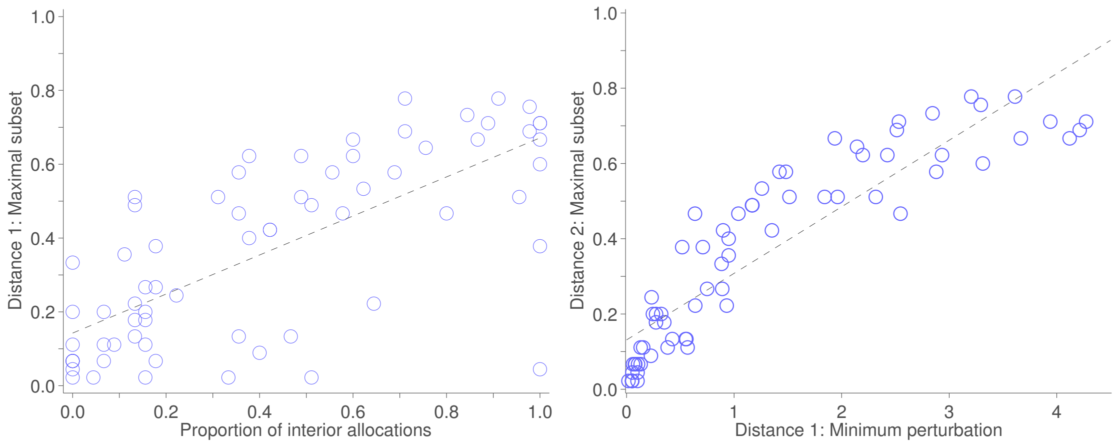


FIGURE 2.5: (Left) Distance to EDU rationality based on maximal subset and proportion of interior allocations. (Right) Correlation between two distance measures. The dotted lines represent the slopes of the least-squares fit.

In the left panel of Figure 2.5, we see a significant positive correlation (Pearson's correlation coefficient $\rho = 0.7239$, $p < 10^{-11}$) between the proportion of interior allocations and the distance to EDU rationality.

The two approaches to measuring the distance to rationality rely on different assumptions: our first measure depends on assuming a distribution of errors, and the interpretation is sensitive to the variance used on the errors. The second measure depends on approximating the minimal violating subset, and is not behaviorally founded. It is therefore comforting that the two measures give the same message. Indeed, the right panel of Figure 2.5 shows that the two measures are highly correlated (Pearson's correlation coefficient $\rho = 0.8999$, $p < 10^{-25}$).

Jittering analysis. Aside from distance, we consider the robustness of our results in a different sense. We studied how "knife edge" the satisfaction of an axiom can be. Is it possible that subjects have preferences in model M, but that they have slightly unstable tastes? Could the violations of QHD be due to small instabilities in tastes? We employ a "jittering" method akin to the one discussed in Andreoni et al. (2013). [16]

We perturb utility to produce data from a synthetic consumer with slightly unstable tastes: more precisely, we assume a CRRA instantaneous utility with QHD of the form (2.7), as in AS. Given a set of estimated parameters $(\widehat{\alpha}, \widehat{\delta}, \widehat{\beta})$, we added normal noise on one of the parameters while fixing the other two, e.g., $(\widehat{\alpha} + \varepsilon, \widehat{\delta}, \widehat{\beta})$ where $\varepsilon \sim N(0, \sigma^2)$. We set σ, the standard deviation of jittering, to

[16] We appreciate insightful comments from Jim Andreoni and Ben Gillen on this subject.

equal the standard error of the corresponding parameter estimate. We simulate choices with such "jittered" parameters, and then apply our test.

First, we take parameters and standard errors from the aggregate estimation in AS: $(\alpha, \delta, \beta) = (0.897, 0.999, 1.007)$, $(\text{se}(\alpha), \text{se}(\delta), \text{se}(\beta)) = (0.0085, 1.8 \times 10^{-4}, 0.0058)$. [17] For each parameter, we simulate 1,000 jittered versions of parameters, predict choices, and perform the QHD test. We observe a 100% pass rate no matter which parameter is jittered, suggesting that our QHD test is robust to small perturbations to the underlying preference parameters.

Secondly, we perform the same exercise using AS's individual parameter estimates and standard errors, restricting our attention to those subjects who pass our QHD test (and whose parameters are estimable by AS). For each subject and each parameter, we draw 100 jittered versions of the parameter using estimated standard errors, predict choices, and perform the QHD test. This procedure gives us pass rates for QHD for each subject. We observe 100% pass rate for 20 out of 22 subjects when α is jittered, all 22 subjects when δ is jittered, and all 22 subjects when β is jittered. As in the case of the aggregate parameter estimates, the QHD test is robust to perturbation of the underlying preference parameters.

We have performed a similar exercise while perturbing choices instead of utility parameters. We prefer the method of perturbing utility because the story of slightly unstable tastes is more appealing than the idea that agents "tremble" when making a choice. The conclusion of this analysis is not as clearly in favor of the robustness of our tests, and it depends on what one takes to be the relevant jittering standard deviations. The results are in Section A.7.

Power of the tests. Finally, we discuss the power of our tests. It is well known that tests in revealed preference theory can have low power when used on certain configurations of budget sets. The low power of GARP is well documented. As a result, it is common to assess the power of a test by comparing the pass rates (the fraction of choices that pass the relevant revealed preference axiom) from purely random choices. [18] Here we report the results from such an assessment using our tests and the experimental design of AS. We find no evidence of low power.

We generate 10,000 datasets in which choices are made at random and uni-

[17] Table 3, column (3) in Andreoni and Sprenger (2012a).

[18] The idea of using random choices as a benchmark is first applied to revealed preference theory by Bronars (1987). This approach is the most popular in empirical application: see, among other studies, Adams et al. (2014), Andreoni and Miller (2002), Beatty and Crawford (2011), Choi et al. (2007), Crawford (2010), Dean and Martin (forthcoming), Fisman et al. (2007). For overview of power calculation, see discussions in Andreoni et al. (2013) and Crawford and De Rock (2014).

formly distributed on the frontier of the budget set (Method 1 of Bronars, 1987). Datasets generated in this way always fail our tests. Next, we apply the simple bootstrap method to look at the power from an ex post perspective, as originally introduced in Andreoni and Miller (2002). For each of 45 budget sets, we randomly pick one choice from the set of choices observed in the entire experiment (i.e., 97 observations for each budget). We generate 10,000 such datasets and apply our revealed preference tests. We again observe high percentages of violation.

The conclusion is that our tests seem to have good power against the (admittedly crude) alternative of random choices. This is a credit to the design of AS.

Afriat inequalities. It should be said that the empirical implementation of our test rests on a set of Afriat inequalities, and not on explicitly checking the axioms. The Afriat inequalities are new to this study, though (see Lemma 1), and different from the standard approach to developing Afriat inequalities in the revealed preference literature. The new form of Afriat inequalities may seem ex-post (now that we know them) like a minor idea, but they were not ex-ante obvious. There are several papers (Adams et al., 2014; Crawford, 2010; Demuynck and Verriest, 2013) in the revealed preference literature that formulate the inequalities in the traditional fashion. The system of inequalities is then not linear (and cannot be linearized like our system can). As a result, these authors resort to a grid search over a finite set of values of the discount factor. The grid search can be a real limitation: we have examples in which our test gives higher pass rates for EDU than what the authors' methods give. Presumably the reason is that the grid does not allow one to conclude with certainty that an agent is not EDU rational, as it does not take full advantage of δ having arbitrary values in $(0, 1]$. So, in a sense, one of the key innovations of this study are the new Afriat inequalities. These are crucial for both the theoretical results and the empirical implementation. [19]

2.6 Proof of Theorem 1

We present the proof of the equivalence between EDU rationality and SAR-EDU.

The proof is based on using the first-order conditions for maximizing a utility with the EDU over a budget set. Our first lemma ensures that we can without loss of generality restrict attention to first-order conditions. The proof of the lemma is the same as that of Lemma 3 in Echenique and Saito (2015) with the changes of S

[19] The other key theoretical insight is the approximation result in Lemma 6.

to T and $\{\mu_s\}_{s\in S}$ to $\{\delta^t\}_{t\in T}$, where μ_s is the subjective probability that state $s \in S$ realizes.

We use the following notation in the proofs: $\mathcal{X} = \{x^k_t : k \in K, t \in T\}$.

Lemma 1. *Let $(x^k, p^k)_{k=1}^K$ be a dataset. The following statements are equivalent:*

1. *$(x^k, p^k)_{k=1}^K$ is EDU rational.*
2. *There are strictly positive numbers v^k_t, λ^k, and $\delta \in (0,1]$, for $t = 1, \ldots, T$ and $k = 1, \ldots, K$, such that*

$$\delta^t v^k_t = \lambda^k p^k_t, \ \ x^k_t > x^{k'}_{t'} \Longrightarrow v^k_t \leq v^{k'}_{t'}.$$

Proof. We shall prove that (1) implies (2). Let $(x^k, p^k)_{k=1}^K$ be EDU rational. Let $\delta \in (0,1]$ and $u : \mathbf{R}_+ \to \mathbf{R}$ be as in the definition of EDU rational data. Then (see, for example, Theorem 28.3 of Rockafellar, 1970), there are numbers $\lambda^k \geq 0$, $k = 1, \ldots, K$ such that if we let

$$v^k_t = \frac{\lambda^k p^k_t}{\delta^t}$$

then $v^k_t \in \partial u(x^k_t)$ if $x^k_t > 0$, and there is $\underline{w} \in \partial u(x^k_t)$ with $v^k_t \geq \underline{w}$ if $x^k_t = 0$. In fact, it is easy to see that $\lambda^k > 0$, and therefore $v^k_t > 0$.

By the concavity of u, and the consequent monotonicity of $\partial u(x^k_t)$ (see Theorem 24.8 of Rockafellar, 1970), if $x^k_t > x^{k'}_{t'} > 0$, $v^k_t \in \partial u(x^k_t)$, and $v^{k'}_{t'} \in \partial u(x^{k'}_{t'})$, then $v^k_t \leq v^{k'}_{t'}$. If $x^k_t > x^{k'}_{t'} = 0$, then $\underline{w} \in \partial u(x^{k'}_{t'})$ with $v^{k'}_{t'} \geq \underline{w}$. So $v^k_t \leq \underline{w} \leq v^{k'}_{t'}$.

In second place, we show that (2) implies (1). Suppose that the numbers v^k_t, λ^k, δ, for $t \in T$ and $k \in K$, are as in (2).

Enumerate the elements in $\mathcal{X}$ in increasing order:

$$y_1 < y_2 < \cdots < y_n.$$

Let

$$\underline{y}_i = \min\{v^k_t : x^k_t = y_i\} \text{ and } \bar{y}_i = \max\{v^k_t : x^k_t = y_i\}.$$

Let $z_i = (y_i + y_{i+1})/2$, $i = 1, \ldots, n-1$; $z_0 = 0$, and $z_n = y_n + 1$. Let f be a correspondence defined as follows:

$$f(z) = \begin{cases} [\underline{y}_i, \bar{y}_i] & \text{if } z = y_i, \\ \max\{\bar{y}_i : z < y_i\} & \text{if } y_n > z \text{ and } \forall i (z \neq y_i), \\ \underline{y}_n/2 & \text{if } y_n < z. \end{cases}$$

By assumption of the numbers v_t^k, we have that, when $y < y'$, $v \in f(y)$ and $v' \in f(y')$, then $v \leq v'$. Then the correspondence f is monotone and there is a concave function u for which $\partial u = f$ (Theorem 24.8 of Rockafellar, 1970). Given that $v_t^k > 0$ all the elements in the range of f are positive, and therefore u is strictly increasing.

Finally, for all (k,t), $\lambda^k p_t^k / \delta^t = v_t^k \in \partial u(v_t^k)$ and therefore the first-order conditions to a maximum choice of x hold at x_t^k. Since u is concave the first-order conditions are sufficient. The dataset is therefore EDU rational. □

2.6.1 Necessity

Lemma 2. *If a dataset $(x^k, p^k)_{k=1}^K$ is EDU rational, then it satisfies SAR-EDU.*

Proof. Let $(x^k, p^k)_{k=1}^K$ be EDU rational, and let $\delta \in (0,1]$ and $u : \mathbf{R}_+ \to \mathbf{R}$ be as in the definition of EDU rational. By Lemma 1, there exists a strictly positive solution v_t^k, λ^k, δ to the system in statement (2) of Lemma 1 with $v_t^k \in \partial u(x_t^k)$ when $x_t^k > 0$, and $v_t^k \geq \underline{w} \in \partial u(x_t^k)$ when $x_t^k = 0$.

Let $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$ be a sequence satisfying the three conditions in SAR-EDU. Then $x_{t_i}^{k_i} > x_{t_i'}^{k_i'}$. Suppose that $x_{t_i'}^{k_i'} > 0$. Then, $v_{t_i}^{k_i} \in \partial u(x_{t_i}^{k_i})$ and $v_{t_i'}^{k_i'} \in \partial u(x_{t_i'}^{k_i'})$. By the concavity of u, it follows that $\lambda^{k_i}\delta^{t_i'} p_{t_i}^{k_i} \leq \lambda^{k_i'}\delta^{t_i} p_{t_i'}^{k_i'}$ (see Theorem 24.8 of Rockafellar, 1970). Similarly, if $x_{t_i'}^{k_i'} = 0$, then $v_{t_i}^{k_i} \in \partial u(x_{t_i}^{k_i})$ and $v_{t_i'}^{k_i'} \geq \underline{w} \in \partial u(x_{t_i'}^{k_i'})$. So $\lambda^{k_i}\delta^{t_i'} p_{t_i}^{k_i} \leq \lambda^{k_i'}\delta^{t_i} p_{t_i'}^{k_i'}$. Therefore,

$$1 \geq \prod_{i=1}^{n} \frac{\lambda^{k_i}\delta^{t_i'} p_{t_i}^{k_i}}{\lambda^{k_i'}\delta^{t_i} p_{t_i'}^{k_i'}} = \frac{1}{\delta^{(\sum t_i - \sum t_i')}} \prod_{i=1}^{n} \frac{p_{t_i}^{k_i}}{p_{t_i'}^{k_i'}} \geq \prod_{i=1}^{n} \frac{p_{t_i}^{k_i}}{p_{t_i'}^{k_i'}},$$

as the sequence satisfies (2) and (3) of SAR-EDU, and hence $\sum t_i \geq \sum t_i'$ and the numbers λ^k appear the same number of times in the denominator as in the numerator of this product. □

2.6.2 Theorem of the Alternative

To prove sufficiency, we shall use the following lemma, which is a version of the Theorem of the Alternative. This is Theorem 1.6.1 in Stoer and Witzgall (1970). We shall use it here in the cases where $\mathbf{F}$ is either the real or the rational numbers.

Lemma 3. *Let A be an $m \times n$ matrix, B be an $l \times n$ matrix, and E be an $r \times n$ matrix. Suppose that the entries of the matrices A, B, and E belong to the commutative ordered field $\mathbf{F}$. Exactly one of the following alternatives is true.*

1. *There is $u \in \mathbf{F}^n$ such that $A \cdot u = 0$, $B \cdot u \geq 0$, $E \cdot u \gg 0$.*

2. *There is $\theta \in \mathbf{F}^r$, $\eta \in \mathbf{F}^l$, and $\pi \in \mathbf{F}^m$ such that $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$; $\pi > 0$ and $\eta \geq 0$.*

We also use the following lemma, which follows from Lemma 3 (See Border (2013) or Chambers and Echenique (2014)):

Lemma 4. *Let A be an $m \times n$ matrix, B be an $l \times n$ matrix, and E be an $r \times n$ matrix. Suppose that the entries of the matrices A, B, and E are rational numbers. Exactly one of the following alternatives is true.*

1. *There is $u \in \mathbf{R}^n$ such that $A \cdot u = 0$, $B \cdot u \geq 0$, and $E \cdot u \gg 0$.*

2. *There is $\theta \in \mathbf{Q}^r$, $\eta \in \mathbf{Q}^l$, and $\pi \in \mathbf{Q}^m$ such that $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$; $\pi > 0$ and $\eta \geq 0$.*

2.6.3 Sufficiency

We proceed to prove the sufficiency direction. An outline of the argument is as follows. We know from Lemma 1 that it suffices to find a solution to the Afriat inequalities (actually first-order conditions), written as statement (2) in the lemma. So we set up the problem to find a solution to a system of linear inequalities obtained from using logarithms to linearize the Afriat inequalities in Lemma 1.

Lemma 5 establishes that SAR-EDU is sufficient for SEU rationality when the logarithms of the prices are rational numbers. The role of rational logarithms comes from our use of a version of the theorem of the alternative (see Lemma 4).

The next step in the proof (Lemma 6) establishes that we can approximate any dataset satisfying SAR-EDU with a dataset for which the logarithms of prices are rational, and for which SAR-EDU is satisfied. This step is crucial, and somewhat delicate. [20]

Finally, Lemma 7 establishes the result by using another version of the theorem of the alternative, stated as Lemma 3 above.

[20] One might have tried to obtain a solution to the Afriat inequalities for "perturbed" systems (with prices that are rational after taking logs), and then considered the limit. This does not work because the solutions to our systems of inequalities are in a non-compact space. It is not clear how to establish that the limits exist and are well-behaved. Lemma 6 avoids the problem.

The statement of the lemmas follow. The rest of the paper is devoted to the proof of these lemmas.

Lemma 5. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-EDU. Suppose that* $\log(p_t^k) \in \mathbf{Q}$ *for all* k *and* t*. Then there are numbers* v_t^k*,* λ^k*,* δ*, for* $t \in T$ *and* $k = 1, \ldots, K$ *satisfying* (2) *in Lemma 1.*

Lemma 6. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-EDU. Then for all positive numbers* $\bar{\epsilon}$*, there exists* $q_t^k \in [p_t^k - \bar{\epsilon}, p_t^k]$ *for all* $t \in T$ *and* $k \in K$ *such that* $\log q_t^k \in \mathbf{Q}$ *and the dataset* $(x^k, q^k)_{k=1}^k$ *satisfy SAR-EDU.*

Lemma 7. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-EDU. Then there are numbers* v_t^k*,* λ^k*,* δ*, for* $t \in T$ *and* $k = 1, \ldots, K$ *satisfying* (2) *in Lemma 1.*

2.6.4 Proof of Lemma 5

We linearize the equation in system (2) of Lemma 1. The result is:

$$\log v(x_t^k) + t \log \delta - \log \lambda^k - \log p_t^k = 0, \tag{2.9}$$

$$x > x' \Longrightarrow \log v(x') \geq \log v(x), \tag{2.10}$$

$$\log \delta \leq 0. \tag{2.11}$$

In the system comprised by (2.9), (2.10), and (2.11), the unknowns are the real numbers $\log v_t^k$, $\log \delta$, $k \in K$ and $t \in T$.

First, we are going to write the system of inequalities (2.9) and (2.10) in matrix form. We shall define a matrix A such that there are positive numbers v_t^k, λ^k, and δ the logs of which satisfy equation (2.9) if and only if there is a solution $u \in \mathbf{R}^{K \times (T+1)+1+K+1}$ to the system of equations

$$A \cdot u = 0,$$

and for which the last component of u is strictly positive.

Let A be a matrix with $K \times (T+1) + 1 + K + 1$ columns, defined as follows: we have one row for every pair (k, t), one column for every pair (k, t), one column for each k, and two additional columns. Organize the columns so that we first have the $K \times (T+1)$ columns for the pairs (k, t), then one of the single columns mentioned in last place, which we shall refer to as the δ-column, then K columns (one for each k), and finally one last column. In the row corresponding to (k, t) the matrix has zeroes everywhere with the following exceptions: it has a 1 in the column for (k,t), t in the δ column, -1 in the column for k, and $-\log p_t^k$ in the very last column.

Thus, matrix A looks as follows:

$$\begin{array}{c} \\ \vdots \\ (k,t) \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} (1,0) & \cdots & (k,t) & \cdots & (K,T) & \delta & 1 & \cdots & k & \cdots & K & p \end{array} \\ \left[\begin{array}{ccccc|c|ccccc|c} \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\ \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \end{array}\right] \end{array}.$$

Consider the system $A \cdot u = 0$. If there are numbers solving equation (2.9), then these define a solution $u \in \mathbf{R}^{K\times(T+1)+1+K+1}$ for which the last component is 1. If, on the other hand, there is a solution $u \in \mathbf{R}^{K\times(T+1)+1+K+1}$ to the system $A \cdot u = 0$ in which the last component is strictly positive, then by dividing through by the last component of u we obtain numbers that solve equation (2.9).

In second place, we write the system of inequalities (2.10) and (2.11) in matrix form. Let B be a matrix with $K \times (T+1) + 1 + K + 1$ columns. Define B as follows: one row for every pair (k,t) and (k',t') with $x_t^k > x_{t'}^{k'}$; in the row corresponding to (k,t) and (k',t') we have zeroes everywhere with the exception of a -1 in the column for (k,t) and a 1 in the column for (k',t'). These rows captures the inequality (2.10). Finally, in the last row, we have zero everywhere with the exception of a -1 at $K \times (T+1) + 1$th column. We shall refer to this last row as the δ-row, which is capturing the inequality (2.11).

In third place, we have a matrix E that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and $K \times (T+1) + 1 + K + 1$ columns. It has zeroes everywhere except for 1 in the last column.

To sum up, there is a solution to system (2.9), (2.10), and (2.11) if and only if there is a vector $u \in \mathbf{R}^{K\times(T+1)+1+K+1}$ that solves the system of equations and linear inequalities:

$$(S1): \quad A \cdot u = 0, \;\; B \cdot u \geq 0, \;\; E \cdot u \gg 0.$$

The entries of A, B, and E are integer numbers, with the exception of the last column of A. Under the hypothesis of the lemma we are proving, the last column consists of rational numbers.

By Lemma 4, then, there is such a solution u to $S1$ if and only if there is no rational vector (θ, η, π) that solves the system of equations and linear inequalities:

$$(S2): \quad \theta \cdot A + \eta \cdot B + \pi \cdot E = 0, \quad \eta \geq 0, \quad \pi > 0.$$

In the following, we shall prove that the non-existence of a solution u implies that the data must violate SAR-EDU. Suppose then that there is no solution u and let (θ, η, π) be a rational vector as above, solving system $S2$.

By multiplying (θ, η, π) by any positive integer we obtain new vectors that solve $S2$, so we can take (θ, η, π) to be integer vectors.

Henceforth, we use the following notational convention: for a matrix D with $K \times (T+1) + 1 + K + 1$ columns, write D_1 for the submatrix of D corresponding to the first $K \times (T+1)$ columns, let D_2 be the submatrix corresponding to the following one column (i.e., δ-column), D_3 correspond to the next K columns, and D_4 to the last column. Thus, $D = [D_1|D_2|D_3|D_4]$.

Claim 1. *(i)* $\theta \cdot A_1 + \eta \cdot B_1 = 0$; *(ii)* $\theta \cdot A_2 + \eta \cdot B_2 = 0$; *(iii)* $\theta \cdot A_3 = 0$; *and (iv)* $\theta \cdot A_4 + \pi \cdot E_4 = 0$.

Proof. Since $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$, then $\theta \cdot A_i + \eta \cdot B_i + \pi \cdot E_i = 0$ for all $i = 1, \ldots, 4$. Moreover, since B_3, B_4, E_1, E_2, and E_3 are zero matrices, we obtain the claim. □

For convenience, we transform the matrices A and B using θ and η. We transform the matrices A and B as follows. Let us define a matrix A^* from A by letting A^* have $K \times (T+1) + 1 + K + 1$ columns that consists of the rows as follows: for each row in $r \in A$ (i) have θ_r copies of the rth row when $\theta_r > 0$; (ii) omit row r when $\theta_r = 0$; and (iii) have θ_r copies of the rth row multiplied by -1 when $\theta_r < 0$.

We refer to rows that are copies of some r in A with $\theta_r > 0$ as *original* rows. We refer to rows that are copies of some r in A with $\theta_r < 0$ as *converted* rows.

Similarly, we define the matrix B^* from B by including the same columns as B and η_r copies of each row (and thus omitting row r when $\eta_r = 0$; recall that $\eta_r \geq 0$ for all r).

Claim 2. *For any (k, t), all the entries in the column for (k, t) in A_1^* are of the same sign.*

Proof. By definition of A, the column for (k, t) will have zero in all its entries with the exception of the row for (k, t). In A^*, for each (k, t), there are three mutually exclusive possibilities: the row for (k, t) in A can (i) not appear in A^*, (ii) it can appear as original, or (iii) it can appear as converted. This shows the claim. □

Claim 3. *There exists a sequence of pairs* $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})_{i=1}^{n^*}$ *that satisfies Condition (1) in SAR-EDU.*

Proof. We define such a sequence by induction. Let $B^1 = B^*$. Given B^i, define B^{i+1} as follows.

Denote by $>^i$ the binary relation on $\mathcal{X}$ defined by $z >^i z'$ if $z > z'$ and there is at least one pair (k,t) and (k',t') for which (i) $x^k_t > x^{k'}_{t'}$, (ii) $z = x^k_t$ and $z' = x^{k'}_{t'}$, and (iii) the row corresponding $x^k_t > x^{k'}_{t'}$ in B has strictly positive weight in η.

The binary relation $>^i$ cannot exhibit cycles because $>^i \subseteq >$. There is therefore at least one sequence $z^i_1, \ldots, z^i_{L_i}$ in $\mathcal{X}$ such that $z^i_j >^i z^i_{j+1}$ for all $j = 1, \ldots, L_i - 1$ and with the property that there is no $z \in \mathcal{X}$ with $z >^i z^i_1$ or $z^i_{L_i} >^i z$.

Observe that B^i has at least one row corresponding to $z^i_j >^i z^i_{j+1}$ for all $j = 1, \ldots, L_i - 1$. Let the matrix B^{i+1} be defined as the matrix obtained from B^i by omitting one copy of the row corresponding to $z^i_j > z^i_{j+1}$, for all $j = 1, \ldots L_i - 1$.

The matrix B^{i+1} has strictly fewer rows than B^i. There is therefore n^* for which B^{n^*+1} either has no more rows, or $B^{n^*+1}_1$ has only zeroes in all its entries (its rows are copies of the δ-row which has only zeroes in its first $K \times (T+1)$ columns).

Define a sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})_{i=1}^{n^*}$ by letting $x^{k_i}_{t_i} = z^i_1$ and $x^{k'_i}_{t'_i} = z^i_{L_i}$. Note that, as a result, $x^{k_i}_{t_i} > x^{k'_i}_{t'_i}$ for all i. Therefore the sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})_{i=1}^{n^*}$ satisfies Condition (1) in SAR-EDU. □

We shall use the sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})_{i=1}^{n^*}$ as our candidate violation of SAR-EDU.

Consider a sequence of matrices A^i, $i = 1, \ldots, n^*$ defined as follows. Let $A^1 = A^*$, $B^1 = B^*$, and $C^1 = \begin{bmatrix} A^1 \\ B^1 \end{bmatrix}$. Observe that the rows of C^1 add to the null vector by Claim 1.

We shall proceed by induction. Suppose that A^i has been defined, and that the rows of $C^i = \begin{bmatrix} A^i \\ B^i \end{bmatrix}$ add to the null vector.

Recall the definition of the sequence

$$x^{k_i}_{t_i} = z^i_1 > \cdots > z^i_{L_i} = x^{k'_i}_{t'_i}.$$

There is no $z \in \mathcal{X}$ with $z >^i z^i_1$ or $z^i_{L_i} >^i z$, so in order for the rows of C^i to add to zero there must be a -1 in A^i_1 in the column corresponding to (k'_i, t'_i) and a 1 in A^i_1 in the column corresponding to (k_i, t_i). Let r_i be a row in A^i corresponding to (k_i, t_i), and r'_i be a row corresponding to (k'_i, t'_i). The existence of a -1 in A^i_1 in the column corresponding to (k'_i, t'_i), and a 1 in A^i_1 in the column corresponding to

(k_i, t_i), ensures that r_i and r'_i exist. Note that the row r'_i is a converted row while r_i is original. Let A^{i+1} be defined from A^i by deleting the two rows, r_i and r'_i.

Claim 4. *The sum of r_i, r'_i, and the rows of B^i which are deleted when forming B^{i+1} (corresponding to the pairs $z^i_j > z^i_{j+1}$, $j = 1, \ldots, L_i - 1$) add to the null vector.*

Proof. Recall that $z^i_j >^i z^i_{j+1}$ for all $j = 1, \ldots, L_i - 1$. So when we add the rows corresponding to $z^i_j >^i z^i_{j+1}$ and $z^i_{j+1} >^i z^i_{j+2}$, then the entries in the column for (k, t) with $x^k_t = z^i_{j+1}$ cancel out and the sum is zero in that entry. Thus, when we add the rows of B^i that are not in B^{i+1} we obtain a vector that is zero everywhere except the columns corresponding to z^i_1 and $z^i_{L_i}$. This vector cancels out with $r_i + r'_i$, by definition of r_i and r'_i. □

Claim 5. *The matrix A^* can be partitioned into pairs (r_i, r'_i), in which the rows r'_i are converted and the rows r_i are original.*

Proof. For each i, A^{i+1} differs from A^i in that the rows r_i and r'_i are removed from A^i to form A^{i+1}. We shall prove that A^* is composed of the $2n^*$ rows r_i and r'_i.

First note that since the rows of C^i add up to the null vector, and A^{i+1} and B^{i+1} are obtained from A^i and B^i by removing a collection of rows that add up to zero, then the rows of C^{i+1} must add up to zero as well.

By way of contradiction, suppose that there exist rows left after removing r_{n^*} and r'_{n^*}. Then, by the argument above, the rows of the matrix C^{n^*+1} must add to the null vector. If there are rows left, then the matrix C^{n^*+1} is well defined.

By definition of the sequence B^i, however, B^{n^*+1} has all its entries equal to zero, or has no rows. Therefore, the rows remaining in $A_1^{n^*+1}$ must add up to zero. By Claim 2, the entries of a column (k, t) of A^* are always of the same sign. Moreover, each row of A^* has a non-zero element in the first $K \times (T + 1)$ columns. Therefore, no subset of the columns of A^*_1 can sum to the null vector. □

Claim 6. *(i) For any k and t, if $(k_i, t_i) = (k, t)$ for some i, then the row r_i corresponding to (k, t) appears as original in A^*. Similarly, if $(k'_i, t'_i) = (k', t')$ for some i, then the row corresponding to (k, t) appears converted in A^*. (ii) If the row corresponding to (k, t) appears as original in A^*, then there is some i with $(k_i, t_i) = (k, t)$. Similarly, if the row corresponding to (k, t) appears converted in A^*, then there is i with $(k'_i, t'_i) = (k, t)$.*

Proof. (i) is true by definition of $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})$. (ii) is immediate from Claim 5 because if the row corresponding to (k, t) appears as original in A^* then it equals r_i for some i, and then $x^k_t = x^{k_i}_{t_i}$. Similarly when the row appears converted. □

Claim 7. *The sequence* $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^{n^*}$ *satisfies conditions (2) and (3) in SAR-EDU.*

Proof. We first establish condition (2). Note that A_2^* is a vector, and in row r the entry of A_2^* is as follows. There must be a raw (k, t) in A of which the raw r is a copy. Hence, the component at the row r of A_2^* is t if r is original and $-t$ if r is converted. Now, by the construction of the sequence when r appears as original there is some i for which $t = t_i$, when r appears as converted there is some i for which $t = t_i'$. So for each r there is i such that $(A_4^*)_r$ is either t_i or $-t_i'$. By Claim 1 (ii), $\theta \cdot A_2 + \eta \cdot B_2 = 0$. Recall that $\theta \cdot A_2$ equals the sum of the rows of A_2^*. Moreover, B_2 is a vector that has zeroes everywhere except a -1 in the δ row (i.e., $K \times (T+1) + 1$th row). Therefore, the sum of the rows of A_2^* equals $\eta_{K\times(T+1)+1}$, where $\eta_{K\times(T+1)+1}$ is the $K \times (T+1) + 1$th element of η. Since $\eta \geq 0$, therefore, $\sum_{i=1}^{n^*} t_i \geq \sum_{i=1}^{n^*} t_i'$, and condition (2) in the axiom is satisfied.

Now we turn to condition (3). By Claim 1 (iii), the rows of A_3^* add up to zero. Therefore, the number of times that k appears in an original row equals the number of times that it appears in a converted row. By Claim 6, then, the number of times k appears as k_i equals the number of times it appears as k_i'. Therefore, condition (3) in the axiom is satisfied. □

Finally, in the following, we show that $\prod_{i=1}^{n^*} p_{t_i}^{k_i} / p_{t_i'}^{k_i'} > 1$, which finishes the proof of Lemma 5 as the sequence $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^{n^*}$ would then exhibit a violation of SAR-EDU.

Claim 8. $\prod_{i=1}^{n^*} p_{t_i}^{k_i} / p_{t_i'}^{k_i'} > 1$.

Proof. By Claim 1 (iv) and the fact that the submatrix E_4 equals the scalar 1, we obtain

$$0 = \theta \cdot A_4 + \pi E_4 = \left(\sum_{i=1}^{n^*} (r_i + r_i') \right)_4 + \pi,$$

where $(\sum_{i=1}^{n^*} (r_i + r_i'))_4$ is the (scalar) sum of the entries of A_4^*. Recall that $-\log p_{t_i}^{k_i}$ is the last entry of row r_i and that $\log p_{t_i'}^{k_i'}$ is the last entry of row r_i', as r_i' is converted and r_i original. Therefore the sum of the rows of A_4^* are $\sum_{i=1}^{n^*} \log(p_{t_i'}^{k_i'} / p_{t_i}^{k_i})$. Then,

$$\sum_{i=1}^{n^*} \log(p_{t_i'}^{k_i'} / p_{t_i}^{k_i}) = -\pi < 0.$$

Thus $\prod_{i=1}^{n^*} p_{t_i}^{k_i} / p_{t_i'}^{k_i'} > 1$. □

2.6.5 Proof of Lemma 6

For each sequence $\sigma = (x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^n_{i=1}$ that satisfies conditions (1), (2), and (3) in SAR-EDU, we define a vector $t_\sigma \in \mathbf{N}^{(K\times T)^2}$ as follows. To make the notation easier, we identify the pair $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})$ with $((k_i,t_i),(k'_i,t'_i))$. Let $t_\sigma((k,t),(k',t'))$ be the number of times that the pair $(x^k_t, x^{k'}_{t'})$ appears in the sequence σ. One can then describe the satisfaction of SAR-EDU by means of the vectors t_σ. Define

$$T = \left\{ t_\sigma \in \mathbf{N}^{(K\times T)^2} : \sigma \text{ satisfies conditions (1), (2), (3) in SAR-EDU} \right\}.$$

Observe that the set T depends only on $(x^k)^K_{k=1}$ in the dataset $(x^k, p^k)^K_{k=1}$. It does not depend on prices.

For each $((k,t),(k',t')) \in (K\times T)^2$ such that $x^k_t > x^{k'}_{t'}$, define

$$\widehat{\gamma}((k,t),(k',t')) = \log\left(\frac{p^k_t}{p^{k'}_{t'}}\right),$$

and define $\widehat{\gamma}((k,t),(k',t')) = 0$ when $x^k_t \le x^{k'}_{t'}$. Then, $\widehat{\gamma}$ is a $(KT)^2$-dimensional real-valued vector. If $\sigma = (x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^n_{i=1}$, then

$$\widehat{\gamma}\cdot t_\sigma = \sum_{((k,t),(k',t'))\in(K\times T)^2} \widehat{\gamma}((k,t),(k',t'))t_\sigma((k,t),(k',t')) = \log\left(\prod_{i=1}^n \frac{p^{k_i}_{t_i}}{p^{k'_i}_{t'_i}}\right).$$

Thus, the data satisfy SAR-EDU if and only if $\widehat{\gamma}\cdot t \le 0$ for all $t \in T$.

Enumerate the elements in $\mathcal{X}$ in increasing order:

$$y_1 < y_2 < \cdots < y_N.$$

Fix an arbitrary $\underline{\xi} \in (0,1)$.

We shall construct by induction a sequence $(\varepsilon^k_t(n))$ for $n = 1,\ldots,N$, where $\varepsilon^k_t(n)$ is defined for all (k,t) with $x^k_t = y_n$.

By the denseness of the rational numbers, and the continuity of the exponential function, for each (k,t) such that $x^k_t = y_1$, there exists a positive number $\varepsilon^k_t(1)$ such that $\log(p^k_t\varepsilon^k_t(1)) \in \mathbf{Q}$ and $\underline{\xi} < \varepsilon^k_t(1) < 1$. Let $\varepsilon(1) = \min\{\varepsilon^k_t(1) : x^k_t = y_1\}$.

In second place, for each (k,t) such that $x^k_t = y_2$, there exists a positive $\varepsilon^k_t(2)$ such that $\log(p^k_t\varepsilon^k_t(2)) \in \mathbf{Q}$ and $\underline{\xi} < \varepsilon^k_t(2) < \varepsilon(1)$. Let $\varepsilon(2) = \min\{\varepsilon^k_t(2) : x^k_t = y_2\}$.

In third place, and reasoning by induction, suppose that $\varepsilon(n)$ has been defined and that $\underline{\xi} < \varepsilon(n)$. For each (k,t) such that $x_t^k = y_{n+1}$, let $\varepsilon_t^k(n+1) > 0$ be such that $\log(p_t^k \varepsilon_t^k(n+1)) \in \mathbf{Q}$, and $\underline{\xi} < \varepsilon_t^k(n+1) < \varepsilon(n)$. Let $\varepsilon(n+1) = \min\{\varepsilon_t^k(n+1) : x_t^k = y_n\}$.

This defines the sequence $(\varepsilon_t^k(n))$ by induction. Note that $\varepsilon_t^k(n+1)/\varepsilon(n) < 1$ for all n. Let $\bar{\xi} < 1$ be such that $\varepsilon_t^k(n+1)/\varepsilon(n) < \bar{\xi}$.

For each $k \in K$ and $t \in T$, let $q_t^k = p_t^k \varepsilon_t^k(n)$, where n is such that $x_t^k = y_n$. We claim that the data $(x^k, q^k)_{k=1}^K$ satisfy SAR-EDU. Let γ^* be defined from $(q^k)_{k=1}^K$ in the same manner as $\widehat{\gamma}$ was defined from $(p^k)_{k=1}^K$.

For each pair $((k,t),(k',t'))$ with $x_t^k > x_{t'}^{k'}$, if n and m are such that $x_t^k = y_n$ and $x_{t'}^{k'} = y_m$, then $n > m$. By the definition of ε,

$$\frac{\varepsilon_t^k(n)}{\varepsilon_{t'}^{k'}(m)} < \frac{\varepsilon_t^k(n)}{\varepsilon(m)} < \bar{\xi} < 1.$$

Therefore,

$$\gamma^*((k,t),(k',t')) = \log \frac{p_t^k \varepsilon_t^k(n)}{p_{t'}^{k'} \varepsilon_{t'}^{k'}(m)} < \log \frac{p_t^k}{p_{t'}^{k'}} + \log \bar{\xi} < \log \frac{p_t^k}{p_{t'}^{k'}} = \widehat{\gamma}(x_s^k, x_{t'}^{k'}).$$

Thus, for all $t \in T$, $\gamma^* \cdot t \leq \widehat{\gamma} \cdot t \leq 0$, as $t \geq 0$ and the data $(x^k, p^k)_{k=1}^K$ satisfy SAR-EDU. Thus the data $(x^k, q^k)_{k=1}^K$ satisfy SAR-EDU. Finally, note that $\underline{\xi} < \varepsilon_t^k(n) < 1$ for all n and each $k \in K, t \in T$. So that by choosing $\underline{\xi}$ close enough to 1 we can take the prices (q^k) to be as close to (p^k) as desired.

2.6.6 Proof of Lemma 7

Consider the system comprised by (2.9), (2.10), and (2.11) in the proof of Lemma 5. Let A, B, and E be constructed from the data as in the proof of Lemma 5. The difference with respect to Lemma 5 is that now the entries of A_4 may not be rational. Note that the entries of E, B, and A_i, $i = 1,2,3$ are rational.

Suppose, towards a contradiction, that there is no solution to the system comprised by (2.9), (2.10), and (2.11). Then, by the argument in the proof of Lemma 5 there is no solution to system $S1$. Lemma 3 with $\mathbf{F} = \mathbf{R}$ implies that there is a real vector (θ, η, π) such that

$$\theta \cdot A + \eta \cdot B + \pi \cdot E = 0 \text{ and } \eta \geq 0, \pi > 0.$$

Recall that $B_4 = 0$ and $E_4 = 1$, so we obtain that $\theta \cdot A_4 + \pi = 0$.

Let $(q^k)_{k=1}^K$ be vectors of prices such that the dataset $(x^k, q^k)_{k=1}^K$ satisfies SAR-

EDU and $\log q_t^k \in \mathbf{Q}$ for all k and t. (Such $(q^k)_{k=1}^K$ exists by Lemma 6.) Construct matrices A', B', and E' from this dataset in the same way as A, B, and E is constructed in the proof of Lemma 5. Note that only the prices are different in (x^k, q^k) compared to (x^k, p^k). So $E' = E$, $B' = B$ and $A'_i = A_i$ for $i = 1, 2, 3$. Since only prices q^k are different in this dataset, only A'_4 may be different from A_4.

By Lemma 6, we can choose prices q^k such that $|\theta \cdot A'_4 - \theta \cdot A_4| < \pi/2$. We have shown that $\theta \cdot A_4 = -\pi$, so the choice of prices q^k guarantees that $\theta \cdot A'_4 < 0$. Let $\pi' = -\theta \cdot A'_4 > 0$.

Note that $\theta \cdot A'_i + \eta \cdot B'_i + \pi' E_i = 0$ for $i = 1, 2, 3$, as (θ, η, π) solves system $S2$ for matrices A, B, and E, and $A'_i = A_i$, $B'_i = B_i$, and $E_i = 0$ for $i = 1, 2, 3$. Finally, $B_4 = 0$ so

$$\theta \cdot A'_4 + \eta \cdot B'_4 + \pi' E_4 = \theta \cdot A'_4 + \pi' = 0.$$

We also have that $\eta \geq 0$ and $\pi' > 0$. Therefore θ, η, and π' constitute a solution $S2$ for matrices A', B', and E'.

Lemma 3 then implies that there is no solution to $S1$ for matrices A', B', and E'. So there is no solution to the system comprised by (2.9), (2.10), and (2.11) in the proof of Lemma 5. However, this contradicts Lemma 5 because the data (x^k, q^k) satisfies SAR-EDU and $\log q_t^k \in \mathbf{Q}$ for all $k = 1, \ldots, K$ and $t = 1, \ldots, T$.

2.7 Proof of Theorem 2

The proofs for QHD and PQHD are similar, so we give a detailed proof for PQHD and then explain how the proof for QHD is different.

Lemma 8. *Let $(x^k, p^k)_{k=1}^K$ be a dataset. The following statements are equivalent:*

1. *$(x^k, p^k)_{k=1}^K$ is PQHD rational.*
2. *There are strictly positive numbers v_t^k, λ^k, $\beta \leq 1$, and $\delta \in (0, 1]$, for $t = 0, \ldots, T$ and $k = 1, \ldots, K$, such that*

$$v_t^k = \lambda^k p_t^k \text{ if } t = 0,\ \beta \delta^t v_t^k = \lambda^k p_t^k \text{ if } t > 0, \text{ and } x_t^k > x_{t'}^{k'} \Longrightarrow v_t^k \leq v_{t'}^{k'}.$$

The proof of Lemma 8 is very similar to the proof of Lemma 1 and omitted.

2.7.1 Necessity

Lemma 9. *If a dataset $(x^k, p^k)_{k=1}^K$ is PQHD rational, then it satisfies SAR-PQHD.*

Proof. Let $(x^k, p^k)_{k=1}^K$ be PQHD rational, and let $\beta \leq 1$, $\delta \in (0,1]$, and $u : \mathbf{R}_+ \to \mathbf{R}$ be as in the definition of PQHD rational. By Lemma 8, there exists a strictly positive solution v_t^k, λ^k, β, δ to the system in statement (2) of Lemma 8 with $v_t^k \in \partial u(x_t^k)$ when $x_t^k > 0$, and $v_t^k \geq \underline{w} \in \partial u(x_t^k)$ when $x_t^k = 0$. Moreover, $v_t^k = \lambda^k p_t^k / D(t)$, where $D(t) = 1$ if $t = 0$ and $D(t) = \beta\delta^t$ if $t > 0$.

Let $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$ be a sequence satisfying the four conditions in SAR-PQHD. Then $x_{t_i}^{k_i} > x_{t_i'}^{k_i'}$. Suppose that $x_{t_i'}^{k_i'} > 0$. Then, $v_{t_i}^{k_i} \in \partial u(x_{t_i}^{k_i})$ and $v_{t_i'}^{k_i'} \in \partial u(x_{t_i'}^{k_i'})$. By the concavity of u, it follows that $v_{t_i}^{k_i} \leq v_{t_i'}^{k_i'}$. Similarly, if $x_{t_i'}^{k_i'} = 0$, then $v_{t_i}^{k_i} \in \partial u(x_{t_i}^{k_i})$ and $v_{t_i'}^{k_i'} \geq \underline{w} \in \partial u(x_{t_i'}^{k_i'})$, so that $v_{t_i}^{k_i} \leq v_{t_i'}^{k_i'}$. Therefore,

$$1 \geq \prod_{i=1}^{n} \frac{\lambda^{k_i} D(t_i') p_{t_i}^{k_i}}{\lambda^{k_i'} D(t_i) p_{t_i'}^{k_i'}} = \prod_{i=1}^{n} \frac{D(t_i') p_{t_i}^{k_i}}{D(t_i) p_{t_i'}^{k_i'}} = \frac{\beta^{\#\{i:t_i'>0\} - \#\{i:t_i>0\}}}{\delta^{(\sum t_i - \sum t_i')}} \prod_{i=1}^{n} \frac{p_{t_i}^{k_i}}{p_{t_i'}^{k_i'}} \geq \prod_{i=1}^{n} \frac{p_{t_i}^{k_i}}{p_{t_i'}^{k_i'}},$$

where the first equality holds by (4) of SAR-PQHD, and the numbers λ^k appear the same number of times in the denominator as in the numerator of this product. Moreover, the last inequality holds by (2) and (3) of SAR-PQHD. □

2.7.2 Sufficiency

Lemma 10. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-PQHD. Suppose that $\log(p_t^k) \in \mathbf{Q}$ for all k and t. Then there are numbers v_t^k, λ^k, β, δ, for $t \in T$ and $k \in K$ satisfying* (2) *in Lemma 8.*

Lemma 11. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-PQHD. Then for all positive numbers $\bar{\varepsilon}$, there exists $q_t^k \in [p_t^k - \bar{\varepsilon}, p_t^k]$ for all $t \in T$ and $k \in K$ such that $\log q_t^k \in \mathbf{Q}$ and the dataset $(x^k, q^k)_{k=1}^k$ satisfy SAR-PQHD.*

Lemma 12. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-PQHD. Then there are numbers v_t^k, λ^k, β, δ, for $t \in T$ and $k \in K$ satisfying* (2) *in Lemma 8.*

Lemma 11 and 12 hold as in the proof for Theorem 1.

2.7.3 Proof of Lemma 10

We linearize the equation in system (2) of Lemma 8. The result is:

$$\log v(x_t^k) - \log \lambda^k - \log p_t^k = 0 \text{ if } t = 0, \tag{2.12}$$

$$\log v(x_t^k) + \log \beta + t \log \delta - \log \lambda^k - \log p_t^k = 0 \text{ if } t > 0, \tag{2.13}$$

$$x > x' \Longrightarrow \log v(x') \geq \log v(x), \tag{2.14}$$

$$\log \beta \geq 0, \tag{2.15}$$

$$\log \delta \leq 0. \tag{2.16}$$

In the system comprised by (2.12), (2.13), (2.14), (2.15), and (2.16), the unknowns are the real numbers $\log \beta$, $\log \delta$, $\log \lambda^k$, and $\log v_t^k$ for all $k = 1, \ldots, K$ and $t = 1, \ldots, T$.

First, we are going to write the system of inequalities from (2.12) to (2.16) in matrix form.

We shall define a matrix A such that there are positive numbers v_t^k, λ^k, β, and δ the logs of which satisfy equations (2.12) and (2.13) if and only if there is a solution $u \in \mathbf{R}^{K \times (T+1)+2+K+1}$ to the system of equations

$$A \cdot u = 0,$$

and for which the last component of u is strictly positive.

Let A be a matrix with $K \times (T+1)$ rows and $K \times (T+1) + 2 + K + 1$ columns, defined as follows: we have one row for every pair (k, t), one column for every pair (k, t), two columns for each k, and two additional columns. Organize the columns so that we first have the $K \times (T+1)$ columns for the pairs (k, t), then two columns, which we shall refer to as the β-column and δ-column, respectively, then K columns (one for each k), and finally one last column. In the row corresponding to (k, t) the matrix has zeroes everywhere with the following exceptions: it has a 1 in the column for (k, t), it has a 1 if $t > 0$ and it has a 0 if $t = 0$ in the β-column, it has t in the δ-column, it has a -1 in the column for k, and $-\log p_t^k$ in the very last column.

Thus, matrix A looks as follows:

$$
\begin{array}{c}
\\ \vdots \\ (k,t{=}0) \\ (k,t'{>}0) \\ \vdots
\end{array}
\begin{array}{c}
\begin{array}{cccccc|c|c|ccccc|c}
(1,1) & \cdots & (k,t) & (k,t') & \cdots & (K,T) & \beta & \delta & 1 & \cdots & k & \cdots & K & p
\end{array}
\\
\left[\begin{array}{cccccc|c|c|ccccc|c}
\vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & 1 & 0 & \cdots & 0 & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\
0 & \cdots & 0 & 1 & \cdots & 0 & 1 & t' & 0 & \cdots & -1 & \cdots & 0 & -\log p_{t'}^k \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots
\end{array}\right]
\end{array}.
$$

Consider the system $A \cdot u = 0$. If there are numbers solving equations (2.12) and (2.13), then these define a solution $u \in \mathbf{R}^{K\times(T+1)+2+K+1}$ for which the last component is 1. If, on the other hand, there is a solution $u \in \mathbf{R}^{K\times(T+1)+2+K+1}$ to the system $A \cdot u = 0$ in which the last component is strictly positive, then by dividing through by the last component of u we obtain numbers that solve equation (2.12) and (2.13).

In second place, we write the system of inequalities (2.14), (2.15), and (2.16) in matrix form. Let B be a matrix with $K \times (T+1) + 2 + K + 1$ columns. Define B as follows: one row for every pair (k,t) and (k',t') with $x_t^k > x_{t'}^{k'}$; in the row corresponding to (k,t) and (k',t') we have zeroes everywhere with the exception of a -1 in the column for (k,t) and a 1 in the column for (k',t'). Finally, we have the last two rows, where we have zeroes everywhere with one exception. In the first row, we have a -1 at $(K \times (T+1) + 1)$-th column; in the second row, we have a -1 at $(K \times (T+1) + 2)$-th column. We shall refer to the first row as the β-row, which captures (2.15). We also shall refer to the second row as the δ-row, which captures (2.16). For (general) QHD, we do not have a β-row.

In third place, we have a matrix E that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and $K \times (T+1) + 2 + K + 1$ columns. It has zeroes everywhere except for 1 in the last column.

To sum up, there is a solution to system (2.12), (2.13), (2.14), (2.15), and (2.16) if and only if there is a vector $u \in \mathbf{R}^{K\times(T+1)+2+K+1}$ that solves the system of equations and linear inequalities:

$$(S1): \; A \cdot u = 0, \; B \cdot u \geq 0, \; E \cdot u \gg 0.$$

The argument now follow along the lines of the proof of Theorem 1. Suppose that there is no solution u and let (θ, η, π) be an integer vector solving system:

$$(S2): \; \theta \cdot A + \eta \cdot B + \pi \cdot E = 0, \; \eta \geq 0, \; \pi > 0.$$

The following has the same proof as Claim 1.

Claim 9. *(i) $\theta \cdot A_1 + \eta \cdot B_1 = 0$; (ii) $\theta \cdot A_2 + \eta \cdot B_2 = 0$; (iii) $\theta \cdot A_3 + \eta \cdot B_3 = 0$; (iv) $\theta \cdot A_4 = 0$; and (v) $\theta \cdot A_5 + \pi \cdot E_5 = 0$.*

We transform the matrices A and B based on the values of θ and η, as we did in the proof of Theorem 1. Let us define a matrix A^* from A and B^* from B, as we did in the proof of Theorem 1. We can prove the same claims (i.e., Claims 2, 3, 4, 5, and 6) as in the proof of Theorem 1. The proofs are the same and omitted. In particular, we can show that there exists a sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ that satisfies (1) in SAR-PQHD. We shall use the sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ as our candidate violation of SAR-PQHD.

Claim 10. *The sequence $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ satisfies (2), (3), and (4) in SAR-PQHD.*

Proof. We first establish (2). Note that A^*_3 is a vector, and in row r the entry of A^*_3 is as follows. There must be a (k,t) of which r is a copy. Then the component at row r of A^*_3 is t if r is original and $-t$ if r is converted. Now, when r appears as original there is some i for which $t = t_i$, when r appears as converted there is some i for which $t = t'_i$. So for each r there is i such that $(A^*_3)_r$ is either t_i or $-t'_i$.

By Claim 9 (iii), $\theta \cdot A_3 + \eta \cdot B_3 = 0$. Recall that $\theta \cdot A_3$ equals the sum of the rows of A^*_3. Moreover, B_3 is a vector that has zeroes everywhere except a -1 in the δ row (i.e., $K \times (T+1) + 2$th row). Therefore, the sum of the rows of A^*_3 equals $\eta_{K\times(T+1)+2}$, where $\eta_{K\times(T+1)+2}$ is the $K \times (T+1) + 2$th element of η. Since $\eta \geq 0$, therefore, $\sum_{i:t_i>0} t_i - \sum_{i:t'_i>0} t'_i = \eta_{K\times(T+1)+2} \geq 0$, and condition (2) in the axiom is satisfied.

Next, we show (3). By Claim 9 (ii), $\theta \cdot A_2 + \eta \cdot B_2 = 0$. Recall that $\theta \cdot A_2$ equals the sum of the rows of A^*_2. Moreover, B_2 is a vector that has zeroes everywhere except a -1 in the β-row (i.e., $K \times (T+1) + 1$th row). Therefore, the sum of the rows of A^*_2 equals $\eta_{K\times(T+1)+1}$, where $\eta_{K\times(T+1)+1}$ is the $K \times (T+1) + 1$th element of η. Since $\eta \geq 0$, therefore, $\#\{i : t_i > 0\} - \#\{i : t'_i > 0\} = \eta_{K\times(T+1)+1} \geq 0$, and condition (3) in the axiom is satisfied. (For (general) QHD, B_2 is a zero vector in the β-row (i.e., $K \times (T+1) + 1$th row). Therefore, $\#\{i : t_i > 0\} - \#\{i : t'_i > 0\} = 0$, and condition (3) in SAR-QHD is satisfied.)

Now we turn to (4). By Claim 9 (iv), the rows of A^*_4 add up to zero. Therefore, the number of times that k appears in an original row equals the number of times that it appears in a converted row. By Claim 6, then, the number of times k appears as k_i equals the number of times it appears as k'_i. Therefore, condition (4) in the axiom is satisfied. □

Finally, we can show that $\prod_{i=1}^{n^*} p_{t_i}^{k_i} / p_{t_i'}^{k_i'} > 1$, which completes the proof of Lemma 5 as the sequence $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^{n^*}$ would then exhibit a violation of SAR-PQHD. The proof is the same as that of the corresponding lemma in the proof of Theorem 1.

Chapter 3

When the Eyes Say Buy: Visual Fixations during Hypothetical Consumer Choice Improve Prediction of Actual Purchases

3.1 Introduction

Real choices are binding consequential commitments to a course of action, like accepting a job or voting in an election. However, scientists and policy makers who are interested in real choices often rely on hypothetical statements about what people *would* choose, rather than what they *do* actually choose. Measurement of hypothetical choice is common in many fields, and is usually done for practical reasons. Examples include pre-election polling in politics (e.g., Gallup presidential election poll), marketing surveys of potential new products to forecast sales (Chandon et al., 2004; Green and Srinivasan, 1990; Infosino, 1986; Jamieson and Bass, 1989; Raghubir and Greenleaf, 2006; Schlosser et al., 2006; Silk and Urban, 1978; Urban et al., 1983), artificial choices about moral dilemmas or measurement of "sacred" values, which cannot be actually enforced for ethical reasons (Berns et al., 2012; FeldmanHall et al., 2012b; Greene et al., 2004, 2001; Hariri et al., 2006; Kühberger et al., 2002; Monterosso et al., 2007), eliciting quality-adjusted-life-years (QALY) to choose medical procedures (Cutler et al., 1997; Garber and Phelps, 1997; Gold et al., 1996; Zeckhauser and Shepard, 1976), and surveys used to estimate dollar value of goods that are not traded in markets (such as clean air or the prevention of oil spills) for cost-benefit analysis (Carson, 2012; Carson and Hanemann, 2005; Shogren, 2005, 2006).

The maintained assumption in all these research areas is that hypothetical choices offer some useful relation to real choice. However, many comparisons show that hypothetical and real choices can differ systematically. The differences are collectively called "hypothetical bias." Typically, it is an upward "Yes bias": people overstate their intentions to buy new products and vote, compared to actual rates of purchase and voting (Blumenschein et al., 2008; Bohm, 1972; Cummings et al., 1995; Johannesson et al., 1998; List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005). A small number of brain imaging studies have found common valuation regions (Kang et al., 2011) or emotion regions (Feldman-Hall et al., 2012a) for both types of choice, as well as distinct regions which are more strongly activated during real choice (Kang and Camerer, 2013).

Given the possibility of hypothetical bias, an important practical challenge is how to accurately forecast real choices from data on hypothetical choices. A good forecasting correction method is also likely to be scientifically valuable, if it can create knowledge about the detailed mechanism that produces bias (and how it varies across types of choices and people).

Different forecasting methods have been tried. Methods can be sorted into two categories: procedures and enhanced pre-choice measurement. Procedural approaches change how questions are asked or choice data are processed. Measurement methods collect more data and use them to improve forecasting.

Procedures Many studies have explored different experimental or statistical procedures that might reduce the bias. Statistical procedures ("calibration") search for a predictable measurable relations between the hypothetical and real choices, and then test how well that relation can be used to forecast actual choices from hypothetical ones within-sample, or in a new case (Blackburn et al., 1994; Fox et al., 1998; Kurz, 1974; List and Shogren, 1998, 2002; Shogren, 1993). For example, in our study we observe that about 55% of subjects choose to purchase consumer goods hypothetically, but only 23% do when choices are real. So one could take a hypothetical purchase rate in a new sample, and multiply it by $0.23/0.55 = 0.40$, to crudely estimate a real purchase rate.

Calibration methods such as these have been extended to account for socio-demographics variables in hypothetical bias. They are useful for many purposes. However, calibration has not been well-tuned to adjust for likely vagaries of specific goods and choice contexts, as flagged by List and Shogren (1998, 2002).

A more ambitious procedure is to search for a way of asking hypothetical questions that gives answers which are closer to real-choice answers. Champ et al. (1997) ask respondents how "certain" they are (on a 10-point scale) about whether they would actually donate the stated amount to a project if asked to do so. Cummings and Taylor (1999) use the "cheap talk" protocol: the design includes an explicit discussion of the hypothetical bias problem (what it is and why it might occur) at the beginning of the experiment. Following findings in social psychology, Jacquemet et al. (2013) use a "solemn oath," asking participants to swear on their honor to give honest answers, as a truth-telling commitment device. Finally, the "dissonance-minimizing format" of Blamey et al. (1999) and Loomis et al. (1999) include additional response categories that permit respondents to express support for a project or policy without having to commit dollars. [1]

Several meta-analyses have been conducted to evaluate effects of diverse experimental methods on hypothetical bias to find variables that account for the variation in bias across goods and contexts (Carson et al., 1996; List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005). Hypothetical bias is influenced by the distinction between willingness-to-pay or willingness-to-accept, public goods and private goods, and elicitation methods.

All calibration methods also rely on extrapolating from a past hypothetical-actual relation to the future. An example of where this can backfire is politics. Historically, polls asking people whom they intend to vote for overestimated the actual vote for black candidates on election day (Keeter and Samaranayake, 2007). However, this so-called "Bradley effect" (also known as "Wilder effect") has gradually eroded over time (Hopkins, 2009).

Further search for ideal procedures to pose hypothetical questions that yield responses that predict real answers is surely worthwhile. However, there is no current consensus on a single method that works effectively across choice contexts. We therefore turn to measuring more variables.

Pre-choice measures Another approach that has been explored more tentatively is to measure psychological or neural variables that are recorded during the process of hypothetical choice, and use those measures to forecast actual choice. [2] These measures will often *precede* choice, so we generally call them "pre-choice" measures. We report new evidence from this approach using measures of vi-

[1]Other procedures, such as asking respondents to consider budget constraints and budgetary substitutes, are shown to be ineffective (Loomis et al., 1994; Neill, 1995).

[2]Recent study by Bernheim et al. (2015) has features of both procedural and measurement methods approaches. Their proposed method, called "non-choice revealed preference," involves estimation of statistical relationships between choices and non-choice variables and prediction of choices using those relationships together with non-choice data under new environment. They propose that non-choice reactions could include from simple ratings (liking, familiarity, certainty, happiness, etc.) to physiological reactions including brain activities.

sual attention—both mouse-based lookup of information (mousetracking), and eyetracking recordings. [3]

We record visual attention as people make hypothetical choices (about consumer products, for example). On some trials people choose to (hypothetically) buy the product, and on others they don't buy. On later trials they are surprised by the opportunity to actually buy some of those same products.

The motivating hypothesis is that what people looked at during the initial hypothetical choice will help forecast whether they will stick with their original hypothetical choice, or will change their minds when making a subsequent real choice. [4] A quick preview of the main result is the following: during hypothetical choice, the more people look at prices, and the longer they take to transition from looking to making a choice, the more likely they are to switch a hypothetical "Buy" to a real "Don't buy." The improvement in prediction is not large in magnitude. However, it provides initial evidence that some improvement is possible using pre-choice measures, and further efforts designed at maximizing the improvement in prediction could do much better.

Note that a few recent studies have measured functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) signals and used them to forecast actual choices, (e.g., Levy et al., 2011; Smith et al., 2014; Tusche et al., 2010). While interesting and promising, none of these studies are specifically designed to make the leap from *hypothetical* pre-choice thinking to later *actual* choices. We discuss their methods and compare them to ours in the concluding discussion.

The remaining of the paper is organized as follows. We describe our experimental design in Section 3.2, report the main results in Section 3.3, and discuss our results and conclude in Section 3.4. Additional results and experimental details are reported in Sections B.1 to B.4.

[3] There is only one study directly comparing results from both measures on a common task (Lohse and Johnson, 1996). As in that study, we find that the main regularities are common across both visual fixation measures.

[4] This is partly motivated by the "mind eye hypothesis" which assumes that what a person is looking at indicates what they are currently thinking about or attending to (Just and Carpenter, 1980). Studies have shown a connection between eyetracking patterns and users' decision making processes (Goldberg et al., 2002). Remarkably, eye movements can be even more accurate than conscious recall in predicting whether people have seen a visual stimulus before (and eye movements are associated with hippocampal activity measured by fMRI; Hannula and Ranganath, 2009).

3.2 Experimental Design

3.2.1 Experiment I: Mousetracking Study

Participants Participants were undergraduate or graduate students recruited from subject pools at Caltech. Twenty-eight male subjects participated in Experiment I. Two additional subjects participated, but their data were excluded from the analysis for reasons described below.

Stimuli One hundred and twenty familiar consumer products (e.g., backpack, watch, flash drive) were selected based on popular categories from a pilot test and an earlier study (Kang et al., 2011). [5] The product images were no larger than 320×320 pixels in size. Stimulus presentation and response recording were controlled by MATLAB (MathWorks, Natick, MA), using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Experimental procedures The experiment consists of four blocks: a willingness-to-pay (WTP) reporting block, a hypothetical purchase block, a real purchase block, and a "surprise" real purchase block. Mousetracking was used only in the three purchase blocks, and participants' viewing time was determined from mouse events (e.g., clicks on certain parts of the screen). Subjects were told that they would earn up to $50 for completing the experiment. Detailed instructions for each part were given immediately prior to that part. Therefore, participants were unaware of the existence of two real purchase blocks while they were in the hypothetical purchase block. [6] [7]

In the WTP reporting block, subjects were shown images of the 120 consumer products, one at a time and in random order. They were asked to state a maximum hypothetical WTP for each item, under the restriction that whatever they would buy must be for themselves (i.e., it cannot be gifted or re-sold). In each

[5] For the complete product list, see Table B.3 in Section B.3.

[6] See Instructions in Section B.4.

[7] We intentionally did not counter-balance the order of the hypothetical and real conditions, following the considerations described in Kang et al. (2011) and Kang and Camerer (2013). There might be an ordering effect in which thinking about real choices first would spill over to affect hypothetical choices. On the other hand, the spill-over effect is expected to be minimal, if any, in the hypothetical-then-real order since in the real condition participants have a strong incentive to change or adjust any behavior carried over from previous hypothetical block. In addition, previous studies that used a within-subject design found no evidence for ordering (Cummings et al., 1995; Johannesson et al., 1998). Notice also that hypothetical decision followed by real decision is a natural order for forecasting purposes since, in most applications, hypothetical decision data are gathered in advance of real decisions.

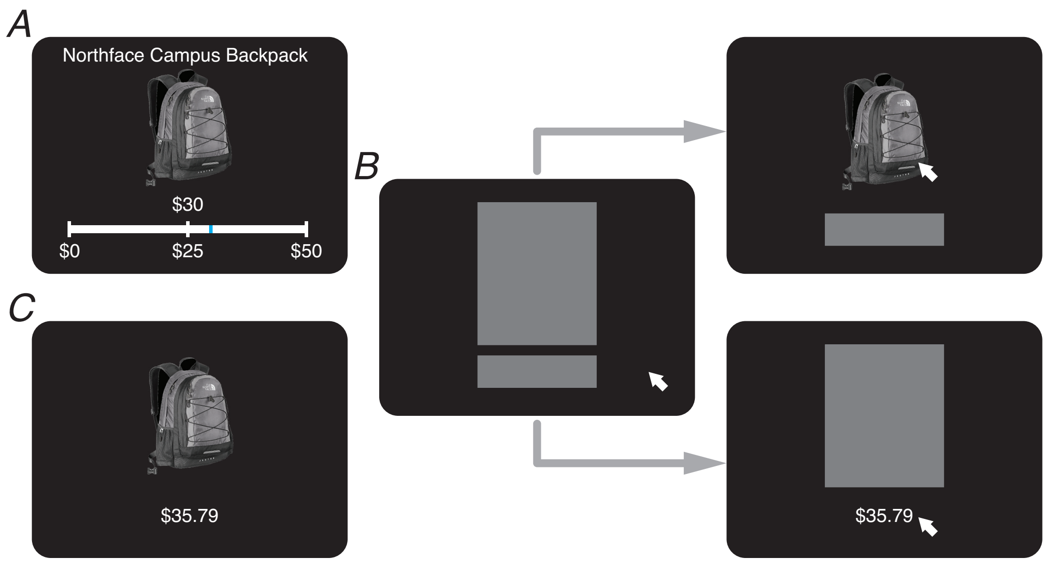


FIGURE 3.1: Example screens for pre-mouse/pre-eyetracking trials (*A*), mousetracking trials (*B*), and eyetracking trials (*C*).

trial, subjects entered an amount between \$0 and \$50 using a sliding scale in \$1 increments (Figure 3.1*A*).

Upon completion of the first part, 60 out of 120 products were selected for each subject, by the computer, for presentation to the subject during the mouse-tracking blocks. More specifically, the computer ranked products in descending order of the subject's WTPs, except for the products with WTP of \$50 (to avoid the ceiling effect), and then paired up each two adjacent products (e.g., {1st, 2nd}, {3rd, 4th}, ...). Among these pairs, the 30 pairs with the highest WTP were selected. One product of each adjacent pair was randomly chosen and assigned to the hypothetical trials, and the other product from each pair was assigned to the real trials. This procedure ensured that the distributions of WTPs in both of hypothetical and real blocks were matched. See Figure B.1 in Section B.3.

In the hypothetical purchase block, subjects were shown a product image with an offer price, one product at a time, and asked to make a hypothetical purchase decision and respond with a Yes or No key press. Each of the 30 products selected in the aforementioned way was presented to the subject three times (for a total of 90 trials in the hypothetical block), and with a different offer price each time. The offer prices for each product were determined as follows: (i) let P_i be the offer price for product i, WTP_i be the WTP for product i, and d be a discounting factor; (ii) sample d from the set $\{0.6+\alpha, 0.9+\alpha, 1.2+\alpha\}$ without replacement for every repeat of product i, where α is a random variable from a uniform distribution over the range $[-0.05, 0.05]$ (i.e., α is to add jitter); and (iii) determine

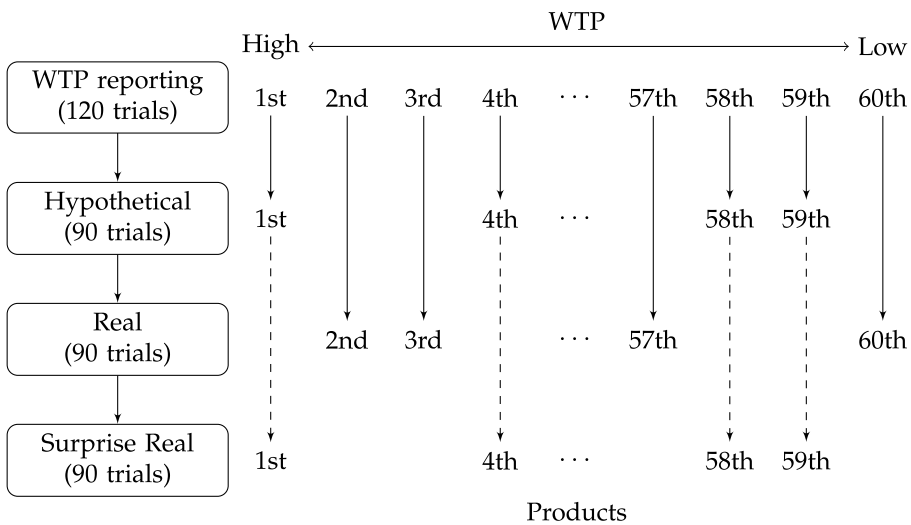


FIGURE 3.2: Timeline of the experiment and an illustration of allocation of products into each choice block.

the offer price by $P_i = WTP_i \times d$. This procedure ensured that there was a balanced distribution of three different price levels (low, middle, and high) so as to increase statistical power and facilitate detection of a treatment effect, if any, on the purchase behavior. The jitter helps to prevent subjects from noticing any type of pricing rule.

The real purchase block was identical in structure to that of the hypothetical block except for one significant difference. Specifically, subjects were informed that one of the 90 trials would be randomly chosen at the end of the experiment, and whatever decision they had made in the chosen trial would be implemented as real, whether that purchase decision was Yes or No. Since only one trial would count as real, subjects were instructed that there was no reason for concern on their part about spreading their budget too thinly over the different items. Instead, they were instructed to treat each trial independently of each other, as if it were the only decision in consideration. The offer prices for each product in this block were determined as in the hypothetical purchase block.

In the final "surprise" real purchase part, the same 90 item-price pairs that had been presented in the earlier hypothetical trials were shown again. This time, subjects were asked to make a real decision on these items. That is, the subjects were told that exactly one trial of the 180 real trials, including 90 from the earlier real part and 90 from this surprise real part, would be randomly selected and implemented, based on their decision made in the selected trial. This surprise real part was designed to measure switches from hypothetical to real decisions for a matched set of items presented once in each condition; therefore, the offer prices for each product remained identical to those in the hypothetical block.

At the end of the experiment, one trial from the real or surprise real block was randomly selected. When the subject had made a purchase decision at the offered price in the selected trial, he paid the offered price out of his $50 budget from experiment participation, and the product was shipped to him and the subject received any remainder of the $50 in cash at the end of the experiment. Otherwise, if the decision in the selected trial was to not buy, the subject received the full $50 in cash and did not receive any product.

During the pre-mousetracking part, the initial location of the anchor on the WTP scale (Figure 3.1*A*) was randomized for each trial and recorded. These data were used as a check for subjects' engagement in the task and possible anchoring effects. Correlations between participants' WTP responses and anchor positions were calculated for each subject. For two subjects, the WTP reports were highly correlated with the anchor positions ($p < 0.0001$) and the number of trials in which $|\text{anchor} - \text{WTP}| \leq 5$ was outside two standard deviations of the group average (i.e., greater than 47 trials). Therefore, these two subjects were excluded from the analysis.

Viewing time data measurement In addition to decision and response time, we also recorded the amount of time participants spent viewing the product image and the offer price in the mousetracking trials. In each of these trials, subjects saw two gray opaque boxes on the screen behind which a product image and the offered price were hidden (Figure 3.1*B*). Subjects had to click and hold the left mouse-button on one of the boxes to see the information behind it (right two panels of Figure 3.1*B*), and they were able to see either the product or the price at any given time. The viewing times of products and prices were recorded by tracking mouse events occurring on the boxes (clicks and releases)—that is, the time elapsed from the moment when the gray box opened to the moment when it closed, and aggregated (summed) within a trial. In addition to viewing times, "latency" was defined as the time between the final box closing and when choice was entered (i.e., key press) to capture last minute computations and contemplations to reach a decision. There was no time limit, and subjects could spend as much time as they wished on the task. Before the start of the hypothetical purchase block, subjects went through five practice rounds to become familiar with the mousetracking task.

The placement of the image and the price was counter-balanced between subjects. That is, one group of subjects saw the screen with a product-top and price-bottom display, and the other saw the screen with a reversed (i.e., price-top and product-bottom) display.

3.2.2 Experiment II: Eyetracking Study

The procedure for Experiment II was similar to that of Experiment I, and thus we provide descriptions only when methodological disparity is present between the two experiments.

Participants Seventeen participants were recruited in the same way as in Experiment I. Participants were screened so that they participated only once in this study (Experiment I and II, inclusive). All subjects had normal or corrected-to-normal vision, and they earned up to $50 for participation.

Experimental procedure The experimental procedure was basically identical to that of Experiment I, except that in purchase decision trials, opaque boxes were removed from the screen and subjects freely viewed information on the screen at the pace they desired (Figure 3.1C). The placement of the image and the price was counter-balanced between subjects as in Experiment I.

Gaze data measurement Gaze data were collected from subjects during the three purchase blocks using the head-mounted EyeLink 2 system (SR Research, Mississauga, Canada) at 250 Hz. Binocular gaze data from both eyes were gathered whenever available (most cases), but when there was a calibration problem with one eye, monocular gaze data were collected from the well calibrated eye. When binocular gaze data were collected, we used the average gaze position between the two eyes for gaze analysis. The system was calibrated at the beginning of each block. Drift correction was performed before each trial to ensure that accuracy of the calibration parameters is maintained. Gaze data acquisition was controlled by MATLAB (MathWorks, Natick, MA), using the Psychophysics and Eyelink Toolbox extensions (Brainard, 1997; Cornelissen et al., 2002; Kleiner et al., 2007; Pelli, 1997). [8]

[8]Eyetracking data of four trials from three participants were not recorded properly. However, decisions in those trials were recorded.

TABLE 3.1: Summary statistics in Experiment I (mousetracking) and Experiment II (eyetracking).

	Experiment I		Experiment II	
Average	Hyp	Real	Hyp	Real
Purchase percentage (%) a, b***	55.99	25.59	55.88	16.67
WTP ($)	23.51	23.51	26.91	26.86
Price ($)	21.06	21.10	24.13	24.07
Response time (sec)	4.01	4.43	2.30	2.16
Cumulative image viewing time (sec) a***	0.95	1.50	1.32	1.33
Cumulative price viewing time (sec) b*	0.70	0.64	0.49	0.43
# of image clicks/fixations a** b*	1.27	1.36	2.44	2.32
# of price clicks/fixations	1.21	1.20	1.54	1.44
Image viewing time per click/fixation (sec) a***	0.72	0.89	0.53	0.54
Price viewing time per click/fixation (sec) a***	0.60	0.53	0.32	0.30

Notes: Asterisks indicate statistical significance between hypothetical and real condition. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.10$, for Experiment I (a) and II (b). In Experiment II, averages for *purchase percentage*, *WTP*, and *price* are calculated including first trial of each condition.

3.3 Results

Behavior and viewing times Table 3.1 shows summary statistics of behavior in hypothetical and real conditions from two experiments. Note that the distribution of the WTPs subjects stated for presented products was approximately matched across hypothetical and real conditions by construction (see Section 3.2). Therefore, if there is no hypothetical bias, then subjects should choose to buy goods at the same rate in the two conditions. As found in previous studies, however, subjects exhibited a significant hypothetical bias (Kang et al., 2011; Little and Berrens, 2004). The hypothetical purchase rates were 56.0% and 55.9% in the two experiments ($SE = 2.36, 3.39$) and the real purchase rates were 25.6% and 16.7% ($SE = 3.90, 1.98$). The reduction in purchase rates is highly significant in both experiments (two-sided paired sample t-test, $p < 0.0001$).

To explore differences in the visual gaze pattern during hypothetical and real purchase that could potentially contribute to improvement of prediction, we constructed gaze distribution maps for different conditions of interest, for the eyetracking experiment II only, using *i*Map (Caldara and Miellet, 2011; Chauvin et al., 2005). Gaze distribution maps allow for statistical testing of differences between conditions in viewing any part of the stimuli. They are thus free of subjectivity and potential error in defining regions of interest (ROIs) a priori, and allow differences on a fine spatial scale. [9]

Figure 3.3 shows the gaze distribution maps (panel *A*), along with a map measuring gaze differences across condition (panel *B*) and decision (panel *C*). Panel *B* shows that subjects fixated more frequently on the price as well as the image in the real condition than in the hypothetical condition before making a Yes decision, while that pattern was reversed before making a No decision (the threshold for two-tailed Pixel test $p < 0.05$ is $|z| > 4.3125$, corrected for multiple comparisons). Similarly, Panel *C* shows that subjects fixated longer on both product and price when they made Yes decisions than No decisions. In the hypothetical condition, however, Yes-dominant areas (red) and No-dominant areas (blue) were both present. [10] These differential patterns of visual attention associated with hypothetical and real decisions are observed also in mouse-based Experiment I, using an ROI-based aggregate measure of fixation duration as shown in Figure 3.4. [11] We speculate that we did not observe any difference in viewing times between Yes and No decisions in the hypothetical condition because we constructed viewing time measures by summing up total duration fixated on the predetermined ROIs, thus discarding all the fine-grained information regarding spatial attention.

In addition to viewing times for image and price, there is also an interesting difference in latency, which is the duration between the last time subjects viewed the price or the image, and the time at which they made a decision (recorded in Experiment I only). This pre-choice latency was significantly longer for real Yes compared to real No decisions (Yes: $M = 1.08$, $SE = 0.19$; No: $M = 0.62$, $SE = 0.07$; two-sided paired sample t-test, $p < 0.01$), but there was no such difference in hypothetical decisions. (Yes: $M = 0.74$, $SE = 0.06$; No: $M = 0.82$, $SE = 0.08$). This extra pre-choice latency plausibly reflects additional last minute contemplation before choosing to actually buy the product (Figure 3.5).

Furthermore, pre-choice latencies in the hypothetical condition were longer when subjects later made a No decision rather than a Yes decision in the surprise real condition (Yes: $M = 0.64$, $SE = 0.07$; No: $M = 0.84$, $SE = 0.07$; two-sided paired sample t-test, $p < 0.001$). That is, subjects who took a longer time post-viewing before making a hypothetical choice were more likely to change their minds and say No when asked to choose for real. This is the first clue that

[9] See Section B.1 for more information about gaze data analysis, construction of ROIs, and gaze distribution maps.

[10] We obtain qualitatively similar results in the comparison between hypothetical and surprise real conditions. See statistical fixation map (Figure B.6) in Section B.3.

[11] We briefly note that number of mouse clicks or eye fixations showed basically the same patterns across hypothetical and real conditions, and across conditions, as measures of visual attention times do. See Figure B.4 in Section B.3.

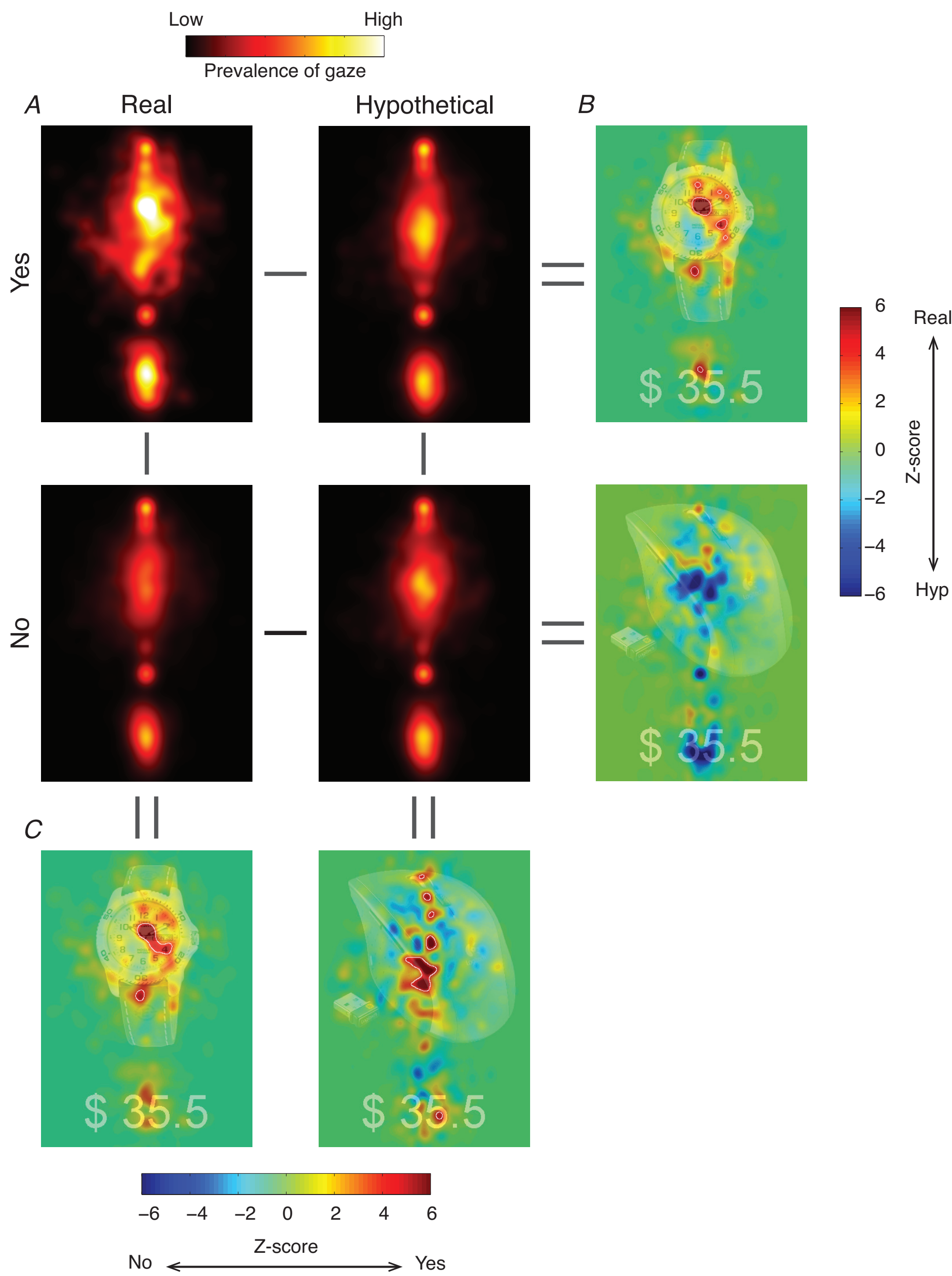


FIGURE 3.3: Gaze distribution maps by condition and decision from Experiment II. *A*. Average gaze prevalence. *B*. Statistical significance of the difference between real and hypothetical conditions. Red indicates gaze bias toward real choice (i.e., longer viewing time in the real than the hypothetical condition) and blue indicates gaze bias toward hypothetical choice. *C*. Statistical significance of the difference between Yes and No decisions within each condition. In Panels *B* and *C*, product images are shown in the background for illustration. Red indicates gaze bias toward Yes and blue indicates bias toward No. The threshold for two-tailed Pixel test $p < 0.05$ is $|z| > 4.3125$, corrected for multiple comparisons.

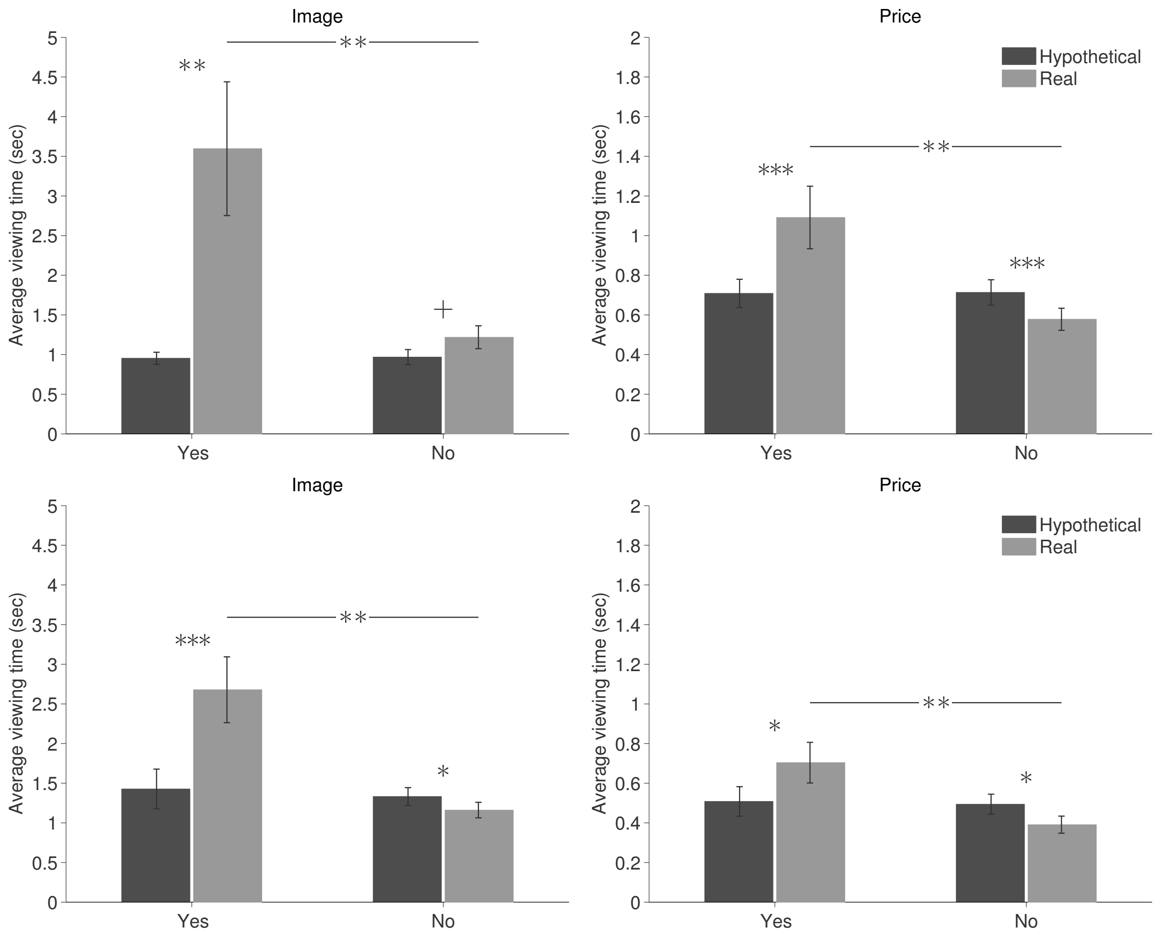


FIGURE 3.4: Average viewing time (sec) by condition and decision. Image and price viewing times in Experiment I (top) and in Experiment II (bottom). Error bars indicate standard errors. $*** : p < 0.001$, $** : p < 0.005$, $* : p < 0.05$, $+ : p < 0.10$, two-sided paired sample t-test. Comparison between hypothetical Yes and No is not significant.

features of hypothetical decisions might have some predictive power for whether hypothetical choices translate into the same real choices, for the same products (see *Classification Analysis* below).

Decision switches By design, participants faced exactly the same pairs of product and price in the hypothetical and surprise real conditions. Although subjects should make the same decisions between these two conditions in the absence of hypothetical bias, we observed frequent decision switches, as shown in Table 3.2: more than half (55.5% and 69.4% in Experiment I and II, respectively) of hypothetical Yes decisions were switched to No decisions in the surprise real condition, compared to a very low rate of switching in the opposite direction (5.2% and 3.6%).

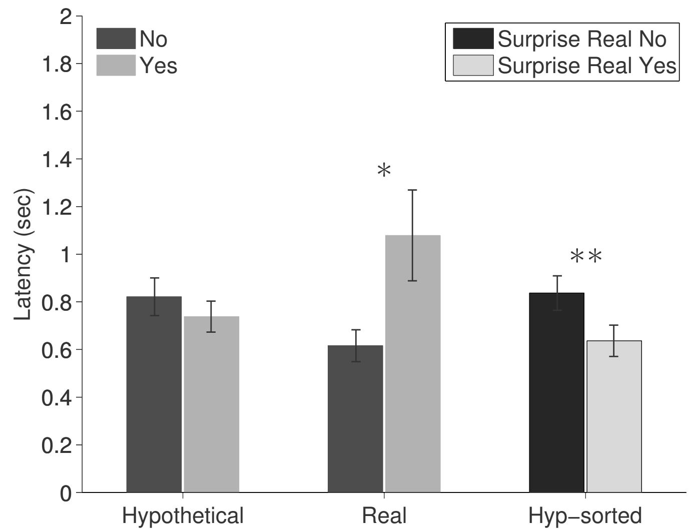


FIGURE 3.5: Pre-choice latency by condition and decision in Experiment I. For "Hyp-sorted" bars, average latencies in hypothetical condition sorted by decisions in matching surprise real trials are plotted. Two additional subjects are dropped, since they did not make any Yes decisions in surprise real trials. Asterisks indicate statistical significance: $** : p < 0.001$, $* : p < 0.01$, two-sided paired sample t-test.

Because the overwhelming majority of switches are from Yes to No, we next focus on whether there are differences in characteristics of decision problems (i.e., WTP and price) and fixation patterns between trials in which subjects made the same Yes decision in the surprise real condition ("stick" trials) and those in which they later changed their minds from Yes to No ("switch" trials). In Experiment I, during switch trials WTPs and prices were significantly higher, subjects fixated significantly longer on both image and price, and pre-choice latency was also significantly longer, than on stick trials. In contrast, we found significant differences only between price fixation time, and standardized fixation time for price in Experiment II. [12]

In Experiment I, fixations on *both* image and price were longer in hypothetical Yes decisions that later resulted in switches to No. This finding is in contrast to the general tendency to pay less attention to price in hypothetical decisions. In switch trials participants also have longer pre-choice latencies. We speculate that in hypothetical trials resulting in a Yes that later is switched, participants collect more information (= longer fixations), are still not really sure they want to buy (= longer latency), but respond with a reluctant Yes decision anyway. The longer fixations and pre-choice latency could be indicators of hesitation, despite choosing Yes.

[12]See Table B.2 in Section B.3.

TABLE 3.2: The percentage of decision switches from Hypothetical to Surprise Real decisions in Experiment I (mousetracking) and II (eyetracking).

Experiment I		Surprise Real		
		Yes	No	Total
Hyp	Yes	24.9	31.1	56.0
	No	2.3	41.7	44.0
	Total	27.2	72.8	100.0

Experiment II		Surprise Real		
		Yes	No	Total
Hyp	Yes	17.1	38.8	55.9
	No	1.6	42.5	44.1
	Total	18.7	81.3	100.0

Notes: The total number of observations: 2520 (Experiment I); 1530 (Experiment II). Numbers appear in the right panel are slightly different from those in Tables 3.1 and B.1 since in this table first trial in each condition and four trials in which eyetracking data were not recorded (but choices were recorded) are included.

Logistic regression Analysis The analyses above show differences between visual processing in hypothetical and real conditions, depending on whether choices are Yes or No. Next we ran several logistic regression models, in which the dependent variable was purchase decision (Yes = 1 and No = 0). Independent variables included viewing times and several dummy variables (Table 3.3, columns *AB* and *DE*). [13] Longer viewing times for price and image were significant predictors of the purchase decision only in the real conditions, in both experiments. The second group of logistic regression models included price, WTP, viewing times, and dummy variables as independent variables (Table 3.3, columns *C* and *F*). WTP and price were generally strong predictors of purchase decisions; this is unsurprising because these are the *only* variables predicted by standard economic theory to guide decisions, and they undoubtedly have an effect on choice.

More surprisingly, image and price viewing times were also significant predictors of purchase even after controlling for WTP and price, but only in the real condition. It is also notable that the magnitudes of the coefficients for price and WTP are smaller in the real condition than in the hypothetical condition, which implies that including viewing times reduces the statistical influence of price and subjective value (WTP) on decisions.

We also ran several logistic regression models in which the dependent variable was decision switch between hypothetical and surprise real condition (switch = 1, stick = 0) and independent variables included viewing times collected during the hypothetical trials. Focusing on hypothetical Yes trials, we found that not only prices but also price viewing time and latency were significant predictors of later

[13] Viewing times, including latency and other viewing time, were standardized within subject, across conditions.

Table 3.3: Random-effects logistic regression of purchase decision (Yes = 1, No = 0; group variable = subject).

	Experiment I			Experiment II		
Independent variables	*A*	*B*	*C*	*D*	*E*	*F*
WTP			0.214 ***			0.153 ***
			(0.017)			(0.016)
Price			−0.206 ***			−0.149 ***
			(0.017)			(0.015)
Real × *WTP*			−0.120 ***			−0.032
			(0.021)			(0.024)
Real × *Price*			0.050 **			0.014
			(0.019)			(0.017)
ImageViewing		−0.034			−0.095	
		(0.085)			(0.083)	
PriceViewing		0.007			0.007	
		(0.050)			(0.096)	
Real	−1.798 ***	−1.804 ***	−0.479	−2.499 ***	−2.542 ***	−2.316 ***
	(0.297)	(0.298)	(0.433)	(0.322)	(0.320)	(0.664)
Real × *ImageViewing*	0.223 ***	0.256 **	0.218 ***	0.402 ***	0.497 ***	0.307 ***
	(0.050)	(0.092)	(0.055)	(0.092)	(0.115)	(0.083)
Real × *PriceViewing*	0.289 ***	0.282 **	0.337 ***	0.240 *	0.232	0.281 **
	(0.077)	(0.093)	(0.078)	(0.105)	(0.140)	(0.108)
Trial	−0.005 **	−0.006 **	−0.006 **	−0.004	−0.005 *	−0.005 *
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Real × *Trial*	0.006 *	0.006 *	0.007 *	0.010 *	0.011 **	0.010 *
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
Constant	0.524 **	0.531 **	−0.067	0.427 **	0.470 *	0.037
	(0.154)	(0.162)	(0.320)	(0.164)	(0.184)	(0.462)
Log likelihood	−2934.45	−2934.28	−2456.79	−1645.50	−1644.50	−1428.37
# Obs	5040	5040	5040	3023	3023	3023

Notes: Level of significance $***: p < 0.001$, $**: p < 0.01$, $*: p < 0.05$. *ImageViewing* (*PriceViewing*) is cumulative fixation time for image (price), standardized across conditions within subject. *Real* is a dummy variable for real trials (1 if real, 0 otherwise). Robust standard errors are reported in parentheses, corrected for subject-level clustering. According to likelihood-ratio tests, adding the variable *ImageViewing* and/or *PriceViewing* to specifications *A* and *D* does not result in a statistically significant improvement in model fit.

decision switches in Experiment I (Table 3.4).

This logistic regression analysis establishes that viewing times do have significant associations with the purchase decision. Next, we investigate whether we can actually predict purchase decisions more accurately by including viewing times, using classification analysis.

Classification analysis A main goal of this study is to investigate whether we could predict consumers' actual purchase decisions using information on visual attention. Note that information of this kind is potentially available to sellers in natural settings (such as website time usage or in-store observation of customer visit times). To provide an answer to this question, we performed a linear discrim-

TABLE 3.4: Random-effects logistic regression of decision switch (switch = 1, stick = 0; conditional on hypothetical Yes; group variable = subject). Independent variables are measured during the hypothetical condition.

	Experiment I			Experiment II		
Independent variables	*A*	*B*	*C*	*D*	*E*	*F*
Price	0.039 ***		0.040 ***	0.062 ***		0.064 ***
	(0.011)		(0.011)	(0.017)		(0.017)
ImageViewing	0.002	0.015		0.189	0.234	
	(0.088)	(0.097)		(0.146)	(0.160)	
PriceViewing	0.242 ***	0.257 ***		0.064	0.079	
	(0.064)	(0.065)		(0.105)	(0.101)	
Latency/Else	0.345 ***	0.355 ***	0.357 ***	0.045	0.029	0.050
	(0.092)	(0.092)	(0.091)	(0.095)	(0.082)	(0.092)
Constant	−0.352	0.367	−0.331	−0.421	0.881 **	−0.442
	(0.415)	(0.372)	(0.422)	(0.502)	(0.314)	(0.493)
Log likelihood	−692.57	−703.16	−697.07	−435.82	−453.61	−438.05
# Obs	1411	1411	1411	842	842	842

Notes: Level of significance $***: p < 0.001$, $**: p < 0.01$, $*: p < 0.05$. *ImageViewing* (*PriceViewing*) is cumulative fixation time for image (price), standardized within subject across conditions. *Latency/Else* are constructed in the similar manner.

inant analysis (LDA) to predict purchase decisions. As independent variables, the model included price and price viewing time, image viewing time, and (in Experiment I) pre-choice latency as well as other viewing time (i.e., the duration of the gaze at blank screen areas which show neither image nor price; Experiment II). [14] [15]

We used the following procedure. Prior to classification analysis, viewing times, including latency and other viewing time, were standardized within subject, across different purchase conditions. For each condition, any subjects who made fewer than five Yes or No decisions were excluded from this classification

[14] It is important to note why WTP information was not used in the classification analyses that were reported above. Not surprisingly, WTP is closely related to purchase decisions. Indeed, including WTP along with price classifies decisions sufficiently well that viewing times no longer add predictive power. However, in most natural settings unbiased subject-specific measures of WTP are more difficult to collect than visual attention (or other kinds of self-report). For example, a user accessing a website will provide a clickstream to a retailer, even if the user does not answer a question about WTP. In many practical settings, consumers have no positive incentive to state their true subjective WTP, and will usually have a natural inclination to understate it. When respondents are especially concerned about their social image with a surveyor (even online), WTPs can be highly misreported, especially for vices and virtues.

[15] Unlike Experiment I, there was no clear-cut way to measure latency as defined above (i.e., not looking at a stimulus but still contemplating before submitting a decision) in this part unless subjects actually closed their eyes after the last fixation until the decision submission. Therefore, we used the total duration of the gaze at blank instead, called “other viewing time” or simply, “else.”

analysis, because classification requires enough observations in both response categories. In Experiment I, this criterion left a total of 23 subjects for each of the real and surprise real conditions, and 28 for the hypothetical condition. In Experiment II, no subjects were excluded for hypothetical classification, one subject was excluded from real classification, and three subjects were excluded from surprise real classification.

We performed a linear discriminant analysis (LDA) for each subject with viewing times, latency, and price as independent variables (i.e., features) to predict purchase decisions (Yes = 1, No = 0). The detailed classification procedure was as follows. First, for each subject, we divided 90 observations into a training sample to estimate a classification model, and a hold-out sample to evaluate predictions based on the estimated model. Specifically, exactly two observations, one Yes decision and one No decision, were randomly selected out of 90 observations and set aside. By construction, one could classify 50% of the decisions correctly by chance. This 50% prediction level serves as the baseline success rate against which our classification results are compared. A classification model was then estimated based on the rest of the 88 samples, and used to predict binary choices associated with the hold-out samples. This procedure was repeated 1,0000 times per subject. We called a correct prediction a "success." For classification of decisions in real (hypothetical) trials, viewing times and latency collected from real (hypothetical) trials were used. However, for classification of decisions in surprise real trials (i.e., real, binding purchase decisions), viewing times and latency from matching *hypothetical* trials were used.

Figure 3.6 shows that viewing times do improve prediction accuracy to a modest extent. Adding viewing times to price in predicting real purchase decisions (using viewing times collected in the real condition) improved accuracy from 62.7% to 69.2%, and from 62.1% to 68.5%, in Experiments I and II, respectively. These success rates for prediction were significantly higher than the baseline success rate of 50%. On the other hand, prediction of hypothetical choices was not improved by adding viewing times collected in the hypothetical condition (66.2% and 65.3% in Experiment I, and 62.3% and 62.8% in Experiment II).

A much more challenging test is whether viewing times collected in hypothetical trials can improve prediction of decisions *for the same products* in the surprise real trials. This improvement is evident, although small in magnitude, using mousetracking data (Experiment I), but there is no improvement using eyetracking data (Experiment II). That is, in Experiment I the average success rate was significantly improved from 62.4% (prices only) to 65.6% by incorporating view-

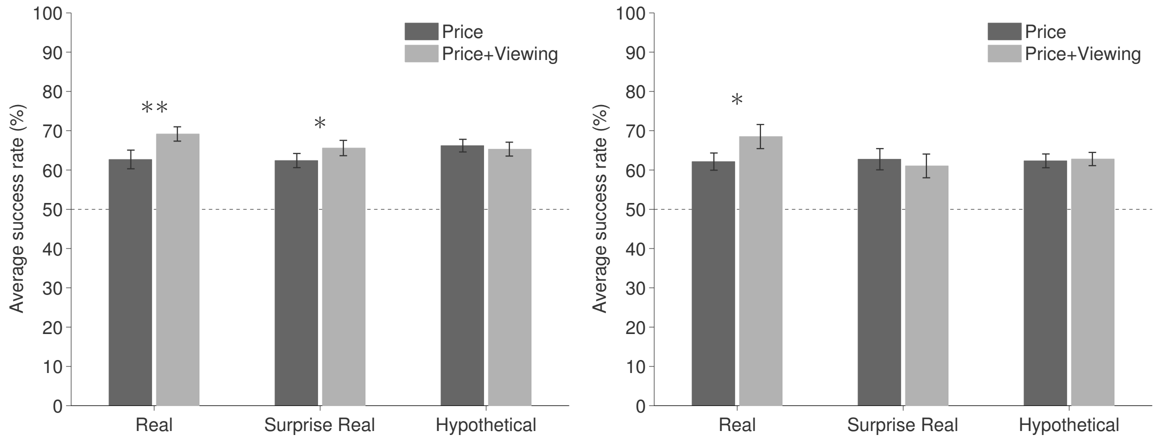


FIGURE 3.6: Average success rates for classification of decisions by condition in Experiment I (left) and Experiment II (right). Viewing times include *ImageViewing*, *PriceViewing* and *Latency* for Experiment I, and *ImageViewing*, *PriceViewing* and *Else* for Experiment II. For classification of decisions in surprise real trials, viewing times from matching *hypothetical* are used. Asterisks indicate statistical significance: $** : p < 0.001$, $* : p < 0.01$, one-sided paired sample t-test. Error bars indicate standard errors. Dashed horizontal lines represent a chance level (50%).

ing times ($SE = 1.8, 1.9$; one-sided paired sample t-test, $p < 0.01$).

We can also test, focusing only on hypothetical Yes trials, whether we can predict their decision switches in the later surprise real condition using hypothetical viewing times. In contrast with the case of predicting surprise real decisions, we do not see significant improvement in predictive accuracy by adding viewing time data in either experiment. Prediction success rates vary only a little across sets of predictors, averaging around 56% and 60% for Experiments I and II. As is common in other classification studies (e.g., Smith et al., 2014) some participants' stick/switch choices can be classified rather accurately and others cannot (see also Figure 3.7). This motivates our next analysis, in which we investigate the possibility of improving prediction by selecting a subset of predictors.

Improve prediction with feature selection In the previous analysis, we apply the same model (price only or price and viewing times) to all subjects' data to measure prediction success rates. Given the heterogeneity in visual fixation patterns, we might expect that different subsets of predictors could improve prediction for different subjects. In order to assess this possibility, we perform the classification analysis with feature selection.

Since we have a small number of predictors, we can exhaustively examine all possible combinations of predictors and compare the performance of linear

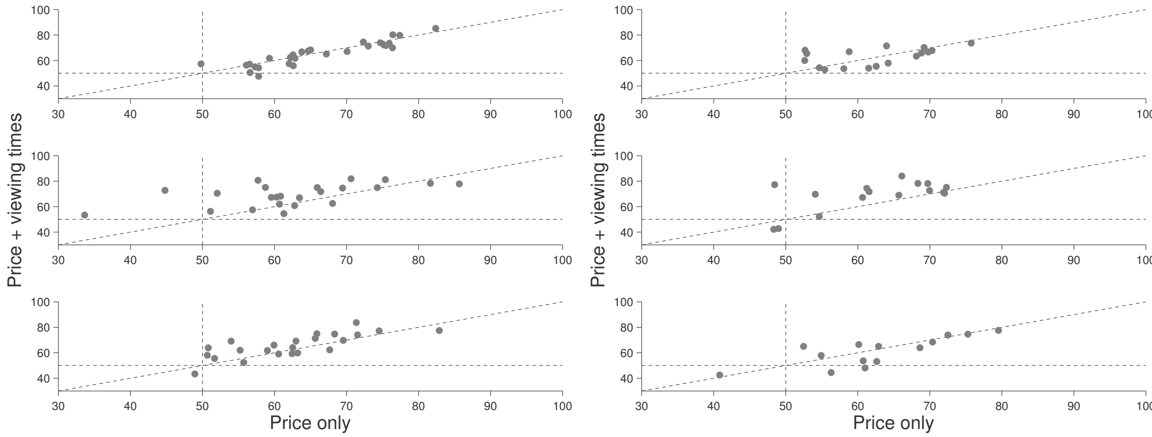


FIGURE 3.7: Individual heterogeneity in improvement of prediction success rates in Experiment I (left panels) and Experiment II (right panels). Each panel compares prediction success rates using price only (x-axis) and using both price and viewing times (y-axis). Each dot represents one subject. Top panels: prediction of hypothetical choices; Middle panels: prediction of real choices; Bottom panels: prediction of surprise real choices (with viewing times recorded during matching hypothetical trials).

classifiers in terms of their resulting success rates. [16] For each subject, we select a subset of viewing times that achieves the highest success rate when added to price information. We call this set of viewing times the best subset. Figure 3.8 shows empirical cumulative distribution functions (CDFs) for prediction success rates in each condition, using (i) price only, (ii) price and all viewing times, and (iii) price and (individual-specific) best subset of viewing times. The figure revealed that for many subjects, including some subset of viewing time information actually works better than including all of them. Although the method employed here is a naïve one (simply picking the model which attains the highest success rate), we observe the potential of improving prediction.

3.4 Discussion

A sizable amount of evidence suggests that answers to hypothetical questions are systematically biased. Using a novel experimental design, we explored whether visual attention that is easy to measure—attention to product images or prices—is associated with the tendency to overstate hypothetical purchase intentions, compared to real purchases (hypothetical bias). The present study showed that viewing times (both for product images and prices) are longer when subjects reported

[16]When the number of predictors is large, one can use methods such as forward- or backward-stepwise selection (see Hastie et al., 2009).

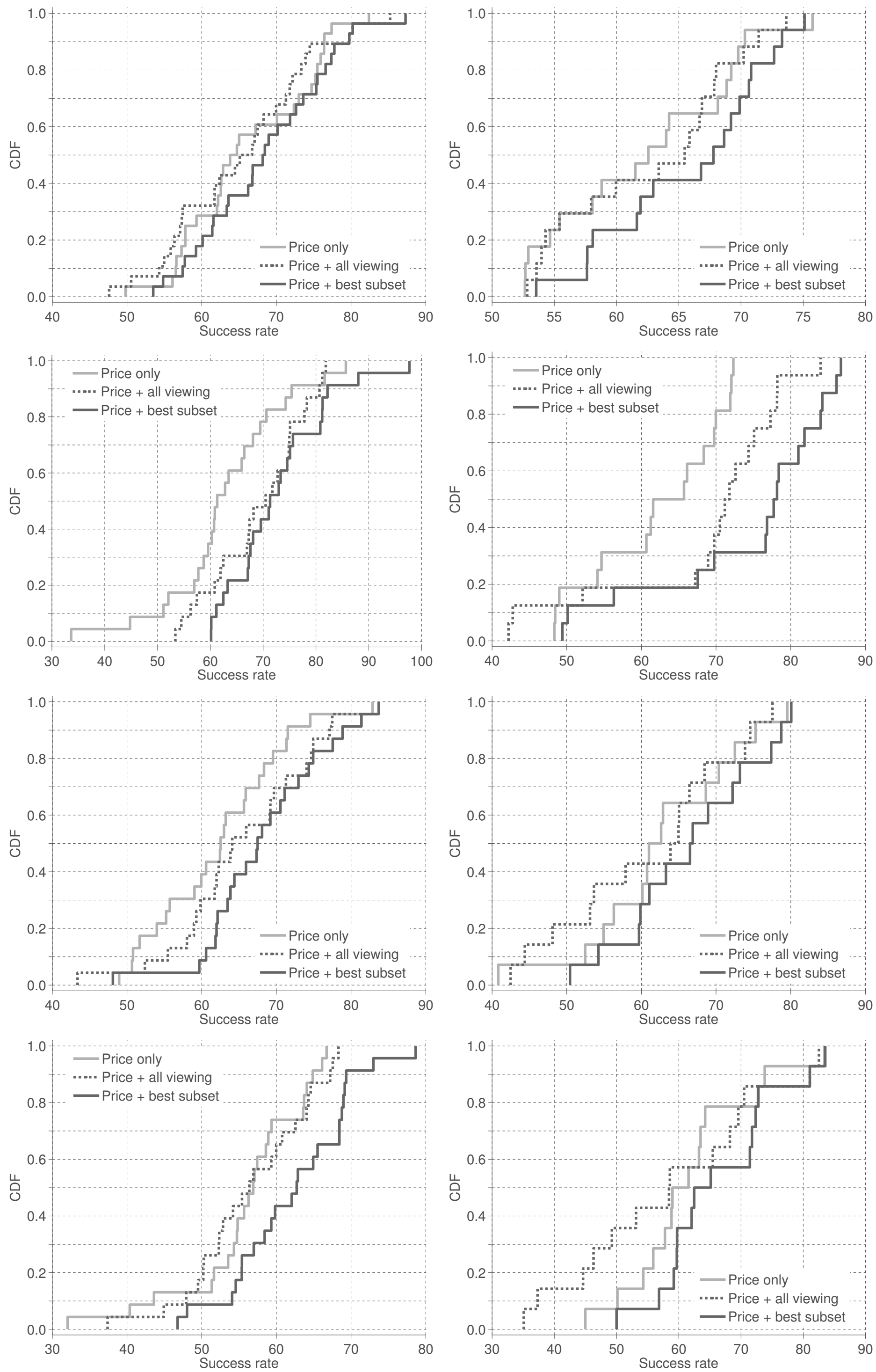


Figure 3.8: Empirical CDFs for prediction success rates using price only, price and all viewing times, and price and the best subset of viewing times (individual-specific). Left column: Experiment I; right column: Experiment II. First row: hypothetical choice; second row: real choice; third row: surprise real choice; fourth row: surprise real switch.

a purchase intention (i.e., a Yes response) than when they did not, but only in the real purchase condition.

A natural hypothesis that states individuals who are looking at image (price) longer would (would not) buy does not seem like to hold in our data. In general, price viewing time is shorter than image viewing time, probably because a number is processed more quickly than a more complex product image. However, we observed differences in viewing times when they answered Yes and when they answered No in the real condition. Such differences were not present in the hypothetical condition. The results that subjects exhibited differential patterns of visual attention in hypothetical and real purchasing decisions complements the previous neuroimaging results, showing that a common set of valuation circuit including areas of the orbitofrontal cortex and the ventral striatum exhibited stronger activation in the real condition than in the hypothetical condition (Kang et al., 2011).

We further investigated to what extent visual fixations help predicting purchasing decisions. We found that adding viewing times to prices in cross-validated linear discriminant analysis of purchase improves accuracy from 62.7% to 69.2% for real purchases, and from 62.4% to 65.6% for surprise real purchases using viewing times in corresponding hypothetical condition (but only when viewing times are recorded by mousetracking). Even though these improvements are not large in magnitude, they are highly significant, and could even be of practical importance for firms forecasting consumer behavior in highly competitive markets.

One may wonder why viewing times recorded in the hypothetical condition could improve prediction of choices in corresponding surprise real trials, even though those same viewing times did not add much in predicting choices in the hypothetical condition (Figures 3.4 and 3.6). Notice that although image and price viewing times did not vary by decision in the hypothetical condition, latency exhibited a significant difference when sorted by the corresponding choices in the surprise real condition (Figure 3.5). It is likely that the predictive power came from this information, which we could obtain only in Experiment I. Participants spent longer on last-second thoughts in hypothetical trials in which they say No in later surprise real trials for the same products. One explanation is that they spent longer before making decisions since they were unsure (or indifferent) between buying and not buying, and break the tie by answering a difficult Yes to a hypothetical question.

An important question is how our classification accuracies compare to those in the recent studies, which also aimed at predicting consumer choice or valuation.

Accuracies in three recent studies using cross-validated neural decoding from fMRI signals (rather than visual attention), were 61%, 70%, and 55-60% in Levy et al. (2011), Smith et al. (2014), and Tusche et al. (2010). Although there is clearly substantial room for improvement, our classifications using prices and viewing times between 65-69% are at the high end of the range of these other studies.

Mousetracking and eyetracking have been used widely in cognitive psychology and experimental economics in order to uncover cognitive processes behind many domains of economic decision making, including consumer choice under time pressure (Reutskaja et al., 2011), information acquisition in complex multi-attribute multi-alternative environment (Gabaix et al., 2006), bargaining (Camerer et al., 1993; Johnson et al., 2002), and strategic thinking (Brocas et al., 2014a,b; Costa-Gomes et al., 2001; Devetag et al., forthcoming; Knoepfle et al., 2009; Polonio et al., 2015; Stewart et al., 2015; Wang et al., 2010). However, little is known about the relative advantages of these two methods since most of the existing studies used either one of the methods. The current study filled this gap, by using two methods and comparing the results (the only exception is a study by (Lohse and Johnson, 1996)). It is notable that the main results in our study are basically the same using either motor-driven mouse movements to reveal information on computer screen boxes (in Experiment I), or video-based eyetracking (in Experiment II). The robustness of these choice patterns to how attention is measured is encouraging for practical applications.

Mousetracking seems to be more sensitive than eyetracking, in two ways. First, choices in surprise real decisions were better predicted with hypothetical viewing times in the mousetracking sessions than in the eyetracking sessions (Figure 3.6). The differences in viewing times between hypothetical and real choices are also larger when measured by mousetracking compared to eyetracking. We speculate that mouse movement is more effortful and deliberative than eye saccades. As a result, there will be fewer random, low-information mouse movements compared to eye saccades. If so, mouse movements are actually clearer evidence of underlying deliberate decision processes than eye saccades are.

Our results have two implications for practice.

As noted in the introduction, the two leading methods for predicting real choice from hypothetical reports are statistical adjustment, and using instructions to respondents that are intended to produce more realistic reported choices. Our method is also a kind of statistical adjustment, but it uses pre-choice cognitive data that are easy to measure. Furthermore, unlike special instructions to respondents, which require substantial internal validity and may not work well for all subjects and choices, measuring their visual attention is relatively effortless and does not require special comprehension or internalization by subjects. Given the small, but promising incremental predictive power of viewing times in predicting real choice, a more finely-tuned version of our method using viewing times could prove to be useful on larger scales.

Second, the fact that viewing times are associated with purchase invites the possibility that manipulating viewing times exogenously can causally change real purchase decisions. Indeed, previous studies have shown the presence of causal effect of manipulation of visual fixations on preferences, on a modest scale, by simply changing the duration of exposure to food items (Armel et al., 2008) or face images (Shimojo et al., 2003).

Chapter 4

Risk Taking in a Natural High-Stakes Setting: Wagering Decisions in *Jeopardy!*

4.1 Introduction

Risk preferences play a key role in the many domain of economic decision making, including insurance selection and activity in financial markets.

Traditionally, economists have been trying to understand risk attitude using several different empirical approaches. The first approach relies on macroeconomic data on consumption and savings to estimate risk attitudes implied by Euler equations derived from lifecycle models (Attanasio and Weber, 1989), survey data on household portfolios, or labor supply data (Chetty, 2006). In the second approach, researchers use controlled laboratory experiments with real monetary incentives (Camerer, 1995; Harrison and Rutström, 2008; Hey, 2014; Starmer, 2000). This strand of literature has established that substantial heterogeneity in risk preferences exist (Bruhin et al., 2010; Choi et al., 2007) and deviations from the standard expected utility theory are evident (Camerer, 1989; Camerer and Ho, 1994; Harless and Camerer, 1994; Hey and Orme, 1994). The third approach, which has been growing more recently, uses field data from choice domains that are well-structured for estimating preferences and comparing theories (see Barseghyan et al., 2015, for an extensive review). Cohen and Einav (2007) and Barseghyan et al. (2013), for example, estimate models of choice under uncertainty using insurance choice data. Jullien and Salanié (2000) and Chiappori et al. (2012) use horse racing betting data to recover a representative agent's risk preference or heterogeneity in bettors' preferences.

Since Gertner's (1993) seminal work, researchers have been using data from

TV game shows to investigate attitude towards risk. Although researchers note potential issues on the use of special sample of participants in TV game shows (e.g., they may participate to be in the limelight), TV game shows often provide a good environment to test expected utility theory as well as non-expected utility theory since they "are presented with well-defined choices where the stakes are real and sizable, and the tasks are repeated in the same manner from contestant to contestant" (Andersen et al., 2008b, p. 361).

Subsequent studies have examined the behavior of contestants, their risk preferences in particular, using data from a variety of game shows such as *Affari Tuoi* (Bombardini and Trebbi, 2012; Botti et al., 2008), *Cash Cab* (Bliss et al., 2012; Keldenich and Klemm, 2014; Kelley and Lemke, 2015), *Deal or No Deal* (Blavatskyy and Pogrebna, 2008; Brooks et al., 2009; de Roos and Sarafidis, 2010; Deck et al., 2008; Mulino et al., 2009; Post et al., 2008), *Hoosier Millionaire* (Fullenkamp et al., 2003), *Lingo* (Beetsma and Schotman, 2001), and *Who Wants to be a Millionaire?* (Hartley et al., 2014).

Using data from 4,810 *Jeopardy!* episodes, we investigate players' risk attitude, strategic thinking, and gender differences in players' behavior. *Jeopardy!* is one of the most popular TV games shows in the United States, which has more than thirty years of history. The final round of the game, which we explain in detail in Section 4.2, involves features of both risk taking and strategic interaction. Some of the analyses reported in this study follow the approach taken in Metrick (1995), who analyzes wagering behavior in the final round and finds that the representative player is risk neutral and that players do not appear to play empirical best responses. We depart from Metrick (1995) by quantifying players' subjective beliefs, which is made possible by richer nature of our dataset as well as recent advances in machine learning techniques.

We obtain the following set of observations. First, unlike Metrick (1995), we find that the representative player is risk averse under the assumption of expected utility and constant absolute risk aversion (CARA) in the situation where leading players do not need to take into account other players' strategies. Second, in strategic situations, trailing players tend to wager more than what "folk" strategies suggest, which are thought to be well known among the community of *Jeopardy!* contestants and fans, while the majority of leading players follow those folk strategies. We also show that trailing players' overbetting is based on their subjective beliefs while leading players' obedience to folk strategies is rather mechanical and not related with their subjective beliefs. Third, we look at gender differences in risk taking and find that female players are in general more risk averse than male contestants, confirming a pattern repeatedly observed in the laboratory and field.

The current study contributes to two strands of literature. First, we use a machine learning techniques, such as Least Absolute Shrinkage and Selection Operator (LASSO; Hastie et al., 2009) and cross-validation, to quantify players' subjective beliefs. It is a methodological advance in the literature on estimating risk preferences using the field data, particularly the data from TV game shows, in which researchers usually do not have access to individuals' subjective beliefs. We confirm the usefulness of the machine learning approach from two observations: (i) we obtain a well-calibrated measure of subjective beliefs (Figure 4.2), and (ii) we obtain precise estimates of the coefficients of absolute risk aversion using the measure of subjective beliefs in standard econometric estimation (Table 4.8). Second, we demonstrate that gender difference in risk taking behavior are context dependent in the TV game show under study. A voluminous empirical evidence in economics suggest that women are more risk averse than men (Byrnes et al., 1999; Charness and Gneezy, 2012; Croson and Gneezy, 2009; Eckel and Grossman, 2002, 2008; Fehr-Duda et al., 2010). We first confirm that female players are indeed more risk averse than male players in a series of analyses. On top of this, we find that male players in female-dominant groups take less risk compared to those in male-dominant groups. Female players, on the other hand, do not exhibit the gender-context effect.

The remainder of this chapter is organized as follows. Section 4.2 describes the basic rules of *Jeopardy!* and our dataset. Section 4.3 presents several descriptive statistics of our dataset. Section 4.4 revisits Metrick's (1995) analyses to examine if one of his main findings (risk neutrality of the representative player) holds in our bigger dataset. In Section 4.5 we quantify players' subjective beliefs about their chances of answering correctly in the final round. In order to do this, we first assemble a rich set of features which captures several aspects of game dynamics and then select "meaningful" subset using LASSO. We then estimate a representative player's risk preferences with these quantified subjective beliefs together with two-limit Tobit approach. In Section 4.6 we turn to strategic situations and look into the top two players' behavior with several "folk" strategies as benchmarks. Finally, in Section 4.7 we look into gender differences.

4.2 Description of *Jeopardy!* and Data

4.2.1 Rules of the Game

Jeopardy! is one of the most popular TV quiz shows in the United States. The show is "ranked number 45 on *TV Guide*'s list of the 60 greatest American television programs of all time" (Cohen, 2015). It debuted on 1964 and the current version of the show started on September 10, 1984 and is still on the air. In the show three contestants compete by answering questions in various general knowledge categories.

The game consists of three stages. In each of the first two stages, called the *Jeopardy!* round and the *Double Jeopardy!* round, a panel of 30 clues is presented. There are six categories and each of them has five clues, which are valued incrementally reflecting the difficulty of the question. The clue values in the *Jeopardy!* round originally ranged from \$100 to \$500 in \$100 increments, and values are twice as large in the *Double Jropardy!* as the name suggests. Those panel values were doubled after November 26, 2001, implying that panel values ranged between \$200 to \$1,000 in the *Jeopardy!* round and \$400 to \$2,000 in the *Double Jeopardy!* round. [1] After a contestant picks a panel, the clue is revealed and read by the host. Then, the first contestant who rings in must provide a response. Differently from other quiz shows, in *Jeopardy!*, contestants must phrase their answers in the form of questions. A correct response adds the clue value to her/his score, while an incorrect response subtracts it. The contestant who responds correctly can select the next clue from the panel. One clue in the *Jeopardy!* round and two clues in the *Double Jeopardy!* round are called *Daily Doubles.* Only the contestant who selects a *Daily Double* panel has an opportunity to respond. The contestant first decides how much to bet (the minimum amount is \$5 and the maximum is the bigger one of her/his current score and the highest dollar clue in the round) and then provides a response. At the end of the second stage, any contestants whose scores are nonpositive are eliminated from the game.

The final stage of the game is called the *Final Jeopardy!*. After the host announces the category, each contestant decides how much to bet between \$0 and her/his total score. The clue is then revealed and contestants are given 30 seconds to write down their responses. Contestants who respond correctly earn their bets.

The contestant who earns the most wins the game and returns as the champion in the next match. If two or more contestants tie for first place, they return on the

[1] We investigate how players responded to this one-time structural change in Section C.1.

Table 4.1: Number of archived games in each year.

Year	# games	Year	# games	Year	# games	Year	# games
1984	26	1992	48	2000	230	2008	232
1985	27	1993	57	2001	218	2009	226
1986	46	1994	25	2002	144	2010	231
1987	58	1995	31	2003	174	2011	224
1988	80	1996	100	2004	230	2012	231
1989	81	1997	225	2005	225	2013	231
1990	92	1998	224	2006	230	2014	231
1991	38	1999	230	2007	231	2015	152

Notes: Last access to the archive is on August 31, 2015, before the start of new season 32.

next match as co-champions. However, if all three contestants have a final score of $0, there is no winner and none of the contestants return on the next game. Originally, a contestant who won five consecutive days "retired" undefeated but this limit was eliminated after September 8, 2003.

4.2.2 Source and Description of the Dataset

We obtain the data from the website *J! Archive*, which is created and managed by former contestants (including Kenneth Jennings, who has the record winning streak of 74 games) and fans of the show. [2] The archive contains clues, answers, contestants who answered correctly and so on. The data is rich enough to reconstruct the dynamics of the game.

The website archives more than 4,800 games from the beginning of the show in 1984. Table 4.1 presents number of archived games aired in each year. The website does not provide a complete list of games broadcasted so far. In fact, many of the games before 1995 are missing, while after 1996, most of the games are archived.

In the dataset, there are nine games which involve four contestants (quarter-final games of *Super Jeopardy!* tournament). [3] The final sample includes 4,810 games which consist of 1,813 games before November 25, 2001 and 2,997 games

[2] `http://j-archive.com/`.

[3] In 13 games we observe situations in which a contestant finishing the *Double Jeopardy!* round with a zero or negative score was granted a nominal score of $1,000 with which to wager for the final round. This is a special rule in *Celebrity Jeopardy!* games. We exclude those games and two additional ("pilot") games before the official start of the first season on September 10, 1984. We also exclude three "IBM Challenge Games" on February 14 to 16, 2011, in which *Watson*, an artificial intelligent computer system developed by IBM, participated as one of the contestants.

after November 26 of the same year, when panel values were doubled.

We obtain demographic information regarding contestants. The *J! Archive* website describes each contestant's occupation (e.g., a college student, a realtor, a software engineer) and which city she or he is from. The website does not list each contestant's gender, but we recover that information using an API provided by Genderize.io. [4] The API judges gender of a first name. It cannot determine gender for 155 (7.9%) of 1,964 distinct names appearing in our dataset, but returns higher than 95% confidence for 1,403 (77.6%) of the remaining 1,809 names. We manually check gender of contestants for the ambiguous ones, with confidence lower than 90%, by looking up their pictures posted on the player profile pages on *J! Archive*. When we use players' gender in our analyses, we restrict samples to a subset of players whose gender is identified by the API with more than 90% confidence or those whose gender is identified manually.

4.3 Preliminary Analyses

4.3.1 Setup and Notations

We use the notation of Metrick (1995). Three players begin the *Final Jeopardy!* round with scores earned during the previous two rounds. We rank these players by their scores and denote them as the endowment vector $X = (x_1, x_2, x_3) \in \mathbf{R}^3_+$ such that $x_1 \geq x_2 \geq x_3$. Let $Y = (y_1, y_2, y_3) \in \times_{i=1}^{3}[0, x_i]$ be the vector of bets that satisfies $y_i \in [0, x_i]$ for all $i \in \{1, 2, 3\}$. Finally, let $A = (a_1, a_2, a_3) \in \{0, 1\}^3$ be the vector of states that represents the correct ($a_i = 1$) or incorrect ($a_i = 0$) answer for each player i.

4.3.2 Descriptive Statistics

As we described in the previous section, players can proceed to the *Final Jeopardy!* round if the total score earned in the first two rounds is positive (except for some special cases such as the *Celebrity Jeopardy!* games). Our dataset contains four games in which only one player made it to the final, 258 games in which two players made it to the final, and the remaining 4,548 games in which all three players proceeded to the final round. Three players were strictly ordered ($x_1 > x_2 > x_3$) in 4,692 games, two players were tied at second place ($x_1 > x_2 = x_3$) in

[4] `https://genderize.io`.

Table 4.2: Frequency of the states for three-players (left) and two-players finals (right).

State (a_1, a_2, a_3)	Number of observations	Frequency $f(a_1, a_2, a_3)$
$(1,1,1)$	899.0	0.198
$(1,1,0)$	506.5	0.111
$(1,0,1)$	424.0	0.093
$(1,0,0)$	545.5	0.120
$(0,1,1)$	373.5	0.082
$(0,1,0)$	496.5	0.109
$(0,0,1)$	393.0	0.086
$(0,0,0)$	910.0	0.200
Total	4548.0	1.000

State (a_1, a_2)	Number of observations	Frequency $f(a_1, a_2)$
$(1,1)$	80.0	0.310
$(1,0)$	67.5	0.262
$(0,1)$	36.5	0.141
$(0,0)$	74.0	0.287
Total	258.0	1.000

Notes: A "1" indicates a correct answer for that player. Thus, all three players answered correctly in $f(1,1,1) = 0.198$ of the games. Fractional observations are due to some ties at the beginning of the final round. Observations from tied games are split between the two possible definitions for the state.

71 games, two players were tied at first place ($x_1 = x_2 > x_3$) in 46 games, and all three players were tied ($x_1 = x_2 = x_3$) in one game.

For those 4,548 games with three players in the final round, the number of outcomes and their frequencies of each state A are presented in the left panel of Table 4.2. The right panel of Table 4.2 presents the same statistics for 258 games involving only two contestants. Finally, Table 4.3 presents the ranking of the player before and after the final round. More than 70% of leading players ended up being the champion, but other players also had chances: about 26% of the champions were at 2nd or 3rd place before starting the final round.

The total of 14,164 players made their bets and responded to the question in the final round. Here the returning champions are counted as different individuals in each game. Their answers were roughly equally split into correct (49.5%) and incorrect (50.5%) responses. The probability of correct responses for each rank of players is presented in Table 4.4. Player 1 has a higher probability of answering correctly than chance level (one-sided binomial test, $p < 0.001$) while player 3 has a lower probability than chance (one-sided binomial test, $p < 0.001$).

The left panel of Figure 4.1 shows the distribution of bet amount (bin size is \$100). We observe several spikes at round numbers (every \$1,000), possibly reflecting saliency or focality of those bets. Note, however, that players' scores, x_i, themselves cluster at those round numbers by design of the game and hence

Table 4.3: Ranking before and after the final round.

	After FJ			
Before FJ	1st	2nd	3rd	Total
1st	3,585	837	436	4,858
2nd	1,010	2,767	1053	4,834
3rd	273	1,323	3,142	4,738
Total	4,868	4,927	4,635	14,430

Notes: This table include 266 players eliminated from the final round (negative scores). Numbers in "Total" columns can exceed the total number of games (4,810) due to ties.

Table 4.4: Outcome of the *Final Jeopardy!*.

	Correct		Incorrect	
	# obs.	Freq	# obs.	Freq
Player 1	2,549	0.525	2,309	0.475
Player 2	2,407	0.499	2,423	0.501
Player 3	2,050	0.458	2,426	0.542
Total	7,006	0.495	7,158	0.505

the existence of spikes does not immediately imply that players were "left-digit biased" (Lacetera et al., 2012). The right panel of Figure 4.1 displays bet share (bet amount divided by endowment) for each player separately. Players 2 and 3 chose to bet all that they had much more frequently than player 1, who rarely did so.

There are 4,716 games in which all three contestants' genders are identified with confidence higher than 95%. Gender compositions for those games are as follows: (i) three male contestants in 492 (10.4%) games, (ii) two male contestants and one female contestant in 2,823 (59.9%) games, (iii) one male contestant and two female contestants in 1,301 (27.6%) games, and (iv) three female contestants in 100 (2.1%) games.

4.4 Metrick (1995) Revisited

The wager-decision problem players face in the *Final Jeopardy!* round is inherently a very complex one. [5] A player must take into account not only her own

[5] Nalebuff (1990) lists *Final Jeopardy!* wagering as one of the puzzles and asks "What advice can economists offer?" His reply to the question is one sentence: "I don't know."

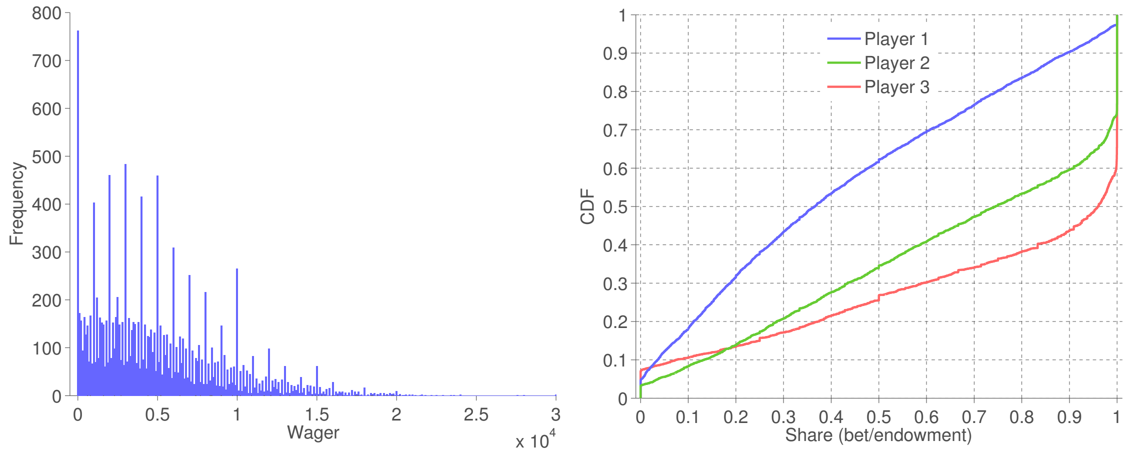


FIGURE 4.1: (Left panel) Distribution of wagers. (Right panel) Empirical CDFs for bet as share of the endowment for each player separately.

subjective belief about getting the right answer and her tolerance to risk but also her estimates of other players' beliefs and betting strategies. In certain situations, however, all of these strategic complications go away. In this section we analyze behavior in these simplified situations following Metrick's (1995) procedure to examine whether his main result (risk neutrality of the representative player) holds in our bigger dataset.

A game is called "runaway" if $x_1 \geq 2x_2$, i.e., player 1's score is so far ahead of player 2's that player 1 is (almost) guaranteed to win. In this type of games, the maximum score player 2 could obtain is $2x_2$ (by betting x_2 and answering correctly). Therefore, player 1 is guaranteed to win by betting any amount $y_1 \in [0, x_1 - 2x_2]$ irrespective of the outcome of her response. There are 1,269 games of this type in our dataset. Player 1's wagering is summarized in Table 4.5. About 95% of bets fall into the range $y_1 \in [0, x_1 - 2x_2]$, and 84.6% of those decisions (1,038 out of 1,227) are in the strict interior of the range. There are 34 situations in which player 1's score is exactly twice as large as player 2's. In 79.4% of those games player 1 chose $y_1 = 0$ to secure a win or tie.

Metrick (1995) estimated the "representative" player's risk preferences in the following manner. Player 1's decision problem is modeled as an expected utility maximization:

$$\begin{aligned} \max_{y_1} \ & p \cdot U(x_1 + y_1 + W_U) + (1 - p) \cdot U(x_1 - y_1 + W_U) \\ \text{s.t.} \ & 0 \leq y_1 \leq x_1 - 2x_2, \end{aligned} \tag{4.1}$$

where p is her subjective probability of responding correctly and W_U is her cer-

TABLE 4.5: Player 1's wagering in "runaway" games.

			Outcome	
Bet amount	# obs.	Freq	Correct	Incorrect
$y_1 = 0$	162	0.128	87	75
$0 < y_1 < x_1 - 2x_2 - 1$	927	0.730	516	411
$y_1 = x_1 - 2x_2 - 1$	111	0.087	69	42
$y_1 = x_1 - 2x_2$	27	0.021	17	10
$x_1 - 2x_2 < y_1 < x_1$	37	0.029	22	15
$y_1 = x_1$	5	0.004	3	2
Total	1,269	1.000	714	555

tainty equivalent of the wealth W for utility function U (here, W is a random variable representing the value of their future winnings for a first-time champion). Assuming a utility function with constant absolute risk aversion (CARA) specification $U(w) = 1 - \exp(-\alpha w)$ and an interior choice, the solution to the above problem (4.1) is

$$y_1^* = \frac{1}{2\alpha} \ln \left(\frac{p}{1-p} \right). \tag{4.2}$$

Since subjective probability p is unobservable (in his dataset), Metrick (1995) estimates risk preference of the "representative" player: the value of α that is most likely to result in the observed sample of bets (y_1) and frequencies of correct and incorrect responses (a_1). Rewriting the above equation, we obtain

$$p = \frac{\exp(2\alpha y_1)}{1 + \exp(2\alpha y_1)}. \tag{4.3}$$

He applies maximum-likelihood estimation with a_1 as the dependent and y_1 as the independent variable and then solves for α. Using 104 uncensored observations, Metrick (1995) reports that the point estimate for α is 6.6×10^{-5} with a standard error of 5.6×10^{-5}. The null hypothesis that the representative player is risk-neutral is not rejected.

Applying the same method to our 1,038 interior (i.e., $0 < y_1 < x_1 - 2x_2$) bets and their associated outcomes, we estimate the same logit model with a_1 as the dependent and y_1 as the independent variable. The point estimate is $\widehat{\alpha} = 1.93 \times 10^{-5}$ with a much smaller standard error of 9.41×10^{-6}. The null hypothesis of risk-neutrality is rejected at a 5% significance level ($p = 0.041$). The model implies

TABLE 4.6: Estimation of CARA utility function with logit specification.

	All	Before	After
α	$1.93 \times 10^{-5*}$	2.42×10^{-5}	$1.83 \times 10^{-5\dagger}$
	(9.41×10^{-6})	(2.45×10^{-5})	(1.05×10^{-5})
Constant	1.36×10^{-1}	1.11×10^{-1}	1.45×10^{-1}
	(8.50×10^{-2})	(1.45×10^{-1})	(1.10×10^{-1})
# Obs.	1,038	390	648
Pseudo R^2	0.0031	0.0018	0.0036
CC $(0/10^1)$	1.000	1.000	1.000
CC $(0/10^2)$	1.000	0.999	1.000
CC $(0/10^3)$	0.995	0.994	0.995
CC $(0/10^4)$	0.952	0.940	0.954
CC $(0/10^5)$	0.578	0.502	0.595
CC $(0/10^6)$	0.072	0.057	0.076

Notes. The table displays the estimation results using whole sample (column 1), subsample before November 25, 2001 (column 2), and subsample after November 26, 2001 (column 3). Standard errors are in parentheses. The implied certainty coefficient (CC; certainty equivalent as a fraction of the expected value) is shown for 50/50 gambles of \$0 or \$$10^z$, $z = 1, \ldots, 6$. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, †: $p < 0.1$.

that there is no intercept term in equation (4.3). The estimation result does not contradict this implication: the estimated intercept term is 0.136 with a standard error of 0.085 ($p = 0.109$).

Since the panel values were doubled after November 26, 2001, we estimate the model for each subsample before and after the panel values were doubled, and obtain results presented in the right two columns in Table 4.6. The estimated $\widehat{\alpha}$ is in the direction of risk aversion but insignificant before panel values were doubled, and it is marginally significant at 10% level after panel values were doubled. The difference is statistically insignificant (χ^2 test $p = 0.8285$).

Following Post et al. (2008), we also calculate the implied certainty coefficient, which is the certainty equivalent as a fraction of the expected value, for 50/50 gambles of \$0 or \$$10^z$ for $z = 1, \ldots, 6$. These values illustrate the shape of the estimated utility function: it exhibits risk neutrality when CC $= 1$ and risk aversion when CC < 1. The values of CC's are similar to those reported in Post et al. (2008), although they estimated the flexible expo-power family of utility functions which nests CARA and CRRA as special cases.

4.5 Estimating the Representative Player's Risk Preferences

Metrick's (1995) approach described in Section 4.4 relies on the assumption that actual outcomes provide a reasonable approximation of players' beliefs about whether they respond correctly or not. Hausman et al. (1998), however, show that misclassification of the left hand side variable results in an attenuated estimate in probit or logit specification. In our context, misclassification occurs when a player answers correctly when she is not expected to do so, or vice versa. In this section, we aim to improve the estimation result by incorporating subjective beliefs quantified by machine learning techniques.

4.5.1 Quantifying Subjective Beliefs

The website *J! Archive* records a detailed game information including the number of correct and incorrect responses from players, values of panels to which players answer correctly, and who gets *Daily Doubles* on panels, and so on.

Using those observables characteristics, we aim to quantify each player's subjective belief about answering the *Final Jeopardy!* question correctly. It is not a trivial task since we do not have structural knowledge regarding the relationship between those observable contents and unobservable subjective beliefs.

We approach this problem by considering a large number of candidate observable features in search of a potentially small set that is predictive, using cross-validation to control for overfitting (Arlot and Celisse, 2010; Geisser, 1975; Hastie et al., 2009; Stone, 1974). Cross-validation avoids overfitting because the training sample is independent from the test sample, as described below. The similar machine learning approach has been used in many applications in computer science and neuroscience, and a few in economics (Bajari et al., 2015; Belloni et al., 2012, 2014; Bernheim et al., 2015; Camerer et al., 2014; Einav and Levin, 2014; Krajbich et al., 2009; Smith et al., 2014; Varian, 2014).

For each player $i = 1, \dots, N$ in the *Final Jeopardy!* round from the entire dataset, let $\boldsymbol{D}_i$ denote a vector of m observable features (e.g., score, number of correct responses, and so on). We model player i's probability of answering correctly $a_i = 1$ with a logistic form:

$$p_i = \Pr[a_i = 1 | \boldsymbol{D}_i] = \frac{\exp(\beta_0 + \boldsymbol{\beta}' \boldsymbol{D}_i)}{1 + \exp(\beta_0 + \boldsymbol{\beta}' \boldsymbol{D}_i)},$$

TABLE 4.7: The set of features used for LASSO.

#	Feature
1	Score at the end of the *Double Jeopardy!* round
2	Score at the end of the *Double Jeopardy!* round (inflation adjusted)
3	Maximum score during the first two round
4	Number of correct responses (including *Daily Doubles*)
5	Number of incorrect responses (including *Daily Doubles*)
6	Percent correct responses (including *Daily Doubles*)
7	Dummy for the 1st place
8	Dummy for the 2nd place
9-18	Number of correct responses (other than *Daily Doubles*) for each of the 10 panel values
19-28	Number of incorrect responses (other than *Daily Doubles*) for each of the 10 panel values
29-38	Squared number of correct responses (other than *Daily Doubles*) for each of the 10 panel values
39-48	Squared number of incorrect responses (other than *Daily Doubles*) for each of the 10 panel values

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$. We chose 48 observable features recorded during the entire game. See Table 4.7 for the list. We choose not to include the actual wager amount or the share of wagering amount with respect to the endowment intentionally, since our primary motivation in this exercise is to quantify subjective beliefs at the beginning of the *Final Jeopardy!* round but before deciding how much to wager.

Our objective in this exercise is out-of-sample prediction (and not causal inference or coefficient estimation). We thus employ cross-validation for model assessment and model selection. In particular, we randomly split the entire dataset into 10 approximately equal-sized groups. For each of those 10 "holdout" samples, we train a model to classify players into correct responses and incorrect responses, using the remaining 90% of the dataset (called the "training" samples). We estimate a logistic regression with a Least Absolute Shrinkage and Selection Operator (LASSO). We maximize penalized log-likelihood function

$$\frac{1}{N}\sum_{i=1}^{N}\Big\{a_i \log(p_i) + (1-a_i)\log(1-p_i)\Big\} - \lambda\|\boldsymbol{\beta}\|_1,$$

where $\|\boldsymbol{\beta}\|_1$ denotes the L_1 norm on $\boldsymbol{\beta}$, i.e., $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{m}|\beta_j|$. Note that p_i depends on β_0 but this term is not penalized in the LASSO specification. By applying these trained models, we then conduct out-of-sample classification of the binary outcomes for each of the 10 holdout samples. In the context of binary models, Cramer (1999) proposes identifying an alternative as the predicted outcome if its predictive probability exceeds its baseline frequency in the population. We classify a response as "correct" if the predicted probability is above the (rank-specific) empirical frequency of correct responses, and "incorrect" otherwise (see

Table 4.4).

The value of λ is determined through cross-validation. We first partition the training samples into ten approximately equal-sized subsets indexed $k = 1, \ldots, 10$. We call each subset a "fold." For each k, we estimate penalized regression given above for each possible value of λ in a pre-specified grid, using only data from the remaining $k - 1$ folds. We use the estimated model to predict outcomes for the left-out fold, and compute accuracy of the predictions by comparing them to the actual outcomes. The value of λ with the highest successful prediction rates across all of the folds, λ^*, is then used to estimate the model with all of the observations in the training sample. We obtain a predicted probability of correct response (which we equate with the player's subjective belief) of the test sample using the estimated model.

The left panel of Figure 4.2 shows the kernel density estimation of predicted subjective beliefs for each rank of players. The mode of distribution is 0.516 for player 1, 0.490 for player 2, and 0.453 for player 3. Those values are close to actual empirical frequency of correct responses (Table 4.4). This closeness is not completely guaranteed by the LASSO method, but indicates that the predicted values correspond to the overall empirical frequencies and that overfitting is minimal. The right panel of Figure 4.2 shows the relationship between predicted probabilities of answering correctly and average probabilities of actually answering correctly (called a calibration curve). In order to generate this plot, we first split predicted probabilities of correct answers in deciles. We then calculate actual probabilities of correct answers within each decile part. The plot suggests that our prediction model is reasonably well-calibrated.

Which subset of features does LASSO select (i.e., $\beta_j \neq 0$)? We find that, from the main features, (i) inflation adjusted score, (ii) number of correct responses, (iii) number of incorrect responses, and (iv) a dummy for the first player are the important ones. Several of the numbers of correct/incorrect responses for *each* position of panel and squared numbers of them also kick in the model, but the interpretation is unclear.

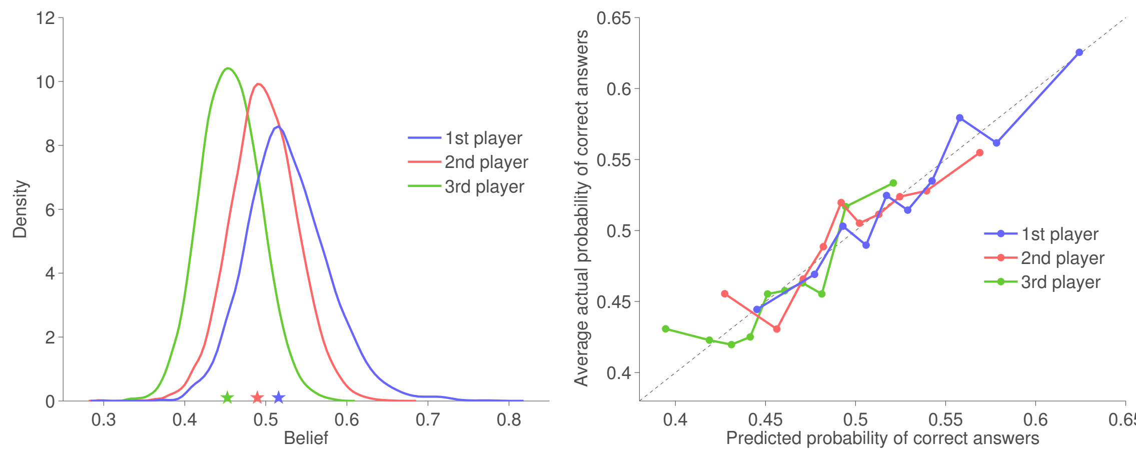


FIGURE 4.2: (Left) Kernel density estimation of subjective beliefs for each player. Stars indicate the modes of the distributions. (Right) Relationship between predicted probabilities of answering correctly and actual probabilities of answering correctly.

4.5.2 Estimation

As in Metrick (1995), we assume that player 1's decision problem is modeled as an expected utility maximization:

$$\begin{aligned} &\max_{y_1} \; p \cdot U(x_1 + y_1 + W_U) + (1-p) \cdot U(x_1 - y_1 + W_U) \\ &\text{s.t.} \;\; 0 \leq y_1 \leq x_1 - 2x_2, \end{aligned} \tag{4.4}$$

where p is her subjective probability of responding correctly and W_U is her certainty equivalent of the wealth W for utility function U. We again assume a CARA utility function $U(w) = 1 - \exp(-\alpha w)$. We are thus able to ignore the wealth term W_U. Rearranging the optimality condition yields

$$y_1^* = \frac{1}{2\alpha} \ln\left(\frac{p}{1-p}\right).$$

Assuming an additive error, this functional form can be estimated at the aggregate level. An issue to consider is the potential of corner solutions. Following Andreoni and Sprenger (2012a) and Augenblick et al. (2015), who estimate discount factor and present-bias parameter in the presence of corner choices, we provide estimates from two-limit Tobit regression. This technique is applicable to the case that the tangency condition does not hold with equality (Wooldridge, 2002). In the two-limit Tobit specification, we define a latent variable $\tilde{y}_1$ as being equal to the theoretically predicted wager for some utility function plus an error term. We then assume that we observe $y_1 = 0$ if $\tilde{y}_1 \leq 0$, $y_1 = 2x_2 - x_1$ if

Table 4.8: Estimation of CARA utility function with two-limit Tobit specification.

	All	Before	After
α	$5.81 \times 10^{-5***}$	$7.90 \times 10^{-5***}$	$5.28 \times 10^{-5***}$
	(1.97×10^{-6})	(4.56×10^{-6})	(2.19×10^{-6})
# Obs.	1,200	450	750
Log-likelihood	−10073.72	−3625.92	−6375.77
CC ($0/10^1$)	1.000	1.000	1.000
CC ($0/10^2$)	0.999	0.998	0.999
CC ($0/10^3$)	0.986	0.980	0.987
CC ($0/10^4$)	0.857	0.807	0.870
CC ($0/10^5$)	0.238	0.175	0.261
CC ($0/10^6$)	0.024	0.018	0.026

Notes. The table displays the estimation results using whole sample (column 1), subsample before November 25, 2001 (column 2), and subsample after November 26, 2001 (column 3). Standard errors are in parentheses. The implied certainty coefficient (CC; certainty equivalent as a fraction of the expected value) is shown for 50/50 gambles of \$0 or \$$10^z$, $z = 1, \ldots, 6$. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

$\widetilde{y}_1 \geq 2x_2 - x_1$, and $y_1 = \widetilde{y}_1$ otherwise. [6]

The estimation results are presented in Table 4.8. Unlike previous analysis based on Metrick's (1995) approach, here we found strong, precise evidence of risk aversion.

In order to assess economic significance of estimated coefficients of absolute risk aversion, we compare our results with estimates obtained in the literature (see Barseghyan et al., 2015, for details in estimation methods). Looking at risk aversion in other TV game shows, Gertner (1993), using two different methods, obtains statistically significant lower bounds on risk aversion 3.1×10^{-4} and 7.11×10^{-5}, Beetsma and Schotman (2001) obtain much higher risk aversion of 0.12, Fullenkamp et al. (2003) obtain mean estimates ranging from 4.8×10^{-6} to 9.7×10^{-6}, Post et al. (2008) obtain 1.58×10^{-5}, and de Roos and Sarafidis (2010) obtain the average lower bound of 7.81×10^{-5} for dynamically consistent players. [7] In the context of insurance (either auto or health) choices, Cohen and Einav (2007) obtain a median coefficient of 3.4×10^{-5}, Handel (2013) obtains estimates ranging from 1.9×10^{-4} to 3.3×10^{-4}, and Einav et al. (2013) obtain an average

[6] In this formulation, negative latent variable means a player is unconfident and would prefer to bet against himself/herself.

[7] Post et al. (2008) estimate a flexible expo-power utility function which has CARA as a special case, but coefficients for the German edition of *Deal or No Deal* reduces to CARA.

coefficient of 1.9×10^{-3}. Those estimates are the order of magnitude smaller than those estimated from lottery choice experiments. For example, Choi et al. (2007) obtain a median estimate of 0.029, Holt and Laury (2002) have 0.032, and von Gaudecker et al. (2011) obtain 0.018 from a laboratory experiment and 0.032 from an online survey. [8]

Taken together, the magnitude of absolute risk aversion estimated in the current study is consistent with those find in the field settings including other TV game shows and insurance choices, but they are much closer to risk-neutral compared to typical utility function estimated from lottery choice experiments.

4.6 Folk Wagering Strategies in Non-Runaway Games

A game is called "non-runaway" if $x_1 < 2x_2$, i.e., player 1's lead is not large enough to guarantee her a victory. In this type of games, players need to take into account how much other players would wager. Instead of trying to fully characterize equilibrium behavior, we investigate players' behavior in those strategic situations by comparing players' wager decisions with several "folk" strategies, which are known to the community of *Jeopardy!* contestants and fans.

There exist several well-known "rules of thumb" for deciding how much to wager in non-runaway situations in the community of *Jeopardy!* contestants and fans. The *J! Archive* website, the main source of our data, has a section called "wager calculation" which lists a suggestion on how much each player should wager to win the game.

Those wager suggestions, which we call *folk strategies*, depend on which subintervals of $(0, 1)$ the ratio of scores x_2/x_1 falls into. In other words, the suggestions depend on the number $m \in \mathbf{N}$ at which

$$\frac{x_2}{x_1} \in \left(\frac{m}{m+1}, \frac{m+1}{m+2} \right].$$

In the community of *Jeopardy!* fans, the threshold ratio $(m+1)/(m+2)$ is known as a "break point" and the game situation is called "two-thirds for first place" if $m = 2$, "three-quarters for first place" if $m = 3$, and "four-fifth for first place" if $m \geq 4$.

There is one simple folk strategy for player 1, which is known as "Boyd's rule"

[8] Since Holt and Laury (2002) do not report a comparable estimate, we take the value reported in Cohen and Einav (2007) who re-estimate risk aversion using Holt and Laury's (2002) "$\times 90$" treatment data.

and is applied to every situation except for some very special cases. Under this rule, player 1 is suggested to wager to cover a doubled score by player 2, i.e., wager $y_1 = 2x_2 - x_1$ or \$1 more of that. This corresponds exactly to the "shutout bet" in Metrick's (1995) terminology. For player 2, folk strategies depend on the threshold number $m \in \mathbf{N}$ but folk strategies can be expressed as intervals of the form $[\underline{w}_2^m, \bar{w}_2^m]$. When $m = 1$, it is suggested that player 2 should wager everything (i.e., $y_2 = x_2$). When $m = 2$, it is suggested that player 2 should wager between $\underline{w}_2^2 = 0$ and $\bar{w}_2^2 = 3x_2 - 2x_1$. When $m = 3$, it is suggested that player 2 should wager between $\underline{w}_2^3 = x_1 - x_2$ and $\bar{w}_2^3 = 3x_2 - 2x_1$. When $m \geq 4$, it is suggested that player 2 should wager between $\underline{w}_2^4 = 2(x_1 - x_2)$ and $\bar{w}_2^4 = 3x_2 - 2x_1$.

Notice that the value of common upper bound $3x_2 - 2x_1$ comes from the fact that player 1's final score would become $2x_1 - 2x_2$ if she wagered $y_1 = 2x_2 - x_1$ and her answer was incorrect. Player 2's wager $y_2 \leq 3x_2 - 2x_1$ guarantees her final score being at least this amount even if her answer was incorrect.

The main objective of the analysis in this section is how often each player obeys those suggested folk strategies. Tables 4.9 and 4.10 present frequencies of wagers falling into each category for each value of $m \in \{1, 2, 3, 4\}$. First, the majority of player 1 chose to wager more than $2x_2 - x_1$ to achieve a higher score than player 2's maximum possible score. However, many of those bets are within \$100 from the threshold. Player 1 rarely bet \$0, and the frequency of the bet somewhere between \$0 and $2x_2 - x_1$ increases as m becomes large (i.e., the difference between player 1's score and player2's score shrinks).

In contrast with player 1's general tendency to obey the folk strategy, player 2's bets fall outside the suggested ranges quite often (Table 4.10). Except for the case of $m = 1$, where "bet everything" (i.e., $y_2 = x_2$) is the suggested strategy, player 2 tends to overbet.

We then ask which subset of player 2 chose to obey folk strategies and who chose to overbet and which subset of player 1 chose to bet the amount that is sufficient to cover player 2's maximum possible score and who chose to bet more aggressively. It is reasonable to hypothesize that players base their decisions on their subjective beliefs. The general regression framework to capture this effect can be written as

$$\Pr[\xi_i = 1 | Z_i] = G(Z_i, \beta),$$

where ξ_i is an indicator variable that takes on the value 1 if the i-th observation of player 2 chooses $y_2 \in (\bar{w}_2^m, x_2]$ and 0 if $y_2 \in [\underline{w}_2^m, \bar{w}_2^m]$ for $m \geq 2$ (we thus drop observations where player 2 chose to underbet) for the case of player 2, and similarly, takes on the value 1 if the i-th observation of player 1 chooses

TABLE 4.9: Frequencies of player 1's wager categories.

	$m=1$	$m=2$	$m=3$	$m \geq 4$
$y_1 = 0$	0.012	0.018	0.013	0.009
$y_1 \in (0, 2x_2 - x_1)$	0.051	0.125	0.139	0.198
$y_1 = 2x_2 - x_1$	*0.060*	*0.049*	*0.040*	*0.045*
$y_1 \in (2x_2 - x_1, 2x_2 - x_1 + 100]$	*0.600*	*0.584*	*0.613*	*0.531*
$y_1 > 2x_2 - x_1 + 100$	0.278	0.225	0.195	0.216
N	1,004	570	375	1,495

Note: Italicized numbers represent the frequencies with which observed wagers match folk strategies suggestions.

TABLE 4.10: Frequencies of player 2's wager categories.

	$m=1$	$m=2$	$m=3$	$m \geq 4$
$y_2 \in [0, \underline{w}_2^m)$	–	–	0.120	0.199
$y_2 \in [\underline{w}_2^m, \bar{w}_2^m]$	0.014	*0.156*	*0.245*	*0.363*
$y_2 \in (\bar{w}_2^m, x_2)$	0.736	0.675	0.488	0.280
$y_2 = x_2$	*0.250*	0.168	0.147	0.157
N	1,004	570	375	1,495

Notes: For $m = 1$, we define $\underline{w}_2^1 = \bar{w}_2^1 = 0$. Italicized numbers represent the frequencies with which observed wagers match folk strategies suggestions.

$y_1 \in [2x_2 - x_1, 2x_2 - x_1 + 100]$ and 0 if $y_1 \in (2x_2 - x_1 + 100, x_1]$ for $m \geq 2$ (similar remark applies). The vector Z_i comprises a set of variables including *Score* x_i, *Diff* $x_1 - x_2$, and a dummy variable *Male*. Since we are primarily interested in whether estimated subjective beliefs can rationalize their wagering decisions, we include *Belief* as well as *Belief*^{2} to capture potential nonlinear effect. We choose a logit specification for the function G.

Table 4.11 presents result from a series of logistic regressions. In specification (2), the coefficient on *Belief* is negative and significant, and the coefficient on *Belief*^{2} is positive and significant. Those two coefficients are close in magnitude (the absolute values are not significantly different each other, $p = 0.657$). Those observations imply that the relationship between probability of overbetting and subjective belief is U-shape with the reflection point around 0.5. In other words, players who are more "confident," in both positive and negative directions, tend to wager more than what folk strategies suggest. Columns (4) to (6) indicate that the model does not fit player 1's behavior well. The coefficients on *Belief* and *Belief*^{2} suggest U-shaper relationship but neither of them are significant. Further-

TABLE 4.11: Estimation result.

	Player 2			Player 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Belief	−18.80	−65.52*	−70.46*	−17.21	−23.11	−21.54
	(29.22)	(31.69)	(31.82)	(21.75)	(21.60)	(22.05)
$Belief^2$	12.28	63.72*	67.94*	16.58	23.65	22.44
	(28.10)	(30.54)	(30.64)	(21.42)	(21.29)	(21.70)
$Belief_{-i}$			28.00			−21.79
			(21.23)			(27.58)
$Belief^2_{-i}$			−25.68			17.53
			(21.09)			(26.77)
Score (*K*)		−0.132***	−0.136***		−0.003	0.005
		(0.015)	(0.016)		(0.014)	(0.016)
Diff (*K*)		0.491***	0.479***		−0.068	−0.059
		(0.040)	(0.041)		(0.036)	(0.045)
Male		−0.030	−0.032		0.143	0.143
		(0.104)	(0.104)		(0.110)	(0.111)
Constant	7.045	17.79	11.73	5.493	6.746	12.68
	(7.576)	(8.198)	(9.285)	(5.508)	(5.472)	(8.619)
N	1,983	1,983	1,983	1,987	1,987	1,987
Pseudo R^2	0.009	0.096	0.097	0.000	0.003	0.005

Notes: For player 2, the dependent variable is a dummy that takes 1 if $y_2 \in (\bar{w}_2^m, x_2]$ and 0 if $y_2 \in [\underline{w}_2^m, \bar{w}_2^m]$. For player 1, the dependent variable is a dummy that takes 1 if $y_1 \in [2x_2 - x_1, 2x_2 - x_1 + 100]$ and 0 if $y_1 \in (2x_2 - x_1 + 100, x_1]$. Only observations in $m \geq 2$ are included. Robust standard errors are in parentheses. Stars indicate significance level. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

more, we do not observe any effect of own score nor score difference.

Those results, taken together, imply that player 2's wagering decisions are in part based on their subjective beliefs, while player 1's are not. It might be the case that player 1 tends to choose some amount close to $2x_2 - x_1$ mechanically, rather than deliberately think about her own chance of answering correctly.

4.7 Gender Differences

In this section, we analyze whether or not gender differences in risk-taking are observed in wagering decisions in several different contexts.

In *Final Jeopardy!* **runaway wagers** We estimate the parameter in CARA specification studied in Section 4.5.2 separately for each gender. The estimated parameter $\widehat{\alpha}_m$ for the representative male player is 5.61×10^{-5} ($SE = 2.16 \times 10^{-6}$) while the parameter $\widehat{\alpha}_f$ for the representative female player is 7.5×10^{-5} ($SE = 6.43 \times 10^{-6}$). The difference $\widehat{\alpha}_f - \widehat{\alpha}_m$ is positive and statistically significant ($p = 0.01$), implying that the representative female player is more risk averse than the representative male player.

In *Daily Double* **wagers** As we described in Section 4.2, special panels called *Daily Doubles* are hidden on the boards during the first two rounds. Unlike the case with usual panels, only the contestant who selects the panel has an opportunity to respond. Before the clue is revealed, the contestant has to decide how much to bet (the minimum amount is $5 and the maximum is the bigger one of her/his current score and the highest dollar clue in the round).

One may wonder the possibility of strategic thinking in wagering decisions for *Daily Double* situations, but the effect may be small due to the following reason. Note that *Daily Doubles* are just one part of dynamic games. Since players' scores are continually changing over time, there is no significant gain to make a strategic move just for one period.

With slight abuse of notation, here we represent score and wagering amounts at the time of *Daily Double* by x_i and y_i.

We first estimate the following model to capture simple male-female difference in wager decisions:

$$W_i = \beta_0 + \beta_1 Male_i + \boldsymbol{\beta} X_i + \varepsilon_i, \tag{4.5}$$

where *Male* is a dummy variable for male players. The control variables include the following. *Timing* $\in (0, 1]$ codes when this *Daily Double* panel was opened (number of panels opened so far, including the current one, divided by the total number of panels opened in the round). *Diff* codes the absolute difference between i's score and the highest score of the other two players (i.e., $|x_i - \max_{j \neq i} x_j|$). *Position* $\in \{1, \ldots, 5\}$ codes on which row the panel was hidden (possibly reflecting difficulty level of the question). *Gain* is the value added to the player's score immediately before picking a *Daily Double* panel. [9] *Score* is the score at the time

[9] Remember that only the contestant who provides the correct answer to the current panel can select a panel for the next round with an exception of "triple stumper" situation where all three players respond incorrectly or none of them respond. Thus, in many cases, the player who picks a *Daily Double* panel has earned some amount right before.

of *Daily Double* opening. *Treatment* takes 1 if the game is after November 26, 2001. The dependent variable is wager share, y_i divided by the maximum possible wager. [10]

Columns (1) to (4) in Table 4.12 present the result of estimation. It is clear that in general male contestants wager more aggressively than female contestants.

Next, we ask whether gender-contextual factors such as the gender of the other players in the game influence wagering decisions. We estimate the following model:

$$W_i = \beta_0 + \beta_1 Male_i + \beta_2 Minority_i + \beta_3 Male_i \times Minority_i + \boldsymbol{\beta} X_i + \varepsilon_i, \qquad (4.6)$$

where *Minority* takes 1 if player i is grouped with two opposite-gender players and 0 otherwise. Again, the dependent variable is wager share, y_i divided by the maximum possible wager.

Columns (5) to (8) in Table 4.12 present the result of estimation. The coefficient β_2 captures the difference between behavior of a female player in a male-dominant group and a female player in a female-dominant group. Similarly, the additive effect $\beta_2 + \beta_3$ captures the difference between behavior of a male player in a female-dominant group and a male player in a male-dominant group.

In all four columns, estimated β_2 is not significantly different from 0: there is no gender-contextual effect in female players' behavior. On the other hand, $\beta_2 + \beta_3$ is negative and significant in all but one case: male players in female-dominant groups wager more conservatively compared to those in male-dominant groups.

[10] Note that the maximum is bigger one of her/his current score and the highest dollar clue in the round. Therefore, players can wager a positive amount even if their scores are negative ($x_i < 0$) at the moment.

TABLE 4.12: Gender difference in *Daily Double* wagers.

Dependent variable *Share* (y_i/x_i)		(1) Round 1 Leading	(2) Round 1 Trailing	(3) Round 2 Leading	(4) Round 2 Trailing	(5) Round 1 Leading	(6) Round 1 Trailing	(7) Round 2 Leading	(8) Round 2 Trailing
Male	β_1	0.053***	0.068***	0.025***	0.092***	0.068***	0.056***	0.030***	0.104***
		(0.010)	(0.011)	(0.004)	(0.007)	(0.012)	(0.014)	(0.005)	(0.009)
Minority	β_2					0.019	−0.028	0.005	0.010
						(0.016)	(0.017)	(0.007)	(0.010)
Male × *Minority*	β_3					−0.059**	0.010	−0.024*	−0.055***
						(0.022)	(0.025)	(0.010)	(0.016)
Control		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N		2,623	1,917	4,354	4,653	2,623	1,917	4,354	4,653
Adjusted R^2		0.487	0.360	0.231	0.362	0.487	0.358	0.230	0.362
H_0: $\beta_2 + \beta_3 = 0$						$p = 0.007$	$p = 0.313$	$p = 0.006$	$p < 0.001$

Notes: Robust standard errors are in parentheses. Stars indicate significance level. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

Discussion In the series of analyses presented above, we consistently find that male players wager more than female players do. This pattern is consistent with the widespread view and its voluminous empirical supporting evidence that women are more risk averse than men (Byrnes et al., 1999; Charness and Gneezy, 2012; Croson and Gneezy, 2009; Eckel and Grossman, 2002, 2008; Fehr-Duda et al., 2010). A recent study by Kelley and Lemke (2015), using data from TV game show *Cash Cab*, also find that a group of male players is more likely to take the double-or-nothing gamble than a group of female players. Our finding that male players in female-dominant games wager less aggressively compared to those in male-dominant games is, however, in the opposite direction from Lindquist and Säve-Söderbergh's (2011) finding: in the Swedish version of *Jeopardy!*, female players wager more conservatively when they are assigned to an all-male group, compared to playing in a mixed or female-only group.

In order to asses whether the observed gender-context effect is driven by a motive for competition or a mere-presence of the other gender, we estimate a model similar to (4.6) using *Final Jeopardy!* wagers in runaway games. As we explained in Section 4.4, in runaway games, player 1's score is so far ahead of player 2's that player 1 is guaranteed to win. Thus, if we observed a gender-context effect in wagering decisions in runaway games, the effect would be driven primarily by a mere-presence of the other gender, rather than a motive for competition.

Table 4.13 presents the result. We observe that neither β_2 nor $\beta_2 + \beta_3$ is significantly different from 0, meaning that the gender-context effect is not present in runaway wagering decisions.

Given this result, the gender-context effect observed among male players in *Daily Double* situations (Table 4.12) may be interpreted as a result of a motive for intrasexual competition. This interpretation is in line with the argument in evolutionary psychology, which suggests that gender difference in risk taking is rooted partially in differences in intrasexual competition. Since Trivers (1972), it has been argued that men have had to compete with one another over access to mating opportunities to a greater extent than women have throughout evolutionary history. Since risk-taking can communicate characteristics such as confidence, it can increase a man's chances of gaining access to desired mating opportunities. Although risky behavior can be costly, evolutionary theories of mate competition suggest that men often are inclined to pursue a relatively high risk-high payoff strategy when it comes to acquiring mates (Baker and Maner, 2009; Wilson and Daly, 1985). This view is further supported by the evidence that male risk-taking is pronounced primarily under social circumstances that prime psy-

TABLE 4.13: Gender-context effect in *Final Jeopardy!* runaway wagers.

Dependent variable	(1) y_1/x_1	(2) y_1/x_1	(3) y_1/δ	(4) y_1/δ
Male	0.036^{**}	0.026^{*}	0.023	0.048
	(0.012)	(0.011)	(0.096)	(0.099)
Minority	0.005	0.009	0.121	0.097
	(0.018)	(0.017)	(0.162)	(0.161)
Male × *Minority*	−0.030	−0.024	−0.215	−0.199
	(0.023)	(0.022)	(0.174)	(0.172)
Constant	0.131^{***}	0.092	0.623^{***}	0.286
	(0.010)	(0.052)	(0.083)	(0.243)
Control	No	Yes	No	Yes
N	1,216	1,216	1,183	1,183
R^2	0.009	0.218	0.001	0.014
H_0: $\beta_2 + \beta_3 = 0$	$p = 0.1025$	$p = 0.2911$	$p = 0.1442$	$p = 0.1137$

Notes: Robust standard errors are in parentheses. Stars indicate significance level. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

chological states associated with intrasexual competition (Daly and Wilson, 2001). Since *Jeopardy!* is one of the most popular TV game show in the U.S., being in the limelight may act as a priming factor of intrasexual competition.

4.8 Conclusion

In the current study, we touch several topics in microeconomics and behavioral economics through a rich behavioral data from a popular TV game show *Jeopardy!*. We find that (i) the representative player in this game is risk averse but the degree of risk aversion is very limited, (ii) players overbet than or conform to folk strategies depending on the contexts, (iii) female players are more risk averse than male contestants, and (iv) male players in female-dominant groups take less risk compared to those in male-dominant groups.

The behavior of contestants in TV game shows like *Jeopardy!* may not immediately be generalized to an ordinary person's decisions under risk. We still believe that investigating choices in this particular game show provides new insights into individual's attitude toward risk and gender differences, because the decision problems are simple and well-defined, have a feature of (in many cases mixed-gender) competition, and more importantly, the amounts at stake are very large.

Our analysis first reveals that the representative player is risk averse in situations where the presence of other players is safely ignored (runaway games). Although the estimated coefficients of absolute risk aversion is statistically significant, the implied degree of risk aversion is moderate. Even with tens of thousands of dollars at stake, the representative player rejects offers in excess of three-quarters of the expected value. In this regard, we note two limitations in the current approach: we assume a representative player and an expected utility of wealth model. The mismatch between small scale risk aversion commonly observed in experimental studies and limited large scale risk aversion found in the current study may thus be a product of those two methodological assumptions.

Our second main observation in this study is the gender-context effect. On top of the "baseline" gender effect, in which female players are more risk averse than male players, we find that male players in female-dominant groups take less risk compared to those in male-dominant groups. In other words, male players become more competitive in the presence of other male players. This finding is consistent with the mate competition story often discussed in evolutionary theories, but we do not observe the effect in games with female-female competition.

Every season, *Jeopardy* hosts "special" games such as *Tournament of Champions*, *College Championship*, *Celebrity Jeopardy!*, *Teachers Tournament*, *Teen Tournament*, and *Kids Week*. Looking at how different groups of demographically similar players respond to gender contexts will be a fruitful avenue for future research.

Chapter 5

Dynamics of Forecast Miscalibration in the U.K. Horse Racing In-Play Betting Data

5.1 Introduction

The favorite-longshot bias (FLB) is one of the most well-documented regularities in betting markets (Camerer, 2004; Coleman, 2004; Jullien and Salanié, 2008; Ottaviani and Sørensen, 2008; Thaler and Ziemba, 1988; Tompkins et al., 2008). Specifically, bettors value longshots (horses with a relatively small chance of winning) more than expected given how rarely they win, and they value favorites too little given how often they actually win. As a result, the expected return from any bet increases with the probability that the event will occur. The bias is considered as an important deviation from the market efficiency hypothesis (Fama, 1965, 1970), which argues that the betting odds for an event provide the best forecast of its probability of occurrence and that the expected return at all odds will be the same.

FLB was first documented by Griffith (1949) and McGlothlin (1956), and subsequently observed in many types of events across many countries. A partial list goes as follows: Dowie (1976), Ali (1977), Snyder (1978), Hausch et al. (1981), Asch et al. (1982), Henery (1985), Bird et al. (1987), Vaughan Williams and Paton (1997), Jullien and Salanié (2000), Snowberg and Wolfers (2010), and Gandhi and Serrano-Padial (2015) document FLB in horse racing, both in parimutuel and fixed-odds markets in the U.S., the U.K., and Australia. Zuber et al. (1985) and Sauer (1998) find evidence in the NFL point spread market. [1] See Table D.1 for a

[1] Although the majority of studies find evidence in favor of FLB, some find the "reverse" FLB

list of studies documenting FLB, the reverse FLB, or no bias.

In contrast with a voluminous evidence showing FLB (or reverse FLB) in betting markets, no attention has been paid to its dynamic aspects: when the bias appears and how it evolves over time. This is in part due to limited availability of data that allow researchers to investigate such questions. Note that most existing studies use parimutuel odds or fixed odds set by bookmakers, where the final closing odds are the only available source of data.

In the current study, we exploit a novel and rich dataset from a U.K. online betting exchange market, Betfair, to provide a new insight on FLB. An online betting exchange is a relatively new form of betting mechanism (see Section 5.2 for details). It is a type of prediction market, in which participants trade a set of all-or-nothing contingent claims on some events (Arrow et al., 2008; Snowberg et al., 2011, 2013; Sonnemann et al., 2013; Wolfers and Zitzewitz, 2004). Under certain conditions, the price of the contingent claim can be interpreted as the market's expectation of probabilistic forecast about the event's likelihood (Manski, 2006; Wolfers and Zitzewitz, 2007). Using a second-by-second limit-order book data from Betfair on all U.K. horse races (including both flat and jump races) between January to early November in 2014, we conduct a clean test for FLB *at each moment in time*, before and during the races.

We obtain two main findings. First, unlike standard findings in parimutuel betting markets which show evidence supporting FLB, we find that betting odds taken from a 10 minute time window prior to races are well-calibrated (that means, there is no FLB). Second, when we look at time window just before (40 seconds to 5 seconds) races finish, market odds exhibit systematic FLB. Furthermore, the degree of bias gets larger as races approach to the finish line.

Methodologically, we first nonparametrically estimate *calibration curves*, which relate objective probability (actual empirical probability that horses win) to subjective probability implied by market odds, through a local polynomial regression. We further follow Page and Clemen's (2013) approach to construct valid confidence bands along the estimated calibration curves. The shape of the nonparametric calibration curve lends itself to a clear evidence of growing FLB prior to the race finish, and then we go one step further by estimating calibration curves assuming specific parametric structures to relate our results with the "probability weighting function" extensively studied in non-expected utility models of decision making under risk, and also to understand potential mechanisms behind the

(e.g., Gillen et al., 2015; Woodland and Woodland, 1994, 2001, 2003) or no bias at all (e.g., Busche, 1994; Busche and Hall, 1988).

result. Our estimation results indicate that number of horses in the given race per se is not related with the degree of FLB. The association with pre-race aggregate trading volume, which reflects liquidity of the market and potentially the degree of participants' attention to the race, is also weak. We found that total race duration, even though we focus on the final 40 seconds from the finish, significantly contributes to the degree of FLB.

5.2 Data

5.2.1 Background on Betting

We first introduce common terminologies on betting and standard mechanisms of betting markets.

Betting The betting environment that is standard in Europe, especially in the U.K., is operated by licensed organizations called *bookmakers*. Bookmakers play a similar role as market makers in financial markets by adjusting *odds* (i.e., prices in betting environment). Their objective is to achieve a *balanced book* so that they will profit from whatever the outcome of the event is. The type of bet in this environment is called a *fixed-odds bet*. Under this form of betting, party A, who wants to bet on (*back*) some event, agrees to pay another party B who wants to bet against (*lay*) the same event a certain amount (*stake*) if the outcome does not hold, and B agrees to pay A the same stake multiplied by the odds if the outcome holds.

There are several conventions in displaying odds. *Fractional odds* quote the net total that will be paid out to the bettor relative to the stake if the event holds. Odds of 3/1 (also displayed as 3:1; read "three-to-one") implies that the bettor stands to make a \$300 profit on a \$100 stake. Odds of 1/1 are also known as *even*. *Decimal odds* quote the winning amount that would be paid out to the bettor (including stake). Therefore, the decimal odds are equivalent to the decimal value of the fractional odds plus one: fractional odds 3/1 are quoted as 4 and even odds of 1/1 are quoted as 2. [2]

[2] *Moneyline odds* is a popular format among U.S. bookmakers. Quotes can be either positive or negative. If the figure quoted is positive, the odds are quoting how much money will be won on a \$100 wager. Fractional odds of 3/1 would be quoted as +300. If the figure quoted is negative, then the odds are quoting how much money must be wagered to win \$100. Fractional odds of 1/3 would thus be quoted as −300.

The second standard betting system, common in the U.S. horse racing in particular, is called *parimutuel betting* (Thaler and Ziemba, 1988). In this system, all bets related to a particular event are pooled together. The final payout is determined by the manner in which winning bets divide the total money on losing bets, less transaction costs including track take. The odds are thus determined only after the pool is closed.

The third betting system, which has become increasingly popular since around 2000, is online *betting exchange*. The market matches bettors who want to back an outcome with those who are willing to lay the same outcome. Notice that in the traditional betting environment, bookmakers and race tracks take the lay side of every bet. Another important feature of this new type of betting market is that bettors are allowed to place bets *in-play*. That means, bettors are able to make transactions even during events of interest (until just before the final outcomes are realized). The betting exchanges are thus also known as in-play betting or live betting.

Betting exchanges are essentially order-driven markets: bettors have the choice of either (i) placing a limit order, which is an order to buy or sell a bet at a specific price (i.e., odds) and wait for another bettor to match the bet or (ii) place a market order and thereby directly match a bet that has already been offered by another bettor.

There are many online betting exchange markets including Bet365, Betfair, Bet Victor, Coral, Ladbrokes, Paddy power, Skybet, Stan James, Totepool, Unibet, and William Hill. Betfair is one of the biggest online betting exchange market, which is located in the U.K. and covers a wide range of events including american football, bandy, baseball, basketball, bowls, boxing, cricket, cycling, darts, financial bets, floorball, football, gaelic games, golf, greyhound racing, handball, horse racing, ice hockey, mixed marital arts, motor sports, politics, pool, rugby, snooker, tennis, volleyball, winter sports. Volumes on the exchange are estimated to have doubled from 5.23 billion USD to 11.06 billion USD between 2003 and 2004, and almost doubled again between 2004 and 2005 (Croxson and Reade, 2014). Total trading volume increased to 83.55 billion USD in 2014. [3]

Market odds as probabilities It is a common wisdom that the odds (or prices) in these betting exchanges reflect the market's expectations about the likelihood of associated events. For example, 3/1 fractional odds (or decimal odds of 4)

[3]Betfair annual report (available at `http://corporate.betfair.com/~/media/Files/B/Betfair-Corporate/pdf/annual-report-2015.pdf`).

would imply a view that the event under betting is three times more likely not to occur than to occur (25% chance of occurrence). Similarly, in the parimutuel betting, the proportion of the money in the pool that is wagered on any given outcome (horse) can be interpreted as the subjective probability that the outcome will occur (horse will win).

Betting exchange is one type of "prediction markets," in which probabilities derived from market prices prove to be useful in forecasting (Arrow et al., 2008; Snowberg et al., 2011, 2013; Sonnemann et al., 2013; Wolfers and Zitzewitz, 2004). Forsythe et al. (1992) and Berg et al. (2008), for example, examine the performance of the Iowa Electronic Market on political outcomes and find that markets outperform polls as predictors of future election results. Firms have also created internal markets to predict outcomes of corporate interest, such as new product sales (Chen and Plott, 2002; Cowgill and Zitzewitz, 2015). More recently, prediction markets have been used to aggregate opinion regarding the "reproducibility" (i.e., whether or not scientific results reported in studies will be reproduced in replication studies) of scientific results (Camerer et al., 2015; Dreber et al., 2015).

However, the interpretation of odds as subjective probabilities is under hot debate. Manski (2006) points out that appropriate theoretical results guaranteeing that interpretation were lacking and shows that, if traders are risk neutral and have heterogeneous beliefs, prediction market prices only partially identify the central tendency of those beliefs. Gjerstad (2005) investigates relationship between coefficients of relative risk aversion, the distribution of traders' beliefs, and equilibrium prices. Wolfers and Zitzewitz (2007) then provide sufficient conditions (such as log utility) under which the inverse of betting odds coincides with average beliefs among traders. Sonnemann et al. (2013) empirically show that prediction market prices closely match the mean and median of traders' subjective beliefs.

5.2.2 Description of Data

We obtained data on 8,834 U.K. horse race events from Fracsoft, a third-party vendor of Betfair betting exchange data. The dataset cover all U.K. races (both flat and jump) between January 1st, 2014 and November 10th, 2014. Our dataset focus only on the win market and do not include the place market.

We merge the limit-order book data from Betfair with race information data we obtain from two sources. The first source is Betfair itself, which provides an indicator of winning horse together with volume-weighted average odds before races start and aggregate trading volume before and after races start. [4] The second source is the Sporting Life website. [5] We extract race information, including official race start time, winning time, finish position, distance, from the "Racecards" pages and "Full results" pages.

The primary variable of interest in the current study is probabilistic forecasts of the winning probability for each horse in a given race implied by market odds at each moment before and after the race starts. Since Betfair uses "decimal odds," which stands for the payout ratio of a winning bet, the inverse of the decimal odds can be interpreted as the market expectation of probability of occurrence of the underlying event. [6]

We calculate the implied (raw) probability $q_{h_r}(t)$ that horse h_r will win race r forecasted at time t (either pre-race of in-race) by inverting the mid-point of back-lay spread (Brown, 2014). More precisely, given the best back odds $O^B_{h_r}(t)$ and the best lay odds $O^L_{h_r}(t)$ at time t, we calculate the implied probability by

$$q_{h_r}(t) = \frac{1}{\left(O^B_{h_r}(t) + O^L_{h_r}(t)\right)/2}. \tag{5.1}$$

These market probabilities on all possible outcomes of an event usually sum to greater than one because of the transaction costs, the so-called "overround." Following the convention in the literature and assuming that the overround is equally distributed over the outcome probabilities, we obtain normalized probabilities by

$$p_{h_r}(t) = \frac{q_{h_r}(t)}{\sum\limits_{i_r \in H_r} q_{i_r}(t)}, \tag{5.2}$$

where H_r is the set of horses in race r. In the sequel we may drop subscript r when no confusion is expected.

5.2.3 Descriptive Statistics

The total of 80,871 horses started in 8,834 races and 72,556 of them (89.7%) finished. There is a large variability in the data. The average number of horses in a race is 9.15. The minimum is two horses (for as many as two races) and the

[4] `http://www.betfairpromo.com/betfairsp/prices/index.php`.

[5] `http://www.sportinglife.com/`.

[6] Betfair allows for quotes from 1.01 to 1,000. The grid becomes increasingly coarser, ranging from one penny increments between odds of 1.01 to 2.00, to 10 pounds increments between odds of 100 to 1,000.

maximum is 34 horses (for only one race). There are only 76 races with more than or equal to 20 horses.

The mean race duration is 173.14 seconds and the median is 129.07 seconds. There are 2,644 races which finished in 90 seconds or shorter, and there are 4,836 races which lasted longer than 120 seconds.

The mean pre-race trading volume is 71,094 (GBP) and the median is 27,387. The mean in-play trading volume is 20,665 (GBP) and the median is 4,089. The event that attracted the largest amount of money in our dataset is Cheltenham Gold Cup. This is a Grade 1 National Hunt race in which horses run three miles and two-and-half furlongs (5,331m) and jump 22 fences on the way. [7] The race is a part of "Cheltenham Festival," which takes place annually at March and is the highlight of the National Hunt season. The total pre-race trading volume in this only one race amounted to over nine million GBP.

The distribution of winning probabilities implied by pre-race volume-weighted average odds is shown in the left panel of Figure 5.1. The mean implied probability is 0.115 and the median is 0.083. The red line represents the maximum likelihood fit of beta distribution $B(\alpha, \beta)$ with the estimated shape parameters $\alpha = 0.975$ and $\beta = 6.988$. The right panel of Figure 5.1 illustrates the dynamics of implied probabilities during Cheltenham Gold Cup. Before the race started, two horses named *Bobs Worth* and *Silviniaco Conti* were strong favorites (volume-weighted average odds for those two horses before the race were 2.66 and 4.03, respectively). Implied probabilities were relatively stable until around 350 seconds after race started. Then, the dynamics became busy: implied probabilities for two favorite horses became volatile but suddenly, the third horse *Lord Windermere* (highlighted red in the right panel of Figure 5.1) took the lead and won the race at the very last moment. [8]

The left panel of Figure 5.2 shows a time series of median (together with 25- and 75-percentile) amount of cumulative money (in the unit of 1,000 GBP) traded on the exchange before races start. Notice that about 80% of the money is bet in the last five minutes before races start. This shape is consistent with the finding in a U.S. racetrack parimutuel betting market reported in Camerer (1998). Similarly,

[7] National Hunt racing is the official name of the type of horse racing in which horses jump fences and ditches.

[8] Note that the European/U.K. style of racing is quite different from the typical American style. American races are faster-paced, and it is common for a few horses to vie for leadership early in the race, and change positions frequently. In European/U.K. style the pace is slower early in the race and horses do not switch positions as frequently; they save energy for a mad scramble in the last 30 seconds of the race. This is reflected in the typical odds profile in the right panel of Figure 5.1 in which the implied subjective probabilities do not change much until late in the race.

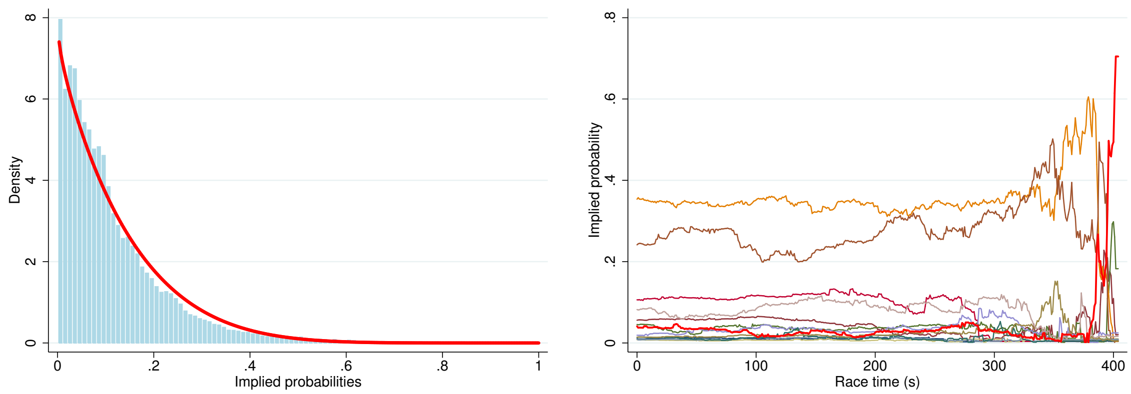


FIGURE 5.1: Probabilities implied by pre-race trading volume weighted average odds with estimated Beta density (left) and the dynamics of implied probabilities during a long race, the Cheltenham Gold Cup (right).

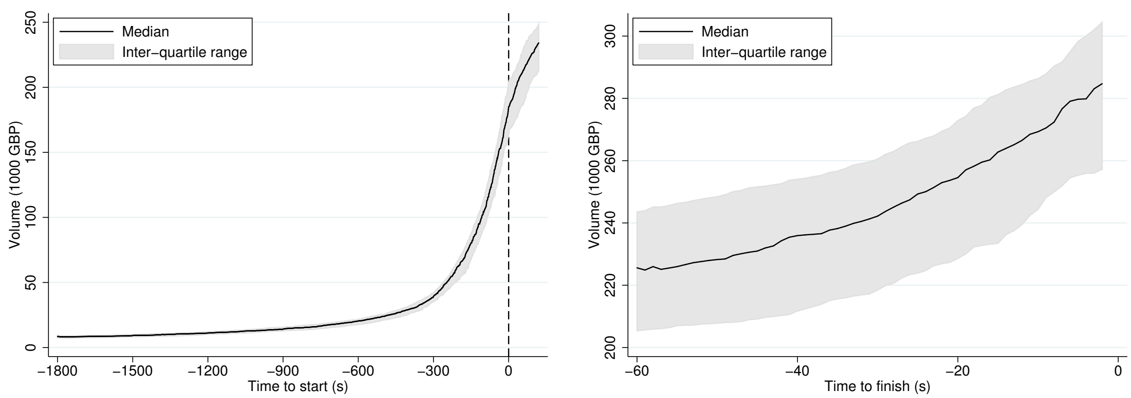


FIGURE 5.2: Median and inter-quartile range of cumulative trading volumes before start (left) and finish (right).

the right panel of Figure 5.2 shows the same series before races finish. While the magnitude of increase is less pronounced compared to the series for pre-race, the shape exhibits slight convexity.

It is a general tendency in this market that we observe relatively stable paths of implied probabilities before races start. This is reasonable since traders in the market receive virtually no additional information in this time window. This tendency holds until 30 to 40 seconds before races finish. Implied probabilities then become much more volatile. Some approach toward 1 while others approach toward 0. See Figure D.1 for an illustration of dynamics of implied probabilities pre- and in-race (an extension of the right panel of Figure 5.1). Given those typical patterns of the dynamics of implied probabilities, in the following analyses we divide the entire race duration into two categories. The first category is pre-race, and in particular we focus on 10 minutes before races start. The second category is the last phase of the races. Since most of the dynamic aspects of horse racing appear in the "last straight," we look at up to 40 seconds before races finish.

5.3 Empirical Tests of the Favorite-Longshot Bias

In this section, we test the existence of FLB and its dynamics over time, using four different approaches (for a review, see Coleman, 2004). All of the approaches share the common idea of comparing the objective win probabilities with the probabilities implied by market odds. It is often useful to visualize the relationship between objective and implied win probabilities, which we call a *calibration curve* (Dawid, 1982; Lichtenstein et al., 1977). In typical empirical studies on FLB, implied probabilities are taken on the x-axis and objective probabilities are taken on the y-axis. In the current study, we flip the axes to make it easily comparable to the shape of the *probability weighting function*, which takes objective probabilities on the x-axis and subjective probabilities (or decision weights) on the y-axis.

The probabilistic forecast is *well calibrated* if, for example, of those horses to which the market assigns a probability 20% of winning, the large-sample empirical proportion that actually win turns out to be close to 20%. This definition implies that we would observe 45-degree-line calibration curve for well-calibrated forecasts. The existence of FLB, on the other hand, implies that we would observe calibration curves that have an inverse-S shape. Formally, a calibration curve $\phi : [0,1] \to [0,1]$ exhibits FLB if $\phi(q) > p$ for $q \in [0, q_0]$ and $\phi(q) < q$ for $q \in [q_1, 1]$ for some $0 < q_0 \leq q_1 < 1$. We reserve the notation p for probabilities implied by market odds. In order to avoid confusion, hereafter we denote objective probability by q.

5.3.1 Bunching Horses with Similar Odds

We start our analysis from the method commonly used in the early literature of FLB (see, for example, Andrikogiannopoulou and Papakonstantinou, 2011; Jullien and Salanié, 2000). We first divide horses into K roughly equal-sized groups by their odds. Let H_k denote the set of horses assigned in group k and N_k denote the cardinality of H_k, namely, the number of horses in group k. We then compute the

TABLE 5.1: A bunching-based test of the Favorite-Longshot Bias.

Group (k)	# Obs.	$\min_{h\in H_k} p_h$	$\max_{h\in H_k} p_h$	Frequency ($\bar{\pi}_k$)	Average implied probability ($\bar{p}_k$)	z-statistic	p-value
1	7256	0.0009	0.0151	0.0061	0.0075	1.5751	0.1152
2	7256	0.0151	0.0304	0.0214	0.0230	0.9449	0.3447
3	7255	0.0304	0.0459	0.0385	0.0377	−0.3158	0.7522
4	7256	0.0459	0.0638	0.0557	0.0545	−0.4506	0.6523
5	7255	0.0638	0.0840	0.0775	0.0736	−1.2308	0.2184
6	7256	0.0840	0.1057	0.0994	0.0945	−1.3896	0.1647
7	7256	0.1057	0.1389	0.1290	0.1212	−1.9939	0.0462
8	7255	0.1389	0.1826	0.1615	0.1593	−0.5299	0.5962
9	7256	0.1826	0.2579	0.2281	0.2169	−2.2700	0.0232
10	7255	0.2579	0.9730	0.3996	0.3798	−3.4423	0.0006

average implied probabilities of horses in group k, $\bar{p}_k$, by

$$\bar{p}_k = \frac{1}{N_k} \sum_{h \in H_k} p_h.$$

Let $\omega_h \in \{0, 1\}$ denote an indicator for whether or not horse h won the race. We then compute the average winning probabilities of horses in group k, $\bar{\pi}_k$, by

$$\bar{\pi}_k = \frac{1}{N_k} \sum_{h \in H_k} \omega_h.$$

With sufficiently large and independent N_k, the average winning probability $\bar{\pi}_k$ approaches a normal distribution. Then, a z-statistic can be computed by

$$z_k = \frac{\bar{p}_k - \bar{\pi}_k}{\sqrt{\frac{\bar{\pi}_k(1-\bar{\pi}_k)}{N_k}}}.$$

We are thus able to perform a series of z-test to examine the difference between implied winning probabilities and actual winning probabilities for each group k (Ali, 1977; Busche and Walls, 2000; Gandar et al., 2001).

The results from those z-tests are reported in Table 5.1. Here, we calculate (normalized) implied probabilities using pre-race volume-weighted average odds to reflect all the information contained in pre-race trading activities. We find that the market significantly underestimates the probabilities of winning by horses in group $k \in \{7, 9, 10\}$. However, we do not find systematic overestimation of small probabilities as suggested by FLB.

TABLE 5.2: A regression-based test of the Favorite-Longshot Bias.

	Before start		Before finish			
Timing	−600 (s)	−60 (s)	−40 (s)	−20 (s)	−10 (s)	−5 (s)
β_1: Slope	1.041	1.031	1.035	1.073	1.089	1.084
	(0.012)	(0.012)	(0.008)	(0.005)	(0.003)	(0.002)
β_0: Constant	0.0001	0.0013	−0.0029	−0.0081	−0.0102	−0.0097
	(0.0014)	(0.0014)	(0.0009)	(0.0006)	(0.0003)	(0.0002)
# observations	72,078	72,091	72,063	72,068	72,068	72,068
# clusters	8,762	8,764	8,759	8,760	8,760	8,760
p-value (H_0: $\beta_0 = 0$)	0.958	0.353	< 0.005	< 0.001	< 0.001	< 0.001
p-value (H_0: $\beta_1 = 1$)	< 0.001	< 0.010	< 0.001	< 0.001	< 0.001	< 0.001
p-value (H_0: $(\beta_0, \beta_1) = (0, 1)$)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Note: Clustered standard errors in parentheses.

5.3.2 Linear Regression

The second approach for testing FLB that is popular in the literature is based on estimating

$$\omega_h = \beta_0 + \beta_1 p_h(t) + \varepsilon_h \tag{5.3}$$

and testing the joint null hypothesis of $H_0 : (\beta_0, \beta_1) = (0, 1)$, which is essentially testing whether implied probabilities are perfectly calibrated (on the 45-degree line). Note that we reverse the axes in this formulation: implied probabilities are on the x- axis and actual probabilities are on the y-axis.

We estimate (5.3) using implied probabilities at 600 and 60 seconds before races start and 40, 20, 10, and 5 seconds before races finish. Results are presented in Table 5.2. The joint hypothesis $H_0 : (\beta_0, \beta_1) = (0, 1)$ is rejected at all timing examined here. Before the race starts, β_1 is significantly larger than one and β_0 is not significantly different from zero, implying that only underestimation of large probabilities is present. After races start, on the other hand, $\beta_1 > 1$ and $\beta_0 < 0$ are both statistically significant, implying that underestimation of large probabilities and overestimation of small probabilities are both present.

5.3.3 Nonparametric Estimation of Calibration Curves

The previous two approaches have two potential drawbacks. First, they do not provide the precise shape of calibration curves. Second, they occasionally include several horses from the same race in one group, even though only one can win, violating independence. Those drawbacks motivate our third approach, which is originally proposed in the recent study by Page and Clemen (2013). The large-sample nature of our data makes it possible to estimate a calibration curve ϕ nonparametrically. Nonparametric regression provides a useful diagnostic tool for detecting FLB.

Let the data be $\{(X_i, y_i)\}_{i=1}^n$ from an unknown joint density f. The regression function for y_i on X_i is

$$m(x_0) = \mathbb{E}_f[y_i|X_i = x_0].$$

We want to estimate this nonparametrically, with minimal assumptions about the structure of m. The idea for local linear estimator is to fit the local model

$$y_i = \beta_0 + \beta_1(X_i - x) + \varepsilon_i$$

through the observations in the same neighborhood. The reason for using the regressor $X_i - x$ rather than X_i is so that the intercept equals $m(x) = \mathbb{E}[y_i|X_i = x]$. Once we get the estimates $\widehat{\beta}_0(x)$ and $\widehat{\beta}_1(x)$, we then set $\widehat{m}(x) = \widehat{\beta}_0(x)$. We can use $\widehat{\beta}_1(x)$ to estimate $\partial m(x)/\partial x$.

Fan (1993) extends the idea of local linear regression to construct a smooth version of a local polynomial: finding α and β to minimize

$$\sum_{i=1}^{n} K\left(\frac{x_0 - X_i}{h_n}\right)\{y_i - \beta_0 - \beta_1(x_0 - X_i)\}^2,$$

where K is a kernel function and h_n is a bandwidth. Let $\widehat{\beta}_0$ and $\widehat{\beta}_1$ be the solution to the weighted least squares problem given above. Simple calculation yields

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^n W_i y_i}{\sum_{i=1}^n W_i}$$

with W_i defined by

$$W_i = K\left(\frac{x_0 - X_i}{h_n}\right)(s_{n,2} - (x_0 - X_i)s_{n,1}),$$

where

$$s_{n,l} = \sum_{i=1}^{n} K\left(\frac{x_0 - X_i}{h_n}\right)(x_0 - X_i)^l, \;\; l = 0, 1, 2, \ldots$$

This idea is an extension of Stone (1977) and is similar in spirit to locally weighted scatterplot smoothers (LOWESS; Cleveland, 1979), but is simpler to implement since it does not require the identification of nearest neighbors.

Page and Clemen (2013) take this idea to estimate local regression line for each implied probability p by solving

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} K_h(p - p_i)\{y_i - \beta_0 - \beta_1(p_i - p)\}, \tag{5.4}$$

where p_i represents the i-th of n observations used in the estimation, h is the width of an estimation window around p and K_h is an Epanechnikov kernel defined by

$$K_h(p - p_i) = \frac{3}{4}\left[1 - \left(\frac{p - p_i}{h}\right)^2\right] \mathbf{1}\{|p - p_i| \leq h\}. \tag{5.5}$$

The estimator of conditional expectation $\mathbb{E}[\omega|p]$ is then given by $\widehat{\beta}_0$.

This approach provides precise estimates for very high and low probabilities (Fan, 1992, 1993; Fan and Gijbels, 1992). [9] This is particularly important in our context since we are interested primarily in whether a calibration curve systematically deviates from perfect calibration at probabilities close to two boundaries, 0 and 1.

An innovation in Page and Clemen's (2013) approach is the use of clustered bootstrap to account for the non-independence of implied probabilities. [10] Using groups of non-independent markets as clusters in a bootstrap resampling, they are able to estimate a confidence interval for the entire calibration curve.

Figure 5.3 shows the nonparametrically estimated calibration curves ϕ_t^s, using implied probabilities at $t \in \{600, 60\}$ seconds before races start. The gray area is the 95% confidence band calculated by the clustered bootstrap procedure with 1,000 replications.

Implied probabilities are well calibrated on the range $[0, 0.5]$ and in particular close to the boundary (see Figure D.2 for the shape of calibration curves around $q \in [0, 0.1]$). The estimated calibration curves ϕ_t^s deviate from perfect calibration outside this range, but the confidence bands are also wider. This is natural since, before a race starts, even the most favored horse usually has an implied probability less than 0.5. The results indicate that FLB is very limited before races start: we observe slight tendency to underestimate objective probabilities larger than 0.9 but the confidence band is relatively wide in this region due to the smaller sample size (it is rare to have pre-race implied probabilities higher than 0.9).

Next, we turn to nonparametric estimation of calibration curves ϕ_t^f, using im-

[9] Härdle (1990) documents boundary problem in kernel estimation.

[10] See Efron and Tibshirani (1993), Härdle and Bowman (1988), Härdle and Marron (1991).

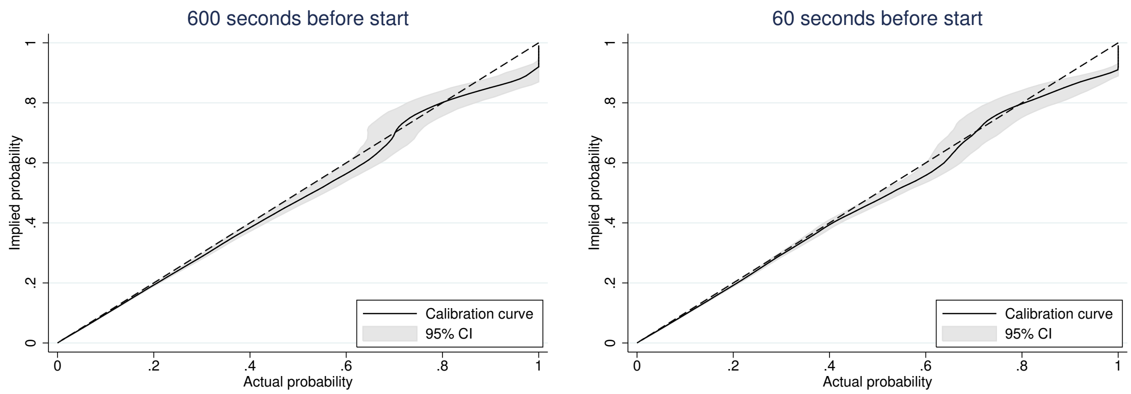


FIGURE 5.3: Nonparametric estimation of calibration curves ϕ_t^s, $t \in \{600, 60\}$ seconds before races start.

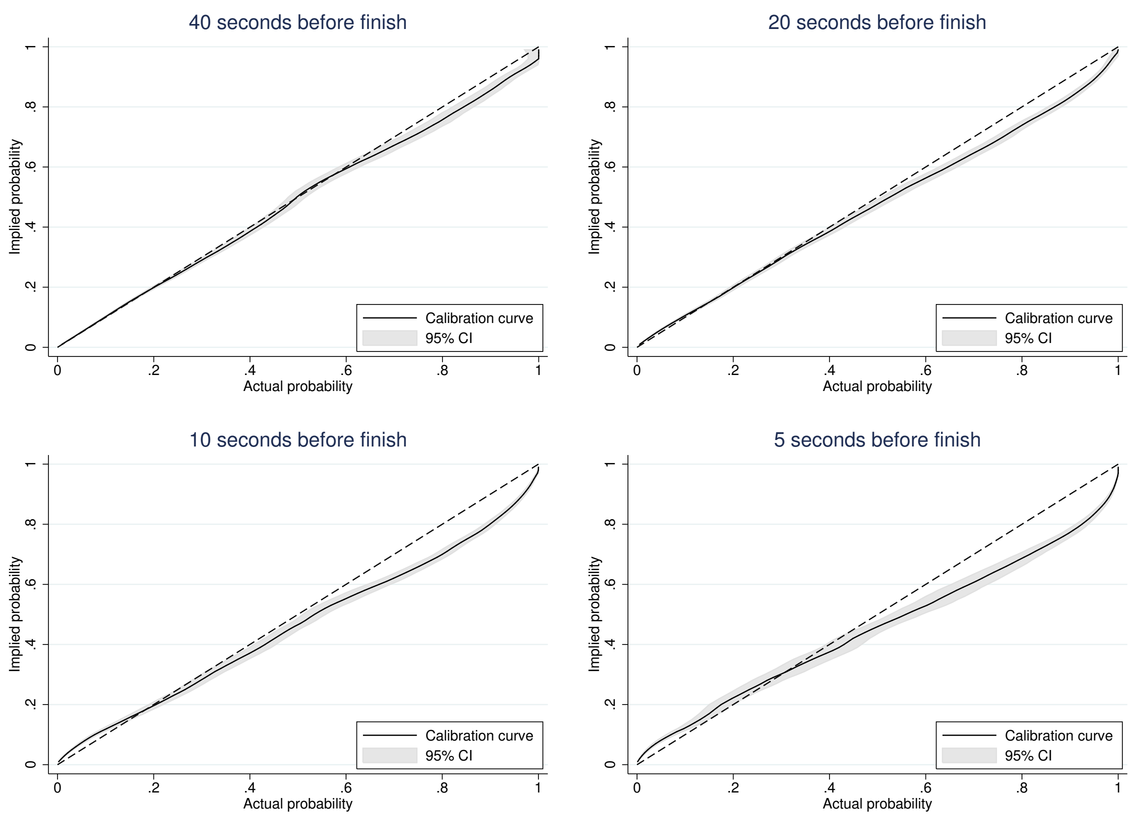


FIGURE 5.4: Nonparametric estimation of calibration curves ϕ_t^f, $t \in \{40, 20, 10, 5\}$ seconds before races finish.

plied probabilities at $t \in \{40, 20, 10, 5\}$ seconds before races finish. Figure 5.4 displays estimated ϕ_t^f with 95% confidence bands. First, we observe limited FLB at 40 seconds before races finish: ϕ_{40}^f coincides with perfect calibration (i.e., 95% confidence band covers the 45-degree line) for objective probability $q \in [0, 0.25]$.

However, ϕ_t^f exhibits bigger deviations from perfect calibration as races approach to finish line ($\phi_{20}^f(q) > q$ for $q \in [0, 0.11]$ and $\phi_{20}^f(q) < q$ for $q \in [0.42, 1]$; $\phi_{10}^f(q) > q$ for $q \in [0, 0.15]$ and $\phi_{10}^f(q) < q$ for $q \in [0.35, 1]$; $\phi_5^f(q) > q$ for $q \in [0, 0.21]$ and $\phi_5^f(q) < q$ for $q \in [0.44, 1]$). Noticeably, the inverse-S shape of calibration curves ϕ_t^f exhibits significant FLB at the implied probabilities extremely close to 0 and 1 (see Figure D.3 for the shape of calibration curves around $q \in [0, 0.1]$). [11]

Our finding that the degree of FLB is magnified as the remaining time horizon of the event shortens is the contrary to Page and Clemen's (2013) finding that the FLB is weaker as the time to contract expiration shortens. Note, however, that the types of events studied in Page and Clemen (2013) and the current study are different. Page and Clemen (2013) use data from Intrade markets on future events, especially on political and sports events, which usually have a much longer time span than horse racing markets examined here.

5.3.4 Parametric Estimation of Calibration Curves

The results from nonparametric estimation already implicated the non-existence of FLB prior to races and increasing FLB just before finishes. In order to understand structures of nonlinear relationships between objective and implied probabilities and investigate potential mechanisms behind them, we add several parametric structures in estimation of calibration curves.

In particular, we borrow ideas from the literature on non-expected utility theories such as prospect theory (Kahneman and Tversky, 1979), cumulative prospect theory (Tversky and Kahneman, 1992; Wakker and Tversky, 1993), and rank-dependent utility (Diecidue and Wakker, 2001; Quiggin, 1982). A key construct in those models is a decision weight, or subjective probability weighting function (Barberis, 2013; Wakker, 2010; Wu et al., 2004). A probability weighting function $w : [0, 1] \to [0, 1]$ maps objective probabilities to (subjective) decision weights. Several functional forms have been proposed to capture probability weighting (see Fehr-Duda and Epper, 2012, for a review). Here we focus on one particular class of two-parameter functional form suggested by Lattimore et al. (1992), which has a clear separation between curvature and elevation (Abdellaoui et al., 2010; Gonzalez and Wu, 1999). [12] Another well-studied two-parameter class of prob-

[11] Polkovnichenko and Zhao (2013) nonparametrically estimate probability weighting functions using prices of the S&P500 index options and find similar inverse-S shape.

[12] Tversky and Kahneman (1992) propose the single-parameter probability weighting function $w(q) = q^\gamma / (q^\gamma + (1-q)^\gamma)^{1/\gamma}$ which is empirically examined in Camerer and Ho (1994), Wu and Gonzalez (1996), Abdellaoui (2000), Bleichrodt and Pinto (2000), Bleichrodt (2001), Harrison and

ability weighting function is suggested (and axiomatized) by Prelec (1998). We obtain qualitatively similar conclusions under this alternative specification and thus results are omitted. Using an adaptive design optimiaztion, Cavagnaro et al. (2013) find that those two two-parameter specifications of the probability weighting function provide the best explanation of experimental data at the individual level while there is heterogeneity across subjects in which one describes the data best.

The probability weighting function studied by Lattimore et al. (1992) has the following form:

$$w(q) = \frac{\delta q^{\gamma}}{\delta q^{\gamma} + (1-q)^{\gamma}}, \quad \gamma, \delta \geq 0. \tag{5.6}$$

The first parameter γ controls the curvature and measures sensitivity towards changes in (intermediate) objective probability. The value $\gamma = 1$ corresponds to linear probability weighting function and smaller γ indicates larger deviation from linearity in a commonly observed inverse-S direction. The second parameter δ controls the elevation ($\delta = 1$ corresponds to a crossing at $q = 0.5$). This form was originally used by Goldstein and Einhorn (1987) as a generalization of Karmarkar (1978, 1979), although not as a probability weighting function.

This class of probability weighting function is also known as linear-in-log-odds specification since it is assumed that the log of the weighted odds and the log probability odds have a linear relationship

$$\ln\left(\frac{w(q)}{1-w(q)}\right) = \gamma \ln\left(\frac{q}{1-q}\right) + \ln \delta.$$

Zhang and Maloney (2012) consider how probabilistic information is used in a wide variety of cognitive, perceptual, and motor tasks and find that the distortion of probability in all cases is well-captured as linear transformations of the log odds of frequency, providing an empirical justification for the use of the linear-in-log-odds specification.

Rutström (2009), Andreoni and Sprenger (2012b), Andrikogiannopoulou and Papakonstantinou (2013), and Callen et al. (2014). Prelec (1998) provides an axiomatization of another one-parameter specification given by $w(p) = \exp(-(-\ln q)^{\gamma})$, which is empirically examined in Wu and Gonzalez (1996), Bleichrodt and Pinto (2000), Donkers et al. (2001), Snowberg and Wolfers (2010), Hsu et al. (2009), Tanaka et al. (2010), and Takahashi et al. (2010).

TABLE 5.3: Estimation of Lattimore et al. (1992) probability weighting function.

	(1)	(2)	(3)	(4)	(5)	(6)
	Before start		Before finish			
Timing	−600 (s)	−60 (s)	−40 (s)	−20 (s)	−10 (s)	−5 (s)
γ: Curvature	0.983	0.997	0.959	0.872	0.774	0.685
	(0.015)	(0.016)	(0.014)	(0.013)	(0.016)	(0.020)
δ: Elevation	0.925	0.946	0.938	0.866	0.803	0.784
	(0.021)	(0.022)	(0.016)	(0.010)	(0.007)	(0.008)
# observations	72,078	72,091	72,063	72,068	72,068	72,068
# clusters	8,762	8,764	8,759	8,760	8,760	8,760
p-value (H_0: $\gamma = 1$)	0.272	0.837	2.70×10^{-3}	4.20×10^{-22}	1.01×10^{-47}	1.73×10^{-54}

Note: Clustered standard errors in parentheses.

By inverting (5.6), we obtain

$$q = \frac{1}{1 + \left(\frac{\delta(1 - w(q))}{w(q)} \right)^{1/\gamma}}. \tag{5.7}$$

We estimate parameters in (5.7) using Nonlinear Least Squares with standard errors clustered at the race level. We then obtain calibration curves by inverting back the function (5.7) and plugging estimated $\widehat{\gamma}_t$ and $\widehat{\delta}_t$:

$$\phi_t(q) = \frac{\widehat{\delta}_t q^{\widehat{\gamma}_t}}{\widehat{\delta}_t q^{\widehat{\gamma}_t} + (1 - q)^{\widehat{\gamma}_t}}. \tag{5.8}$$

Here, parameters are indexed by time t to make the timing of implied probabilities used for estimation explicit.

Table 5.3 presents the estimation results. The first two columns estimate ϕ_t^s for $t \in \{600, 60\}$. The results indicate that the curvature parameter γ is not significantly different from 1 in both cases, indicating non-existence of FLB prior to races. Furthermore, the relationship between implied probabilities and objective probabilities are stable in this range of time: three estimated ϕ_t^s, $t \in \{600, 60, 10\}$ are indistinguishable (left panel of Figure 5.5) and $\widehat{\gamma}_t$'s are not significantly different from 1 on the entire time interval from 600 seconds until one second before races start (right panel of Figure 5.5; gray area represents 95% confidence band).

Columns (3) to (6), which present estimated ϕ_t^f for $t \in \{40, 20, 10, 5\}$ seconds before races finish, also echo our nonparametric results. Estimated curvatures are significantly different from 1 in all cases, and as races approach to the finish, the deviation of $\widehat{\gamma}_t$ increases (Figure 5.6).

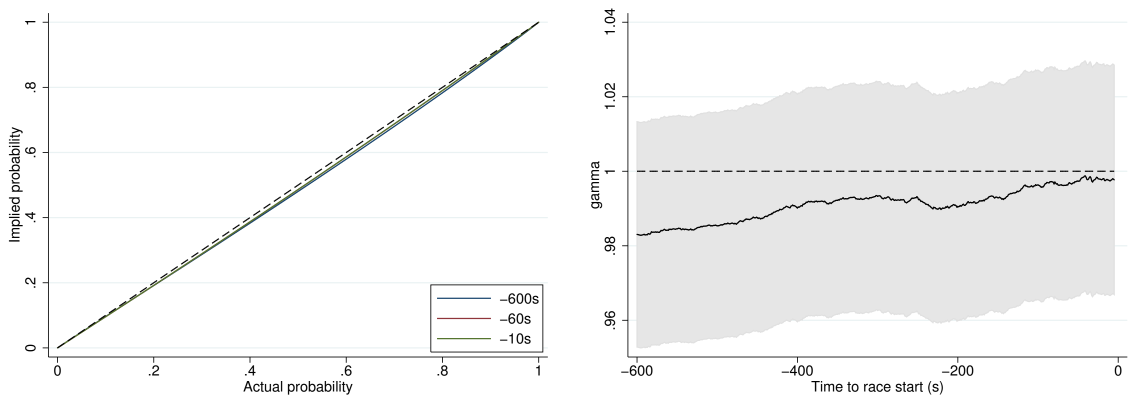


FIGURE 5.5: Estimated calibration curves before races start, ϕ_t^s.

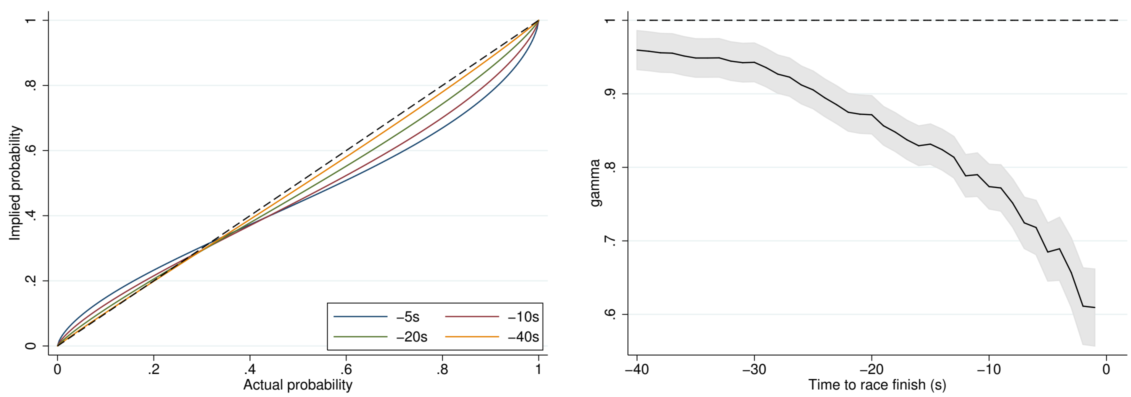


FIGURE 5.6: Estimated calibration curves before races finish, ϕ_t^f.

In order to understand potential mechanisms driving those results, we next estimate the same model while allowing the curvature parameter γ to capture different characteristics of the races. We specify

$$\gamma = \gamma_0 + \gamma X,$$

where X includes variables such as the number of horses finishing the given race, overall duration of the race, and pre-race aggregate trading volume. Since there is no FLB prior to races, we focus on timing before the finish. In this estimation we stack all observations at $t \in \{40, 20, 10, 5\}$ seconds before finish.

We include three dummies indicating $t \in \{20, 10, 5\}$ as well as number of finishing horses, total race duration (in seconds), and aggregated pre-race trading volumes (for each horse separately; in unit of 1,000 GBP). First, inclusion of those control variables (race duration and pre-race trading volume), do not influence the direction, magnitude, and significance of the dummies coding for the remaining

Table 5.4: Structure of curvature.

		(1)	(2)	(3)	(4)
δ:	Elevation	0.865	0.867	0.867	0.869
		(0.007)	(0.007)	(0.007)	(0.007)
γ_0:	Curvature at $t = 40$	0.919	0.949	0.930	0.967
		(0.021)	(0.014)	(0.009)	(0.024)
γ_1:	$\mathbf{1}\{t = 20\}$	−0.049	−0.050	−0.050	−0.049
		(0.011)	(0.011)	(0.011)	(0.011)
γ_2:	$\mathbf{1}\{t = 10\}$	−0.144	−0.145	−0.143	−0.144
		(0.017)	(0.017)	(0.017)	(0.017)
γ_3:	$\mathbf{1}\{t = 5\}$	−0.244	−0.245	−0.244	−0.245
		(0.022)	(0.022)	(0.022)	(0.022)
γ_4:	Number of finishing horses	1.57×10^{-4}			9.04×10^{-4}
		(1.74×10^{-3})			(1.75×10^{-3})
γ_5:	Race duration (s)		-1.92×10^{-4} *		-1.88×10^{-4} *
			(7.92×10^{-5})		(7.97×10^{-5})
γ_6:	Pre-race trading volume (1000 GBP)			-8.52×10^{-5} *	-8.19×10^{-5}
				(4.22×10^{-5})	(4.25×10^{-5})
# observations		28,8267	28,8267	28,8267	28,8267
# clusters		8,760	8,760	8,760	8,760

Notes: Clustered standard errors in parentheses. For γ_4, γ_5, and γ_6, a star indicates $p < 0.05$.

race time, which validates our previous estimation results (p-values for the null hypothesis H_0: $\gamma_0 + \gamma_i = 1$ as well as H_0: $\gamma_i = 0$, $i \in \{1, 2, 3\}$, are all less than 0.001).

In the in-play betting markets, thinking about multiple assets simultaneously under extreme time pressure is a cognitively (and operationally) demanding task. We include number of finishing horses in a given race as an explanatory variable to capture a potential relationship between FLB and limited attention as well as partition dependence. However, the estimated γ_4 is positive and not significantly different from 0 (see columns (1) and (4) in Table 5.4). This may imply that number of finishing horses does not influence the degree of FLB observed right before races finish per se, but the interpretation needs some caution. In particular, in this approach we are not distinguishing the number of horses in a given race and the number of horses "effectively competing" at the moment of our interests, even though traders stop trading on some horses before races finish when they judge that those horses have no chance of winning.

Next, we look at the effect of pre-race aggregated trading volume on curvature. The estimated γ_6 is negative but no significantly different from 0 after controlling for other race characteristics (compare columns (3) and (4) in Table 5.4). The larger pre-race trading volumes indicate higher liquidity in the markets. Using Intrade markets on future events, Page and Clemen's (2013) also find that trading volumes do not have a significant effect on calibration. Tetlock (2008) also find the calibration of TradeSports market prices to event (both financial and sporting) probabilities does not improve with increases in liquidity. other traders' knowledge and unwittingly bet against them, which can slow the response of prices to information.

5.4 Discussion

5.4.1 Magnitude of Estimated Curvature

We find that market-implied probabilities exhibit significant FLB during last 40 seconds of the races and the degree of the bias appears to be increasing as races approach the finish lines. Those results are evident in the inverse-*S* shape of the estimated probability weighting function. Although the curvatures are significantly different from 1 in a statistical sense, it is unclear how big the biases are. Here we compare estimated parameters in our dataset and those in the literature on experimental and empirical studies of non-expected utility theory.

Table 5.5 lists 17 sets of parameter estimates from 10 different studies (taken from the online supplementary material for Hsu et al. (2009) and extended by including more recent studies). Those studies all involve lottery choices, but each uses different elicitation methodologies, different domains (either gains or losses), and different estimation techniques. We thus need caution in comparing those estimates at face values, but it still provides useful information. [13]

The estimated curvature γ exhibits heterogeneity across studies, and the median of those 17 estimates is 0.60. Notice that estimated γ which exhibits the strongest bias in our dataset, namely, that derived from five seconds prior to goal, is 0.685 (Table 5.3, column (6)). This means that the curvature of the probability weighting function in our dataset, even though statistically significant, is much less pronounced compared to the majority of probability weighting functions estimated with lottery choice experiments (see Figure D.4).

We have slightly different picture in field evidence of probability weighting functions. Jullien and Salanié (2000) estimate several probability weighting functions using 34,443 flat horse race run in the U.K. between 1986 and 1995. They obtain $(\widehat{\gamma}, \widehat{\delta}) = (0.97, 0.88)$, which is close to linearity (in fact they cannot reject

[13]The online supplementary material for van de Kuilen and Wakker (2011) provides an extensive list of studies both in favor of and against inverse-*S* shape probability weighting functions, both under risk and uncertainty.

Table 5.5: Empirical studies on Lattimore et al. (1992) probability weighting function.

Study		Domain	Estimates γ	δ
Tversky and Fox (1995)		Gains	0.69	0.77
Wu and Gonzalez (1996)		Gains	0.68	0.84
Gonzalez and Wu (1999)		Gains	0.44	0.77
Abdellaoui (2000)		Gains	0.60	0.65
Abdellaoui (2000)		Losses	0.65	0.84
Etchart-Vincent (2004)	Small stakes	Losses	0.84	1.02
Etchart-Vincent (2004)	Large stakes	Losses	0.85	1.18
Stott (2006)		Gains	0.96	1.40
Fehr-Duda et al. (2006)	Female	Gains	0.47	0.74
Fehr-Duda et al. (2006)	Male	Gains	0.56	0.88
Fehr-Duda et al. (2006)	Female	Losses	0.47	1.10
Fehr-Duda et al. (2006)	Male	Losses	0.57	1.00
Hsu et al. (2009)		Gains	0.79	0.80
Booij et al. (2010)		Gains	0.62	0.77
Booij et al. (2010)		Losses	0.59	1.02
Bruhin et al. (2010)		Gains	0.38	0.93
Bruhin et al. (2010)		Losses	0.40	0.99
Median		All	0.60	0.88
		Gains	0.61	0.79
		Losses	0.59	1.02
This study	600 (s) before start		0.98	0.93
	60 (s) before start		1.00	0.95
	40 (s) before finish		0.96	0.94
	20 (s) before finish		0.87	0.87
	10 (s) before finish		0.77	0.80
	5 (s) before finish		0.69	0.78

expected utility). Page and Clemen (2013), the closest to the current study, use transaction data from 1,883 Intrade markets on future events including both political and sports markets and obtain $(\widehat{\gamma}, \widehat{\delta}) = (0.80, 0.94)$. The null hypothesis of linearity, $\gamma = 1$, is rejected at 1% significance level. Feess et al. (2014) use more than five million fixed-odds bets placed at the New Zealand Racing Board and find that $(\widehat{\gamma}, \widehat{\delta}) = (0.99, 1.28)$ for gains and $(\widehat{\gamma}, \widehat{\delta}) = (0.98, 0.81)$ for losses. They reject the null hypothesis of linearity thanks to large sample size, but curvature in the gain domain is not significantly different from that in the loss domain. Andrikogiannopoulou and Papakonstantinou (2015) use a novel individual-level trading data from a sports wagering market (their dataset cover 336 randomly selected traders over a five-year period). The median of population estimates is $(\widehat{\gamma}, \widehat{\delta}) = (0.91, 1.17)$. Those values imply that the probability weighting function is concave for most probabilities and has a slight inverse-*S* shape. There are not many studies investigating calibration of in-play implied probabilities. Hartzmark and Solomon (2012) is one exception, who obtain $(\widehat{\gamma}, \widehat{\delta}) = (0.76, 0.96)$ from NFL in-play betting prices at TradeSports. [14] Taken together, those estimated values from field dataset are close to our estimates from pre-race implied probabilities (see Figure D.4).

5.4.2 Explanations for the Favorite-Longshot Bias

A number of theories have been proposed to explain FLB. Ottaviani and Sørensen (2008) summarize them into seven major categories: (i) misestimation of probabilities, (ii) preference for risk, (iii) heterogeneous beliefs, (iv) market power by informed bettors, (v) market power by uninformed bookmakers, (vi) limited arbitrage by informed bettors, and (vii) timing of bets. Of those, we are particularly interested in (i), (ii), (iii), and to some extent (vii). We do not cover explanations based on market power or arbitrage here, primarily because those theories consider an environment with fixed-odds betting, while our data are from a betting exchange market. [15]

Weitzman (1965), who estimates utility function using over 12,000 horse races, suggests that local risk loving (i.e., convex utility function) is consistent with observed FLB. Quandt (1986) proves that FLB is a necessary condition for equilibrium in the parimutuel betting market with risk-loving traders. Jullien and Salanié (2000) later use 34,000 horse races in the U.K. and estimate expected utility model with constant absolute risk aversion (CARA) utility function. Their estimation result indicates risk loving. Next, they estimate cumulative prospect theory and find that utility function is convex, probability weighting function for gains is convex but not significant, and probability weighting function for losses

[14] Hartzmark and Solomon (2012) relate this finding with the *disposition effect*, the tendency of investors to sell stocks trading at a gain relative to purchase price, rather than stocks trading at a loss (Odean, 1998; Shefrin and Statman, 1985). Their reasoning goes as follows. When the price is above the pre-game price and there is positive news pushing up the price (including touchdowns and intercepts), there will be an excess supply, causing short-term negative returns and prices being pushed below their fundamental value. This will be followed by positive long-term returns as trading prices return to their true equilibrium probability. They find patterns of apparent overreaction followed by underreaction that depend on whether the price is above or below the pre-game price.

[15] For example, Shin (1991, 1992) explain FLB based on the response of an uninformed bookmaker to insider's private information.

is concave, a rejection of expected utility maximization. [16]

The earliest explanation for the FLB can be found in Griffith (1949), who argues that a simple psychology of overestimation of low probabilities can explain the bias. Using data from 6.4 million horse races started in the U.S. between 1992 and 2001, Snowberg and Wolfers (2010) conduct a crucial test of distinguishing preference for risk and probability misperceptions. Their innovation is in the use of the *exotic bets* (compound lotteries) to differentiate two theories, while most other studies focus only on the *win bets*. They find that misperceptions of probability explain FLB better than preferences for risk.

Ali (1977) proves that FLB can be explained with bettors that are risk-neutral expected utility maximizers but have heterogeneous risk perceptions and capital constraints. Based on this idea, Gandhi and Serrano-Padial (2015) show that differences in agents' beliefs lead to a pricing pattern consistent with FLB in a competitive market for Arrow-Debreu securities. Using data from 176,000 U.S. horse races (and assuming that bettors are risk neutral), they estimate that about 70% of the bettors have roughly correct beliefs and the remaining 30% have dispersed beliefs. [17] In a similar spirit, Ottaviani and Sørensen (2015) theoretically examine how the market price aggregates the bettors' posterior beliefs and how the equilibrium price reacts to information that becomes publicly available to all bettors in the environment with heterogeneous prior beliefs, common knowledge of information structure, and wealth effects. They find that the price underreacts to new information, which implies FLB: outcomes favored by the market occur more often than probability implied by market prices and, conversely, longshots win less frequently than the price indicates. They further show that wider dispersion of beliefs corresponds to more pronounced FLB.

In addition to those relatively standard set of explanations, Ottaviani and Sørensen (2009, 2010) advance an additional theoretical explanation based on information and timing of bets in parimutuel markets. In Ottaviani and Sørensen's (2009) model, a large number of privately informed bettors who share a common prior belief take simultaneous positions just before post time (i.e., simultaneous

[16]Jullien and Salanié (2000) estimate three functional forms for the probability weighting function, power function, a specification due to Cicchetti and Dubin (1994):

$$\frac{w(q)}{1-w(q)} = \left(\frac{q}{1-q}\right)^{\gamma}\left(\frac{q_0}{1-q_0}\right)^{1-\gamma},$$

where γ controls curvature and q_0 specifies the crossing point, and the one proposed by Lattimore et al. (1992). They obtain similar results from all of those specifications.

[17]Chiappori et al. (2012) estimate heterogeneity in preferences, rather than beliefs, from aggregated betting data.

TABLE 5.6: Potential explanations of the evolution of FLB.

Explanation	In-play horse race betting
Misestimation of probabilities	◯
Preference for risk	×
Heterogeneous beliefs	×
Timing of bets	×

move game). FLB arises because bettors are not allowed to condition their behavior on the final odds: they are statistically "surprised" ex post, by which favorite emerges when the results of the simultaneous betting are announced. Bettors would prefer to bet more on the revealed favorite and cancel bets on the revealed longshot, but simultaneous-move assumption prohibits them from doing so. Ottaviani and Sørensen (2010) extend the analysis and show that the direction and magnitude of FLB depend on the signal-to-noise ratio of private information in the market.

What insights do those explanations provide in interpreting our results on the observed evolution of FLB over time? We informally argue that misperception of probabilities is the only likely candidate of explanation to our findings (Table 5.6).

First thing to note is that stronger FLB is observed during the last 40 seconds of the races, the time window in which bettors' beliefs would become increasingly aligned. Thus, heterogeneous beliefs story a la Ali (1977) and Gandhi and Serrano-Padial (2015) cannot explain our findings.

Ottaviani and Sørensen (2009, 2010) obtain their results on FLB as an ex-post "surprise" mainly from the simultaneous-move assumption that implies inability of changing trades after observing market odds. It is clear that traders in betting exchanges are clearly not moving simultaneously and are able to make countertrades anytime if they wish to do so. Furthermore, information regarding which horse is currently favored (and which ones are longshots) are continually updated on the limit-order book, leaving no room for traders to experience "surprise." Therefore, this line of explanation is also unlikely to hold.

Preferences for risk is an often-cited explanation for FLB as described above, but this will not be the full story. Golec and Tamarkin (1998), for example, show that FLB is in fact consistent with risk averse traders with preferences for skewness. Camerer (1998) points out that anecdotal evidence on bettors' typical preference for a "dead heat," in which two horses finish the race almost exactly tied (and the track stewards cannot declare a single winner) and people who bet on either horse share the total pool of money, suggests that their wagering is not due to convex utility per se (if they have convex utility, they should prefer to flip a coin and either win alone or lose than to a dead heat declared).

Furthermore, if we try to explain our findings on the dynamics of FLB using risk-loving framework, it is required to assume that either (i) bettors become increasingly risk-loving over time or (ii) less risk-loving bettors gradually exit from the market and only risk-loving ones remain. Further elaboration of data analysis is necessary, although nontrivial given the aggregated nature of our data, to prove or disprove this line of explanation.

One fruitful avenue for future analysis is to test if difference in race types explains our finding on the dynamics of FLB. A *maiden* race is a race for horses that have never won before (including those who are racing for the first time) and a *non-maiden* race is a race for horses with a racing (and winning) history. Thus, there is much less information about horses in maiden races as compared to non-maiden races (Camerer, 1998). This analysis, if we are in fact able to observe differences, potentially serves as a test discriminating belief-based story and preference-based story, since the amount of information change beliefs but not preferences for risk. [18]

Misperception of probabilities is a likely explanation to our findings. In particular, we speculate that an affect-based story proposed by Rottenstreich and Hsee (2001), who argue that the probability weighting function becomes more inverse-*S* shaped for lotteries involving affect-rich than affect-poor outcomes, is a key driver of the dynamics of FLB. [19] They propose that some outcomes are relatively affect-rich and others are relatively affect-poor, even when the monetary values associated with those outcomes are controlled. An example of affect-rich outcomes is a $100 coupon redeemable for payment toward dinner for two at a fancy French restaurant. They claim that this coupon is likely to evoke relatively strong emotional reactions compared with a $100 coupon redeemable for payment toward one's phone bill.

Horse racing is a particularly affect-rich environment. [20] Wulfert et al. (2005),

[18]Exploiting this natural variation in race types, Gandhi and Serrano-Padial (2015) in fact observe that the magnitude of FLB is much more pronounced in maiden compared to non-maiden races. They do not look at the dynamics, since their data are from parimutuel markets.

[19]Barberis (2013) points out that "it is interesting to think about the psychological foundations of probability weighting. Tversky and Kahneman (1992) and Gonzalez and Wu (1999) offer an interpretation based on the principle of diminishing sensitivity, while Rottenstreich and Hsee (2001) give an affect-based interpretation. More recently, Bordalo et al. (2012) argue that salience is an important driver of probability weighting."

[20]One source of affective experiences may be "suspense" and "surprise" from dynamics of

for example, measure experimental subjects' heart rates while showing them a videotaped horse race with an exciting neck-to-neck finish. Half of the subjects bet $1 for a chance of winning $7 if they picked the winning horse while the other half only predicted the winning horse. They find that subjects with a chance to win money exhibited greater heart rate elevations and reported more subjective excitement while watching the race compared with those who did not wager. Furthermore, Coventry and Norman (1997) find that the average heart rate was at a peak during the last 30 seconds of the race. Those observations support our view that increasing affective reactions, especially during the late in the races, are partly responsible for the observed dynamics of FLB.

One direction for future research is to test this affect-based interpretation directly, using a controlled laboratory experiment in which subjects make trades in a short-horizon laboratory prediction market (as in Sonnemann et al., 2013) while their physiological reactions, such as skin conductance responses, are recorded (as in Kang et al., 2012).

5.5 Conclusion

We have shown in this study that the betting data from an online exchange exhibit a dynamically increasing pattern of the favorite-longshot bias (FLB). More precisely, we find that betting odds prior to races provide well-calibrated probabilistic forecasts for winning horses, while market odds exhibit systematic FLB and the degree of bias increases as races approach to the finish line.

These observations are interesting for following reasons. First, a substantial amount of evidence in the literature (Table D.1) supports the existence of FLB in many markets in several countries. Thus, the non-existence of FLB prior to races in Betfair is somewhat surprising and is implicating the relative efficiency of this new form of market, online betting exchanges, compared to traditional systems such as parimutuel betting and bookmakers. Second, an increasing degree of FLB over time is a new empirical fact in the literature. The availability of "in-play" betting data makes this analysis possible.

What explanations can we offer? We argue, albeit informally, that our "dynamic" FLB is hard to reconcile with many of the leading theories behind the "static" FLB such as preferences for risk and heterogeneous beliefs. One possibility is that bettors distort probabilities due to affective reactions. There is

subjective beliefs (Ely et al., 2015).

an experimental evidence suggesting that the probability weighting function becomes more inverse-*S* shaped for lotteries involving affect-rich than affect-poor outcomes (Rottenstreich and Hsee, 2001). Further analyses are necessary to validate this line of explanation. For example, comparing the evolution of FLB across several categories of races (such as competitive versus non-competitive), classified based on text-mining methods applied to race descriptions on news media, is one potential approach.

Another possibility is the market environment itself. In this regard, Ottaviani and Sørensen's (2015) dynamic market of Arrow-Debreu securities in which information arrives to the market sequentially is a good starting point for a theoretical investigation of behavior in online betting exchanges.

Appendix A

Appendix to Chapter 2

A.1 Proof of Theorem 3

The proof that GTD rationality is equivalent to SAR-GTD is identical to the result in Echenique and Saito (2015) with the changes of T to S and $\{D(t)\}_{t\in T}$ to $\{\mu_s\}_{s\in S}$. In the following, we show the proofs for MTD and TSU.

A.1.1 MTD

The proof that SAR-MTD is equivalent to MTD rationality requires the following modification of the argument in Echenique and Saito (2015).

To see that SAR-MTD is necessary, let $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^n_{i=1}$ be a sequence under the conditions of the axiom. We present the proof under the assumption that u is differentiable, but it is straightforward to use the concavity and the corresponding monotonicity of the superdifferential of u, as we did in the proof of Theorem 1. The first-order condition is $D(t)u'(x^k_t) = \lambda^k p_t$. Then,

$$
\begin{aligned}
1 &\geq \prod_{i=1}^{n} \frac{u'(x^{k_i}_{t_i})}{u'(x^{k'_i}_{t'_i})} \\
&= \prod_{i=1}^{n} \frac{\lambda^{k_i} D(t'_i) p^{k_i}_{t_i}}{\lambda^{k'_i} D(t_i) p^{k'_i}_{t'_i}} = \prod_{i=1}^{n} \frac{D(t'_i) p^{k_i}_{t_i}}{D(t_i) p^{k'_i}_{t'_i}} = \prod_{i=1}^{n} \frac{D(t'_i)}{D(t_i)} \prod_{i=1}^{n} \frac{p^{k_i}_{t_i}}{p^{k'_i}_{t'_i}} = \prod_{i=1}^{n} \frac{D(t'_{\pi(i)})}{D(t_i)} \prod_{i=1}^{n} \frac{p^{k_i}_{t_i}}{p^{k'_i}_{t'_i}}.
\end{aligned}
$$

Since $t_i \geq t'_{\pi(i)}$ and D is decreasing it follows that $D(t'_{\pi(i)})/D(t_i) \geq 1$. Therefore we must have that $\prod_{i=1}^n p^{k_i}_{t_i}/p^{k'_i}_{t'_i} \leq 1$.

For the proof of sufficiency, consider the setup in the proof of Theorem 1 of Echenique and Saito (2015). Note that the GTD model is the same as the model of subjective expected utility. Let A and B be the matrices as constructed in the proof of Theorem 1 of Echenique and Saito (2015). We need to add rows to B to reflect that $D(t') \geq D(t)$ when $t \geq t'$. To put it precisely, we need an additional row for each pair t, t' such that $t \geq t'$. In the row, we have -1 in the column of t and have 1 in the column of t'. Remember in the matrix A, we have a column for each $t \in T$, as we do for each $s \in S$ in Echenique and Saito (2015). In the solution to the dual, we follow the steps of the proof until we construct a balanced sequence $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^n_{i=1}$. Such a sequence corresponds to a decomposition of A^* into pairs of rows $(r_i, r'_i)^n_{i=1}$ in which r_i is original and r'_i is converted.

Now consider the column corresponding to t. In the characterization of GTD (and of SEU), the entries in that column are all zero. For MTD, the entries of B are no longer all zero in that column. The sum of the rows of $A^* + B^*$ equals zero. As usual we can eliminate pairs of rows of B such that $1_{t'} - 1_t + 1_t - 1_{t''} = 1_{t'} - 1_{t''}$. So in the matrix B^* all the entries in the column for t will be of the same sign. Let us say that they are all -1.

Recall that each row of A is identified with a pair (k, t). If r is a row say that t *appears* in row r if there is k such that r is the row associated with (k, t). In A^* we may have multiple copies of the same row.

Since the rows of $A^* + B^*$ is zero, the number of times that t appears in an original row minus the number of times that t appears in a converted row equals the number of rows in B^* in which t has a -1. Since we have assumed that there are -1s in the column for t, then there are more original rows in which t appears than converted rows. Let $\pi(i)$ be an arbitrary original row, for each converted row in which t appears. This defines π for all converted rows in which t appears.

There are then original rows in which t appears that are not the image through π of some converted row. For each such row ρ of A^* there is some -1 in B^*, as $A^* + B^* = 0$. Let $\sigma(\rho)$ be the the row of B^* with -1.

The construction is the same for columns t' in which B^* only has 1. Let σ be defined in the same way. This defines π for some rows. For the remaining rows, define π as follows. Let ρ be original, such that t appears in ρ, and t is not in the image of π. There is t' in row $\sigma(\rho)$ with $t' \leq t$ (the row $\sigma(\rho)$ is $1_{t'} - 1_t$). There is a unique converted row ρ' with $\sigma(\rho) = \sigma(\rho')$, a row in which t' appears. So let $\pi(\rho) = \rho'$. This defines π for all rows.

A.1.2 TSU

The proof that SAR-TSU is equivalent to TSU rationality is similar to the proof of Theorem 1. In the following, we explain the differences.

Lemma 13. *Let $(x^k, p^k)_{k=1}^K$ be a dataset. The following statements are equivalent:*

1. $(x^k, p^k)_{k=1}^K$ *is TSU rational.*
2. *There are strictly positive numbers v_t^k and λ^k for $t = 0, \ldots, T$ and $k = 1, \ldots, K$, such that*
$$v_t^k = \lambda^k p_t^k \text{ and } x_t^k > x_t^{k'} \Longrightarrow v_t^k \leq v_t^{k'}.$$

The proof of Lemma 13 is very similar to the proof of Lemma 1 and omitted.

To see that SAR-TSU is necessary, let $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$ be a sequence under the conditions of the axiom. We present the proof under the assumption that u_t is differentiable, but it is straightforward to use the concavity and the corresponding monotonicity of the superdifferential of u_t, as we did in the proof of Theorem 1. The first-order condition is $u_t'(x_t^k) = \lambda^k p_t$. Since $t_i = t_i'$ for each i, we obtain

$$1 \geq \prod_{i=1}^n \frac{u_{t_i}'(x_{t_i}^{k_i})}{u_{t_i}'(x_{t_i}^{k_i'})} = \prod_{i=1}^n \frac{\lambda^{k_i} p_{t_i}^{k_i}}{\lambda^{k_i'} p_{t_i}^{k_i'}} = \prod_{i=1}^n \frac{\lambda^{k_i}}{\lambda^{k_i'}} \prod_{i=1}^n \frac{p_{t_i}^{k_i}}{p_{t_i}^{k_i'}} = \prod_{i=1}^n \frac{p_{t_i}^{k_i}}{p_{t_i}^{k_i'}},$$

where the last equality holds because each k appears as k_i' the same number of times it appears as k_i.

In the following, we prove the sufficiency. The outline of the proof is the same as in the proof of Theorem 1.

Lemma 14. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-TSU. Suppose that $\log(p_t^k) \in \mathbf{Q}$ for all k and t. Then there are numbers v_t^k, λ^k, β, δ, for $t \in T$ and $k \in K$ satisfying* (2) *in Lemma 13.*

Lemma 15. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-TSU. Then for all positive numbers $\bar{\varepsilon}$, there exists $q_t^k \in [p_t^k - \bar{\varepsilon}, p_t^k]$ for all $t \in T$ and $k \in K$ such that $\log q_t^k \in \mathbf{Q}$ and the dataset $(x^k, q^k)_{k=1}^k$ satisfy SAR-TSU.*

Lemma 16. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-TSU. Then there are numbers v_t^k and λ^k for all $t \in T$ and $k \in K$ satisfying* (2) *in Lemma 13.*

Lemma 15 and 16 hold as in the proof for Theorem 1.

A.1.3 Proof of Lemma 14

We linearize the equation in system (2) of Lemma 13. The result is:

$$\log v_t(x_t^k) - \log \lambda^k - \log p_t^k = 0, \tag{A.1}$$

$$x_t^k > x_t^{k'} \Longrightarrow \log v_t(x_t^k) \leq \log v_t(x_t^{k'}). \tag{A.2}$$

In the system comprised by (A.1) and (A.2), the unknowns are the real numbers λ^k and $\log v_t^k$ for all $k = 1, \ldots, K$ and $t = 1, \ldots, T$.

We shall define a matrix A such that there are positive numbers v_t^k and λ^k, the logs of which satisfy equation (A.1) if and only if there is a solution $u \in \mathbf{R}^{K\times(T+1)+K+1}$ to the system of equations

$$A \cdot u = 0,$$

and for which the last component of u is strictly positive.

Let A be a matrix with $K \times (T+1)$ rows and $K \times (T+1) + K + 1$ columns. The matrix A is similar to the matrix A defined in the proof of Theorem 1, only the difference here is that we no longer have the δ-column. Thus, matrix A looks as follows:

$$\begin{array}{c} \\ \vdots \\ (k,t) \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccccccc} (1,0) & \cdots & (k,t) & \cdots & (K,T) & 1 & \cdots & k & \cdots & K & p \end{array} \\ \left[\begin{array}{ccccc|ccccc|c} \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\ \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \end{array}\right]. \end{array}$$

Consider the system $A \cdot u = 0$. If there are numbers solving equation (A.1), then these define a solution $u \in \mathbf{R}^{K\times(T+1)+K+1}$ for which the last component is 1. If, on the other hand, there is a solution $u \in \mathbf{R}^{K\times(T+1)+K+1}$ to the system $A \cdot u = 0$ in which the last component is strictly positive, then by dividing through by the last component of u we obtain numbers that solve equation (A.1).

In the second place, we write the system of inequality (A.2) in matrix form. Let B be a matrix with $K \times (T+1) + K + 1$ columns. Define B as follows: one row for every pair (k,t) and (k',t) with $x_t^k > x_t^{k'}$; in the row corresponding to (k,t) and (k',t) we have zeroes everywhere with the exception of a -1 in the column for (k,t) and a 1 in the column for (k',t).

In the third place, we have a matrix E that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and $K \times (T+1) + K + 1$ columns. It has zeroes everywhere except for 1 in the last column.

To sum up, there is a solution to system (A.1) and (A.2) if and only if there is a vector $u \in \mathbf{R}^{K\times(T+1)+K+1}$ that solves the system of equations and linear inequalities

$$(S1): \; A \cdot u = 0, \; B \cdot u \geq 0, \; E \cdot u \gg 0.$$

The entries of A, B, and E are integer numbers, with the exception of the last column of A. Under the hypothesis of the lemma we are proving, the last column consists of rational numbers.

By Lemma 4, then, there is such a solution u to $S1$ if and only if there is no vector (θ, η, π) that solves the system of equations and linear inequalities

$$(S2): \; \theta \cdot A + \eta \cdot B + \pi \cdot E = 0, \; \eta \geq 0, \; \pi > 0.$$

In the following, we shall prove that the non-existence of a solution u implies that the data must violate SAR-TSU. Suppose then that there is no solution u and let (θ, η, π) be a rational vector as above, solving system $S2$.

By multiplying (θ, η, π) by any positive integer we obtain new vectors that solve $S2$, so we can take (θ, η, π) to be integer vectors.

For convenience, we transform the matrices A and B using θ and η. We now transform the matrices A and B based on the values of θ and η, as we did in the proof of Theorem 1. Let us define a matrix A^* from A and B^* from B, as we did in the proof of Theorem 1. We can prove the same claims (i.e., Claims 2, 3, 4, 5, and 6) as in the proof of Theorem 1. The proofs are the same and omitted. In particular, we can show that there exists a sequence of pairs $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^{n^*}$ that satisfies (1) in SAR-TSU. Moreover, by the definition of B matrix, we have $t_i = t_i'$ because in matrix B we have $z >^i z'$ if there exist $t \in T$ and $k, k' \in T$ such that there exist $x_t^k = z$ and $x_t^{k'} = z'$. Moreover, as in Claim 7, we can show that in the sequence $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^{n^*}$, each k appears k_i the same number of times it appears as k_i'. Finally, we can show that $\prod_{i=1}^{n^*} p_{t_i}^{k_i} / p_{t_i'}^{k_i'} > 1$, which finishes the proof of Lemma 14 as the sequence $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^{n^*}$ would then exhibit a violation of SAR-TSU. The proof is the same as in the proof of Theorem 1 and omitted.

A.2 Implementing Revealed Preference Tests

This section presents a method to implement the revealed preference tests for time discounting models using Matlab R2014b (MathWorks). We use Andreoni and Sprenger's (2012a) experimental choice data as the model case, but our method is applicable to other empirical/experimental data sets.

Dataset. Subjects in the Andreoni and Sprenger's (2012a) experiment completed 45 intertemporal decisions with varying starting dates τ, delay lengths d, and gross interest rates $a_{\tau+d}/a_\tau$ and, in particular, they complete 5 decision problems for each pair of (τ, d). See Figure A.1 for an illustration of budgets. For each subject, the decision in every trial is characterized by a tuple $(\tau, d, a_\tau, a_{\tau+d}, c_\tau)$ where c_τ is the number of tokens allocated to sooner payment.

The following figure illustrates the budgets faced by the subjects in AS's experiment, fixing one time frame at (τ, d).

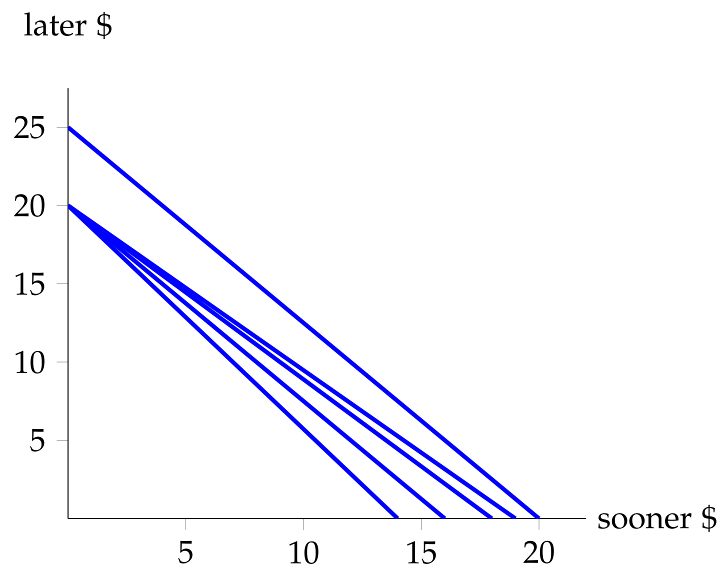


FIGURE A.1: An illustration of the CTB design in Andreoni and Sprenger (2012a). Budget sets are represented in blue lines, fixing one time frame at $(\tau, d) = (0, 35)$.

In order to rewrite our data in price-consumption format as in the theory, we set prices $p_\tau = 1 + r = a_{\tau+d}/a_\tau$ and $p_{\tau+d} = 1$ (normalization), and define consumptions $x_\tau = c_\tau \cdot a_\tau$ and $x_{\tau+d} = (100 - c_\tau) \cdot a_{\tau+d}$. This gives us a dataset $(x^k, p^k)_{k=1}^{45}$.

As we explained in Section 2.5, we implicitly set prices of consumption in periods that were not offered to a subject as very high in order to ensure that consumption is zero. The idea is as follows. Think of EDU for concreteness. We use first-order conditions, so that we are looking for a rationalizing u and δ such that $\delta^t u'(x_t^k) = \lambda^k p_t^k$ if $x_t^k > 0$ and $\delta^t u'(x_t^k) \leq \lambda^k p_t^k$ if $x_t^k = 0$. In setting up such a system of equations we can ignore the t that was not offered to the agents in trial k. Then whatever u we construct will have a finite derivative $u'(0)$. Therefore, we can set p_t^k to be high enough so that the agent finds it optimal to consume $x_t^k = 0$. By this argument it is clear that one can ignore the (zero) consumption in the periods that were not offered in trial k, as we think of consumption in those periods as prohibitively expensive. This is of course consistent with the fact that AS did not offer subjects any consumption in those periods; consumption in those periods is infeasible. The set of time periods we are looking at is thus $T = \{0, 7, 35, 42, 70, 77, 98, 105, 133\}$.

We are able to check whether a given dataset is consistent with TSU, GTD, MTD, QHD, PQHD, or EDU, by solving the corresponding linear programming problem. The construction of linear programming problems closely follows the argument in the proofs of Theorems 1, 2, and 3. In particular, the key to this procedure is to set up a system of linear inequalities of the form:

$$S: \begin{cases} A \cdot u = 0 \\ B \cdot u \geq 0 \\ E \cdot u > 0 \end{cases},$$

which, in the case of EDU for example, is a matrix form of the linearized system:

$$\begin{aligned} \log v(x_t^k) + t \log \delta - \log \lambda^k - \log p_t^k = 0, \\ x > x' \Longrightarrow \log v(x') \geq \log v(x), \\ \log \delta \leq 0. \end{aligned}$$

A system of linear inequalities. We now construct three key ingredients of the system, matrices A, B, and E, starting from those necessary for testing EDU. The first matrix A looks as follows:

$$\begin{array}{c} \\ \vdots \\ (k,t) \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} (1,0) & \cdots & (k,t) & \cdots & (45,133) & \delta & 1 & \cdots & k & \cdots & 45 & p \end{array} \\ \left[\begin{array}{ccccc|c|ccccc|c} \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\ \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \end{array} \right]. \end{array}$$

Since we can ignore the t that was not offered to the agents in trial k, the matrix has $45 \times 2 = 90$ rows and $45 \times 2 + 1 + 45 + 1 = 137$ columns. In the row corresponding to (k, t) the matrix has zeroes everywhere with the following exceptions: it has a 1 in the column for (k, t), it has a t in the δ column, it has a -1 in the column for k, and $-\log p_t^k$ in the very last column. This finalizes the construction of A.

Next, we construct matrix B that has 137 columns and there is one row for every pair (k, t) and (k', t') with $x_t^k > x_{t'}^{k'}$. In the row corresponding to (k, t) and (k', t') we have zeroes everywhere with the exception of a -1 in the column for (k, t) and a 1 in the column for (k', t'). Finally, in the last row, we have zeroes everywhere with the exception of a -1 at 91st column. We shall refer to this last row as the δ-row.

Finally, we prepare a matrix that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and 137 columns. It has zeroes everywhere except for 1 in the last column.

Constructing matrices for other tests. In order to test models other than EDU, we need to modify matrices A, B, and E appropriately.

For the QHD test, we insert another column capturing the present/future bias parameter β, which we shall refer to the β-column. Therefore, three matrices A, B, and E have $45 \times 2 + 1 + 1 + 45 + 1 = 138$ columns. In the row corresponding to (k, t) of the matrix A, the β-column has a 1 if $t > 0$ and a 0 if $t = 0$, indicating "now" or "future".

$$
\begin{array}{c} \\ \vdots \\ (k,t{=}0) \\ (k,t'{>}0) \\ \vdots \end{array}
\begin{array}{c}
\begin{array}{cccccc|c|c|ccccc|c}
(1,0) & \cdots & (k,t) & (k,t') & \cdots & (45,133) & \beta & \delta & 1 & \cdots & k & \cdots & K & p
\end{array} \\
\left[\begin{array}{cccccc|c|c|ccccc|c}
\vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & 1 & 0 & \cdots & 0 & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\
0 & \cdots & 0 & 1 & \cdots & 0 & 1 & t' & 0 & \cdots & -1 & \cdots & 0 & -\log p_{t'}^k \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots
\end{array}\right]
\end{array} .
$$

The construction of matrix B for testing general QHD is the same as above (although the size is now different). For the PQHD test, we add the β-row which has zeroes everywhere except -1 in the β-column to capture $\beta \leq 1$.

For the MTD and GTD tests, we have 9 columns capturing time-varying discount factors $D(t)$'s:

$$\begin{array}{c} \\ \vdots \\ (k,t) \\ \vdots \end{array}
\begin{array}{c}
\begin{array}{ccccc|ccc|ccccc|c}
(1,0) & \cdots & \tilde{x}_\ell & \cdots & (45,133) & \cdots & D(t) & \cdots & 1 & \cdots & k & \cdots & 45 & p
\end{array} \\
\left[\begin{array}{ccccc|ccc|ccccc|c}
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 1 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots
\end{array}\right]
\end{array} .$$

In the matrix B, we add rows

$$\begin{array}{c}
\begin{array}{ccccc|cccc|ccccc|c}
(1,0) & \cdots & (k,t) & \cdots & (45,133) & \cdots & D(t) & D(t+1) & \cdots & 1 & \cdots & k & \cdots & 45 & p
\end{array} \\
\left[\begin{array}{ccccc|cccc|ccccc|c}
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & -1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots
\end{array}\right]
\end{array}$$

in testing MTD to impose the monotonicity restriction on $D(t)$'s.

The matrix A for testing TSU is similar to that appears in testing EDU. The difference is that we no longer have the δ-column:

$$\begin{array}{c} \\ \vdots \\ (k,t) \\ \vdots \end{array}
\begin{array}{c}
\begin{array}{ccccc|ccccc|c}
(1,0) & \cdots & (k,t) & \cdots & (K,T) & 1 & \cdots & k & \cdots & K & p
\end{array} \\
\left[\begin{array}{ccccc|ccccc|c}
\vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & 1 & \cdots & 0 & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\
\vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots
\end{array}\right]
\end{array} .$$

Next, we construct B as follows: one row for every pair (k,t) and (k',t) with $x_t^k > x_t^{k'}$; in the row corresponding to (k,t) and (k',t) we have zeroes everywhere with the exception of a -1 in the column for (k,t) and a 1 in the column for (k',t).

Solve the system. Our task is to check if there is a vector u that solves the following system of linear inequalities corresponding to a model M

$$S_M : \begin{cases} A \cdot u = 0 \\ B \cdot u \geq 0 \\ E \cdot u > 0 \end{cases} .$$

If there is a solution u to this system, we say that the dataset is M rational.

We use the function `linprog` in the Optimization Toolbox of Matlab to find a solution. More precisely, we translate the systems of linear inequalities S_M into constraints in a linear programming problem and solve

$$LP_M : \begin{cases} \min & z \cdot u \\ \text{s.t.} & A \cdot u = 0 \\ & -B \cdot u \leq 0 \\ & -E \cdot u < 0 \end{cases},$$

where z is a zero vector.

It is not possible, however, to specify strict inequality constraints in `linprog`. As an alternative, we find a solution u that has 1 in the last element, i.e., $u_p = 1$. In other words, we solve a normalized version of the problem,

$$LP'_M : \begin{cases} \min & z \cdot u \\ \text{s.t.} & A \cdot u = 0 \\ & -B \cdot u \leq 0 \\ & u_p = 1 \end{cases},$$

where z is a zero vector as above. Here, the constraint $E \cdot u > 0$ is omitted since it is automatically satisfied by our normalization $u_p = 1$.

If the given dataset is EDU rational, we can recover upper and lower bounds of the daily discount factor consistent with the observed choice data. Remember that we include the δ-row in B. The constraint $B \cdot u \geq 0$ then implies that the 91st element of any solution u^* of LP'_M, called u^*_δ, captures the daily discount factor. To be more precise, we can recover the daily discount factor δ by $\exp(u^*_\delta)$ since we normalize u^*_p to be 1. Therefore, a solution (if any) of LP'_M in which the 91st element of z is 1 and 0 elsewhere suggests an lower bound of δ and a solution (if any) of LP'_M in which the 91st element of z is -1 and 0 elsewhere suggests an upper bound of δ. In a similar manner, we can recover bounds of present/future biasedness β.

A.3 Ground Truth Analysis: Test Performance and Parameter Recovery

We assess the basic performance of our revealed preference tests using simulated choices. As in Andreoni and Sprenger (2012a), we assume a decision maker has a utility function (CRRA with quasi-hyperbolic discounting) of the form:

$$U(x_0, \ldots, x_T) = \frac{1}{\alpha} x_0^\alpha + \beta \sum_{t \in T \setminus \{0\}} \frac{1}{\alpha} \delta^t x_t^\alpha.$$

We simulate synthetic subjects' choice data in Andreoni and Sprenger's (2012a) environment (i.e., time frames and budgets are identical to those actual subjects faced in their experiment) under all combinations of parameters

$$\alpha \in \{0.8, 0.82, \ldots, 1\}, \;\; \delta \in \{0.95, 0.951, \ldots, 1\}, \;\; \beta \in \{0.8, 0.82, \ldots, 1.2\},$$

resulting the total of 11,781 such synthetic subjects. We then perform our revealed preference tests, in particular, tests for EDU and QHD rationality, and ask following questions: (i) do our tests correctly identify EDU or QHD rational datasets?, and (ii) can our tests recover "true" underlying model parameters?

A few remarks are in order. (1) For some parameter specifications, it is possible that the slope of (linear) indifference curves coincide with those of budget lines. This happens 21 times when $(\alpha, \delta) = (1, 1)$. [1] If the slope of indifference curve coincides with the budget line (i.e., every point on the budget yields the same level of utility), we randomly pick one point from the budget as the optimal choice as a tie-breaking rule. (2) In order to avoid the rounding issue in Matlab, we treat numbers less than 10^{-10} to be 0. In other words, if the predicted allocation is sufficiently close to a corner, we treat it as a corner choice. (3) Unlike Andreoni and Sprenger's (2012a) original experiment where subjects made choices from "discrete" budget sets by allocating 100 tokens, we allow simulated choices to be at any point on the continuous budget lines. We also prepare another set of simulated choices (with the same set of parameters) which mimic behavior of the Andreoni and Sprenger's (2012a) experimental subjects for the purpose of comparison.

Test results. The results are presented in Table A.1. We first look at our baseline simulation in which choices were made from continuous budget sets. Of the 11,781 synthetic subjects, 3,950 (33.5%) passed the EDU test and 11,781 (100.0%) passed the QHD test.

[1] For example, consider the case when $(\alpha, \delta, \beta) = (1, 1, 0.8)$ and $(1, 1, 0.9)$. Since the utility function has the form $x_\tau + \beta x_{\tau+d}$ when $\tau = 0$, indifference curve coincides with budget line when prices are 1.11 or 1.25. Another possibility is in the time frame $(\tau, d) = (7, 70)$, where the price of 1 (tokens allocated to sooner and later payments have the same exchange rate) is offered. In this case, indifference curve coincides with budget line as long as $(\alpha, \delta) = (1, 1)$.

TABLE A.1: Test results using simulated choice data from continuous budgets (top panel) and discrete budgets (bottom panel).

	Parameters			
Continuous budget	$\alpha = 1$	$\alpha < 1$ $\beta = 1$	$\alpha < 1$ $\beta \neq 1$	Total
No interior choice	1,050	38	700	1,788
Pass EDU	939	510	2,501	3950
Pass QHD	1,071	510	10,200	11,781
Sample size	1,071	510	10,200	11,781

	Parameters			
Discrete budget	$\alpha = 1$	$\alpha < 1$ $\beta = 1$	$\alpha < 1$ $\beta \neq 1$	Total
No interior choice	1,050	252	4,746	6,048
Pass EDU	939	510	6,913	8,362
Pass QHD	1,071	510	8,319	9,900
Sample size	1,071	510	10,200	11,781

We then split the sample into three groups. The first group of subjects have the linear utility function ($\alpha = 1$). They made no interior choices (except for the knife edge case described above), and 939 of them passed the EDU test. The second group of subjects have nonlinear utility and no present/future bias ($\alpha < 1$, $\beta = 1$). They all passed the EDU test (and hence the QHD test, too), as expected. The third group of subjects have nonlinear utility and present/future bias ($\alpha < 1$, $\beta \neq 1$). We find that 2,501 of them passed the EDU test, even though their underlying preferences were strictly present/future biased.

The bottom panel of Table A.1 presents the results with simulated data when choices are assumed to be on the discrete points on the budget lines. As one can imagine, the number of synthetic subjects who make no interior choices increases and accordingly the pass rate for the EDU test increases from 33.5% to 71.0%. We also find that "perturbations" induced by discretization of budget sets is powerful enough for some of the subjects to become QHD non-rational.

Parameter recovery. Next we investigate how precise we can recover underlying preference parameters using our revealed preference tests. Remember that the revealed preference tests boil down to linear programming problems. As we describe in Section A.2, we can find bounds of daily discount factor δ or present-biasedness β, which can be used to rationalize the observed choice data.

In this exercise we restrict our attention to the case of choices from continuous budgets.

1. We look at subset of synthetic subjects who have non-linear instantaneous utility ($\alpha < 1$), no present/future bias ($\beta = 1$), and pass the EDU test. We exclude synthetic subjects who make no interior allocation from this sample. There are 510 subjects in this category. Of those, 304 have $(0, \bar{\delta}_i]$ for some $\bar{\delta}_i < 1$, 10 have $[\underline{\delta}_i, 1]$ for some $\underline{\delta}_i > 0$, and 196 have $[\underline{\delta}_i, \bar{\delta}_i]$ for some combination of $\bar{\delta}_i < 1$ and $\underline{\delta}_i > 0$. Furthermore, within the last category of subjects, the true underlying discount factors are always covered by the ranges $[\underline{\delta}_i, \bar{\delta}_i]$. See Figure A.2, left panel.

2. We focus on those who have non-linear instantaneous utility ($\alpha < 1$), present or future bias ($\beta \neq 1$), and pass the QHD test. We exclude synthetic subjects who make no interior allocation from this sample. There are 10,200 subjects in this category. Of those, 9,737 have $(0, \bar{\delta}_i]$ for some $\bar{\delta}_i < 1$, 43 have $[\underline{\delta}_i, 1]$ for some $\underline{\delta}_i > 0$, and 420 have $[\underline{\delta}_i, \bar{\delta}_i]$ for some combination of $\bar{\delta}_i < 1$ and $\underline{\delta}_i > 0$. Within the last category of subjects, the true underlying discount factors are covered by the ranges $[\underline{\delta}_i, \bar{\delta}_i]$ in 362 cases (86.2%). See Figure A.2, right panel. Next we turn to present bias. Of a total 10,200 in this sample, 700 have $(0, \infty)$ (i.e., any value is possible), 76 have $(0, \bar{\beta}_i]$ for some $\bar{\beta}_i > 0$, 60 have $[\underline{\beta}_i, \infty)$ for some $\underline{\beta}_i > 0$, and 9,364 have $[\underline{\beta}_i, \bar{\beta}_i]$ for some combination of $\bar{\beta}_i, \underline{\beta}_i > 0$. Within the last category of subjects, the true underlying present biases are covered by the ranges $[\underline{\beta}_i, \bar{\beta}_i]$ in 9,354 cases (99.9%). See Figure A.3.

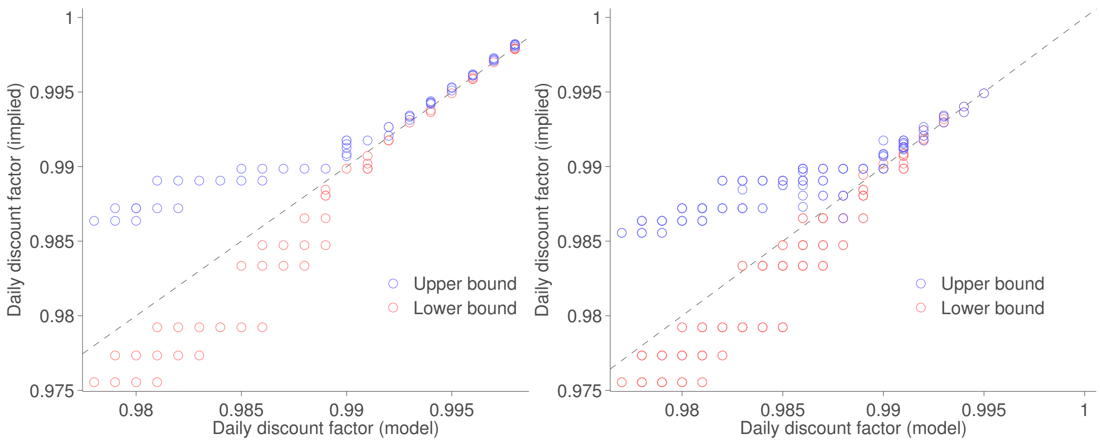


FIGURE A.2: Upper and lower bounds of daily discount factor implied by the revealed preference test. Each synthetic subject has one pair of a blue circle (upper bound) and a red circle (lower bound). (Left) The sample is 196 synthetic subjects who (i) have non-linear utility ($\alpha < 1$) and no present/future bias ($\beta = 1$), (ii) pass the EDU test, and (iii) have recovered range $[\underline{\delta}_i, \bar{\delta}_i]$ with $\underline{\delta}_i > 0$ and $\bar{\delta}_i < 1$. (Right) The sample is 420 synthetic subjects who (i) have non-linear utility ($\alpha < 1$) and present/future bias ($\beta \neq 1$) and (iii) have recovered range $[\underline{\delta}_i, \bar{\delta}_i]$ with $\underline{\delta}_i > 0$ and $\bar{\delta}_i < 1$. The dotted line represents the 45-degree line.

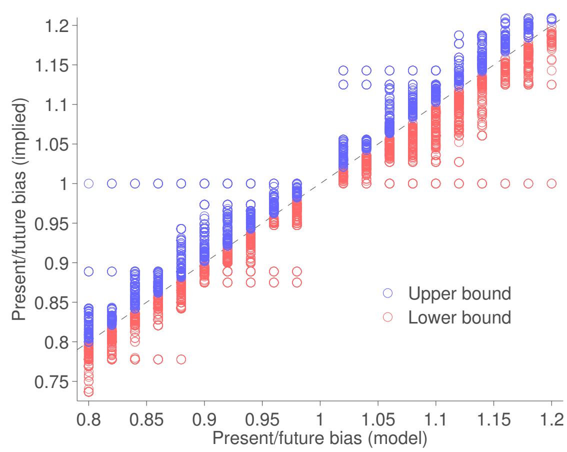


FIGURE A.3: Upper and lower bounds of present/future biasedness implied by the revealed preference test. Each synthetic subject has one pair of a blue circle (upper bound) and a red circle (lower bound). The sample is 9,364 synthetic subjects who (i) have non-linear utility ($\alpha < 1$) and present/future bias ($\beta \neq 1$), (ii) pass the QHD test, and (iv) have recovered range $[\underline{\beta}_i, \bar{\beta}_i]$ with $\underline{\beta}_i, \bar{\beta}_i > 0$. The dotted line represents the 45-degree line.

A.4 Additional Results from Empirical Application

In this section we provide additional results supporting the argument in Section 2.5.2 where we compare AS's parametric estimation of a QHD model and results from our nonparametric revealed preference tests and present our measure of distance from M rationality.

Remember that AS estimate the per-period discount factor, present bias, and utility curvature assuming a QHD model with CRRA utility over money:

$$U(x_0, \ldots, x_T) = \frac{1}{\alpha} x_0^{\alpha} + \beta \sum_{t \in T \setminus \{0\}} \delta^t \frac{1}{\alpha} x_t^{\alpha}.$$

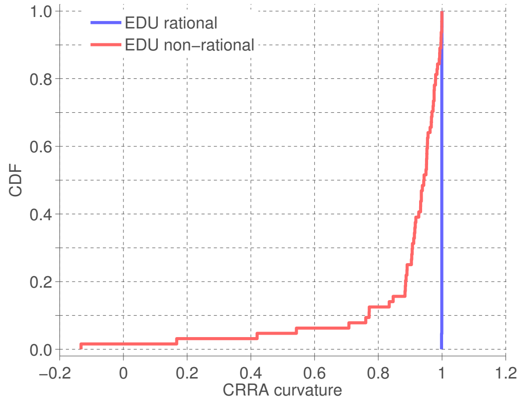
Here we focus on AS's individual level nonlinear least squares (NLS) estimation.

We classify subjects in two groups, those who violate and those who satisfy EDU based on the revealed preference tests. Panels (A)-(C) of Figure A.4 present empirical cumulative distribution functions (CDFs) for the estimated preference parameters in the EDU rational and EDU non-rational groups. Similarly, panels (D)-(F) compare properties of individual's choices (e.g., proportion of interior choices) for the same two groups of subjects.

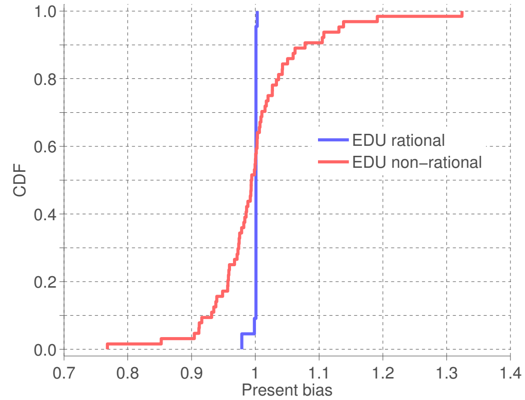
The figure shows how our test is consistent with AS's estimates. Consider panel (B). The CDF for EDU rational subjects concentrates a large mass at $\beta = 1$. The non-EDU group has no such jump in mass at $\beta = 1$, and instead exhibits a substantial fraction of subjects with estimated β different from 1. The CDF for EDU-rational subjects is significantly different from the CDF for EDU non-rational subjects: the null hypothesis of equality-of-distribution is rejected by the two-sample Kolmogorov-Smirnov test ($p < 0.01$).

Figure A.4 panel (B) also shows that subjects who fail our EDU test have estimates of β that differ clearly from 1. An OLS regression of the absolute difference between estimated present bias and 1, $|\widehat{\beta} - 1|$, on a dummy variable for EDU rationality (takes 1 if that subject fails the EDU test) reveals that $\widehat{\beta}$ for EDU non-rational subjects is further away from 1 compared to EDU rational subjects (Table A.2, column 1). A similar result holds for subjects who are not EDU rational but TSU rational and those who are not TSU rational (Table A.2, column 2).

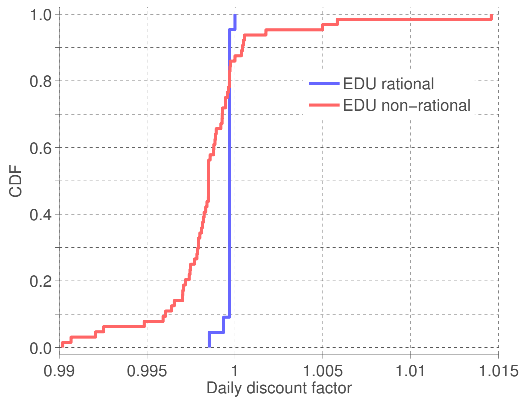
However, $\beta \neq 1$ is not immediately translated into evidence for present or future bias. As we have shown above, most of the subjects who fail the EDU test also fail the QHD test (no additional subject passes the test for PQHD, and most of the subjects who failed EDU even fail MTD). In this sense, the interpretation of estimated β for EDU non-rational subjects in Figure A.4 panel (B) requires some



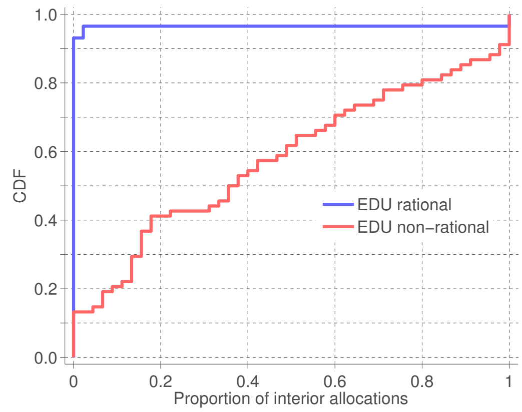
(A) CRRA curvature



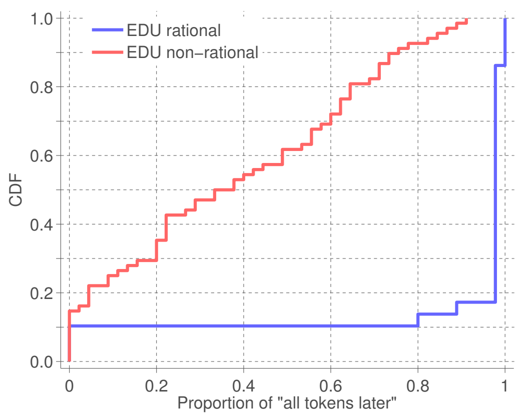
(B) Present bias



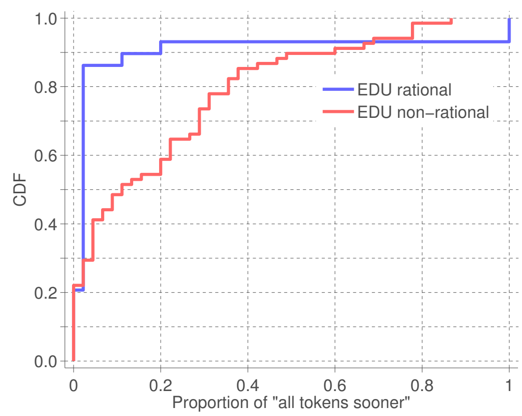
(C) Daily discount factor



(D) Proportion of interior allocations



(E) Proportion of “all later” allocations



(F) Proportion of “all sooner” allocations

FIGURE A.4: Empirical CDFs for preference parameters and properties of choices. Panels (A)-(C) include 86 subjects whose preference parameters are estimable. Panels (D)-(F) include all 97 subjects.

caution. The model is arguably misspecified for such subjects.

One of the advantages of our revealed preference tests is that we can go beyond

TABLE A.2: OLS regression of $|\widehat{\beta}-1|$ on rationality dummies.

	(1)	(2)
nonEDU	0.046 ***	
	(0.007)	
$TSU\setminus EDU$		0.055 ***
		(0.019)
nonTSU		0.043 ***
		(0.007)
Constant	0.002 **	0.002 **
	(0.001)	(0.001)
R^2	0.139	0.147
# Obs.	86	86

Notes: nonEDU is a dummy for subjects who fail the EDU test, $TSU\setminus EDU$ is a dummy for those who fail the EDU test but pass the TSU test, and *nonTSU* is a dummy for those who pass the TSU test. Robust standard errors are reported in parentheses. Level of significance. $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.10$.

the class of QHD utility function by weakening the restrictions in the relevant revealed preference axioms.

Consider Figure 2.2 again. It is interesting to note that the estimated values of $\widehat{\beta}$ for subjects who fail our EDU test are symmetrically distributed around 1.[2] The "average" subject looks, in some sense, as an EDU agent, even though the majority of subjects are not consistent with that model according to our test. It is therefore possible that AS's finding in favor of EDU in their aggregate preference estimation reflects the choice behavior of such an average subject.

A.4.1 Estimated Daily Discount Factors

As in Figure 2.2 where we show AS's estimated present-bias parameter $\widehat{\beta}$ for each class of rationality, Figure A.5 demonstrates the similar comparison for the case of AS's estimated daily discount factor $\widehat{\delta}$. The subjects who pass the EDU test have estimated $\widehat{\delta}$ very close to 1 (many of them have $\widehat{\delta} = 0.9997$). The subjects who do not pass any of the tests (i.e., TSU non-rational subjects) have estimated $\widehat{\delta}$ which are far from 1 in magnitude compared to the other groups of subjects.

[2] We test symmetry using the two-sample Kolmogorov-Smirnov (K-S) test. We first sort estimated $\widehat{\beta}$ in an ascending order, calculate $|\widehat{\beta}-1|$, and split them into the first half (smaller $\widehat{\beta}$) and the last half (larger $\widehat{\beta}$). We apply K-S test for equality of distribution for those two empirical distributions of $|\widehat{\beta}-1|$. The null hypothesis of equal distribution is not rejected ($p = 0.132$).

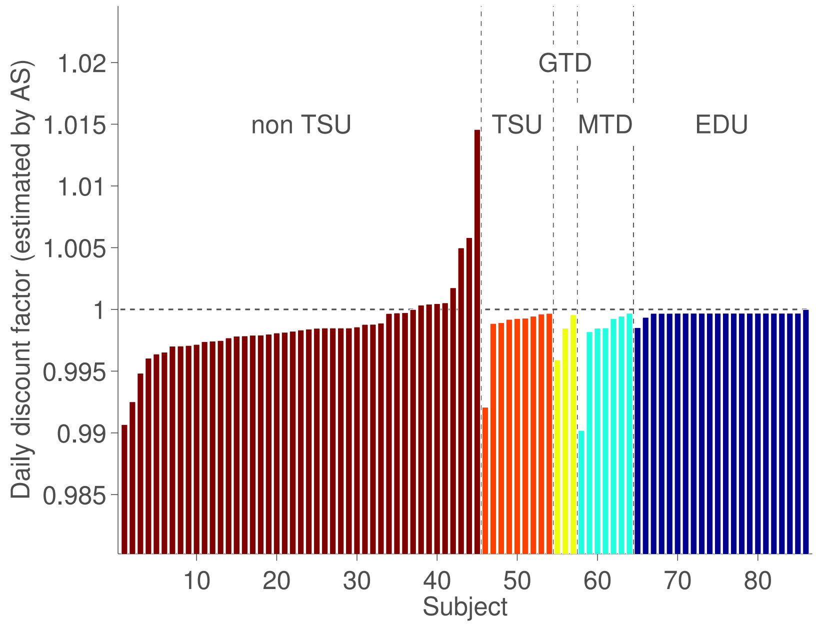


FIGURE A.5: Estimated daily discount factor for each category of subjects.

Furthermore, those who have $\widehat{\delta} > 1$ are all in this category.

A.4.2 Parameter Recovery

As we described in Section A.2, we can find bounds of daily discount factor δ or present-biasedness β, which can rationalize the observed choice data.

Table A.3 lists bounds of discount factor (together with estimated values provided by AS) for 29 EDU rational subjects, and Table A.4 lists bounds of present-biasedness for the same 29 QHD rational subjects.

TABLE A.3: Recovered bounds for daily discount factor (29 EDU rational subjects).

Upper bound	Lower bound	AS estimates
0.9899	0.0000	0.6951
0.9899	0.0000	N.A.
0.9985	0.9985	0.9985
0.9993	0.9993	0.9994
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9997	1.0000
1.0000	1.0000	0.9981
1.0000	1.0000	0.9989
1.0000	1.0000	0.9997
1.0000	1.0000	1.0000
1.0000	1.0000	1.1118

Table A.4: Recovered bounds for present-biasedness (29 QHD rational subjects).

Upper bound	Lower bound	AS estimates
1.0000	0.9676	0.9788
1.0000	0.9972	0.9986
1.3194	0.9500	1.0139
∞	0.9500	0.9856
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0030
∞	0.9500	1.0030
∞	0.9500	1.0781
∞	0	N.A.
∞	0	N.A.

A.5 Implementing Minimum Price Perturbation Test

We describe details on calculation of the distance measure based on minimum price perturbation. Here we focus only on the case of EDU.

Let $D_{\text{true}} = (q^k, x^k)_{k=1}^K$ denote a "true" dataset and $D_{\text{obs}} = (p^k, x^k)_{k=1}^K$ denote an "observed" dataset. The true and observed datasets are connected by the relationship $q_t^k = p_t^k \varepsilon_t^k$ for all $t = 0, \ldots, T$ and $k = 1, \ldots K$, where $\varepsilon_t^k > 0$ is a random variable.

Let H_0 and H_1 denote the null hypothesis that the true dataset D_{true} is EDU rational and the alternative hypothesis that D_{true} is not EDU rational. Consider a test statistic, which is the solution to the following optimization problem given a dataset $D_{\text{obs}} = (p^k, x^k)_{k=1}^K$:

$$\begin{aligned} \min_{(\delta, v_t^k, \lambda^k, \varepsilon_t^k)_{t,k}} \quad & \sum_{k=1}^K \sum_{t=0}^T \frac{1}{K(T+1)} \left| \log \varepsilon_t^k \right| \\ \text{s.t.} \quad & t \log \delta + \log v_t^k - \log \lambda^k - \log p_t^k - \log \varepsilon_t^k = 0 \qquad (\star) \\ & x_t^k > x_{t'}^{k'} \Longrightarrow \log v_t^k \leq \log v_{t'}^{k'}. \end{aligned}$$

Under the null hypothesis, the true dataset $D_{\text{true}} = (q^k, x^k)_{k=1}^K$ is EDU rational. Lemma 1 then implies that there exist strictly positive numbers $\widetilde{\delta}$, $\widetilde{v}_t^k$, and $\widetilde{\lambda}^k$ for $t = 0, \ldots, T$ and $k = 1, \ldots, K$ such that

$$t \log \widetilde{\delta} + \log \widetilde{v}_t^k - \log \widetilde{\lambda}^k - \log q_t^k = 0 \quad \text{and} \quad x_t^k > x_{t'}^{k'} \Longrightarrow \log \widetilde{v}_t^k \leq \log \widetilde{v}_{t'}^{k'}.$$

Substituting the relationship $q_t^k = p_t^k \varepsilon_t^k$ for all $t = 0, \ldots, T$ and $k = 1, \ldots K$ yields

$$t \log \widetilde{\delta} + \log \widetilde{v}_t^k - \log \widetilde{\lambda}^k - \log p_t^k = \log \varepsilon_t^k \quad \text{and} \quad x_t^k > x_{t'}^{k'} \Longrightarrow \log \widetilde{v}_t^k \leq \log \widetilde{v}_{t'}^{k'},$$

which implies that the tuple $(\widetilde{\delta}, \widetilde{v}_t^k, \widetilde{\lambda}^k, \varepsilon_t^k)_{t,k}$ satisfies the constraint in problem $(\star)$ but not necessarily the one that minimizes the objective function. Letting $\Phi^* \left((p^k, x^k)_{k=1}^K \right)$ denote the optimal value of the problem $(\star)$, we have

$$\Phi^* \left((p^k, x^k)_{k=1}^K \right) \leq \sum_{k=1}^K \sum_{t=0}^T \frac{1}{K(T+1)} \left| \log \varepsilon_t^k \right| = \widehat{\Phi}$$

under the null hypothesis. Then, we can construct a test as follows:

$$\begin{cases} \text{reject } H_0 & \text{if } \int_{\Phi^*\left((p^k,x^k)_{k=1}^K\right)}^{\infty} f_{\widehat{\Phi}}(z)dz < \alpha \\ \text{accept } H_0 & \text{otherwise} \end{cases},$$

where α is the size of the test and $f_{\widehat{\Phi}}$ is the pdf of the distribution of $\widehat{\Phi} = \sum_{k,t} |\log \varepsilon_t^k|/(K(T+1))$. Given a nominal size α, we can find a critical value C_α satisfying $P(\widehat{\Phi} > C_\alpha) = \alpha$; we set $C_\alpha = F_{\widehat{\Phi}}^{-1}(1-\alpha)$, where $F_{\widehat{\Phi}}$ denotes the cumulative distribution function of $\widehat{\Phi}$. However, because $\Phi^*\left((p^k,x^k)_{k=1}^K\right) \leq \widehat{\Phi}$, the "true size" of the test is $P(\Phi^* > C_\alpha) \leq P(\widehat{\Phi} > C_\alpha) = \alpha$.

Given an observed dataset $D_{\text{obs}} = (p^k,x^k)_{k=1}^K$, we solve the optimization problem ($\star$) using `fmincon` function in Matlab. We set up a problem

$$\begin{cases} \min & g(z) \\ \text{s.t.} & A \cdot z = 0 \\ & B \cdot z \geq 0 \\ & E \cdot z > 0 \end{cases}$$

to run the function. We now construct three key ingredients of the problem, matrices A, B, and E, focusing on the case of EDU model.

The first matrix A has $K \times (T+1)$ rows and $K \times (T+1) + 1 + K + 1 + K \times (T+1)$ columns, defined as follows: we have one row for every pair (k,t), two columns (for $\log v_t^k$ and $\log \varepsilon_t^k$) for every pair (k,t), one column for δ, one column for each k, and one column for p. In the row corresponding to (k,t) the matrix has zeroes everywhere with the following exceptions: it has a 1 in the column for $\log v_t^k$, it has t in the column for δ, it has a -1 in the column for k, $-\log p_t^k$ in the column for p, and it has a -1 in the column for $\log \varepsilon_t^k$. This finalizes the construction of A. The resulting matrix looks as follows:

$$\begin{array}{c} \\ \vdots \\ (k,t) \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc|c|ccccc|c|ccccc} (1,0) & \cdots & (k,t) & \cdots & (K,T) & \delta & 1 & \cdots & k & \cdots & 45 & p & (1,0) & \cdots & (k,t) & \cdots & (K,T) \end{array} \\ \left[\begin{array}{ccccc|c|ccccc|c|ccccc} \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k & 0 & \cdots & -1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \end{array}\right] \end{array}.$$

Next, we construct matrix B that has $K \times (T+1) + 1 + K + 1 + K \times (T+1)$ columns and there is one row for every pair (k,t) and (k',t') for which $x_t^k > x_{t'}^{k'}$. In the row corresponding to $x_t^k > x_{t'}^{k'}$ we have zeroes everywhere with the exception of a -1 in the column for (k, t) and a 1 in the column for (k', t').

Finally, we prepare a matrix that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and $K \times (T+1) + 1 + K + 1 + K \times (T+1)$ columns. It has zeroes everywhere except for 1 in the column for p.

Consider the system of linear equalities and inequalities

$$\begin{cases} A \cdot z = 0 \\ B \cdot z \geq 0 \\ E \cdot z > 0 \end{cases} .$$

If there is a solution z to the system, then by dividing through by the $K \times (T + 1) + 1 + K + 1$-th component of z, we obtain numbers $(\delta, v_t^k, \lambda^k, \varepsilon_t^k)_{t,k}$ that satisfy the constraints in the problem $(\star)$.

A.6 Distance Measure Based on Maximal Subset: A Robustness Check

In Section 2.5.2, we introduce a measure to characterize the distance from a given dataset to rationality, be it EDU, QHD, and so on. The ideal method for obtaining such a measure is to check all the possible sequences of dropping observations, starting from dropping one observation, until we can find a largest subdata that pass the test. However, exhaustive checking is computationally extremely challenging. Therefore, we take an alternative approach: randomly drop observations and iterate this procedure. We demonstrate that the distance measure obtained by our approach does not depend on the random procedure heavily.

We prepare three different sets of distance measures, each of which is obtained from 10,000 iterations, for each distance measure d'_{EDU}, d'_{QHD}, and d'_{TSU}. As we see in Figure A.6, three sets result in statistically indistinguishable distributions of distance measures: the null hypothesis of equal distribution is not rejected at all conventional levels in the two-sample Kolmogorov-Smirnov test. For the analyses in Section 2.5.2, we merge three sets and take the shortest path from the total of 30,000 iterations.

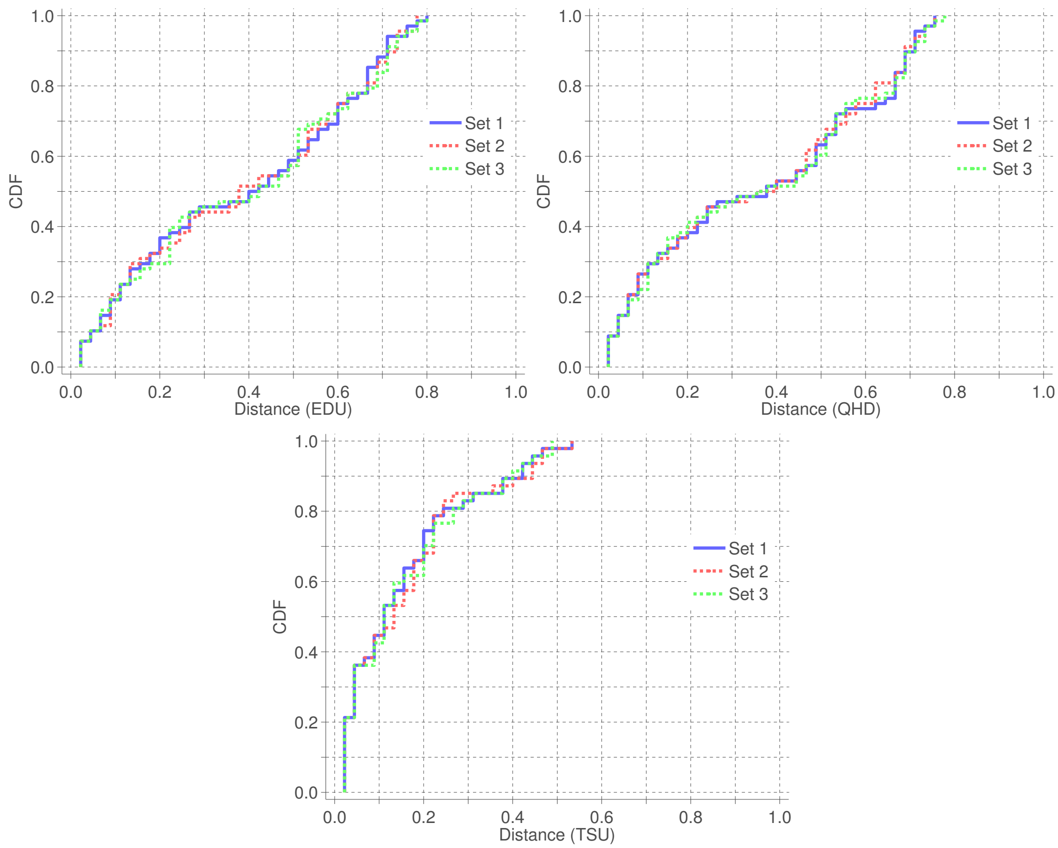


FIGURE A.6: Comparing the distance measures obtained from three sets of 10,000 iterations.

A.7 Jittering: Perturbing Choices

We demonstrate robustness of revealed preference tests to small perturbation in underlying preferences in Section 2.5.2. Here, instead of perturbing preference parameters, we add jitters on choices predicted by a QHD model with a fixed set of parameters.[3]

Assume a QHD model

$$U(x_0, \ldots, x_T) = \frac{1}{\alpha} x_0^\alpha + \beta \sum_{t \in T \setminus \{0\}} \delta^t \frac{1}{\alpha} x_t^\alpha$$

as in AS. For each budget in the AS experiment (there are 45 of those), the model predicts demand for sooner payment, $x(p, \tau, d; \alpha, \delta, \beta)$. We then add "jitters" to these predicted demands so that we observe $\widehat{x}(p, \tau, d; \alpha, \delta, \beta, \sigma) = x(p, \tau, d; \alpha, \delta, \beta) +$

[3] Andreoni et al. (2013) introduce and discuss this way of assessing the goodness-of-fit in the context of revealed preference tests, which they call the jittering measure.

ε. Jitters are assumed to be drawn from a normal distribution, but we ensure that the jittered demand $\widehat{x}(p, \tau, d)$'s are on the budget line. In other words, jitters are drawn from a truncated normal distribution. [4]

In this exercise, we take parameters from AS aggregate estimates: $\alpha = 0.897$, $\delta = 0.999$. For the present bias parameter, we take AS aggregate estimate $\beta = 1.007$ together with other "reasonable" values such as 0.974 (aggregate estimate from Augenblick et al., 2015), 0.995, 1, and 1.05. As for standard deviation of the normal distribution, we use $\sigma \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$.

For each set of parameters and standard deviation of white noise $(\alpha, \delta, \beta, \sigma)$, we simulate 1,000 sets of observations $\{\widehat{x}(p_b, \tau_b, d_b; \alpha, \delta, \beta, \sigma)\}_{b=1}^{45}$. We then perform our EDU and QHD tests.

Table A.5 reports pass rates for the QHD test for each set of parameters and standard deviation. When the standard deviation is $\sigma = 0.001$, the simulated dataset always pass the QHD test. As the standard deviation increases, pass rates decrease at the speed depending on the parameter configuration.[5]

Table A.6 reports the same statistics for the EDU test. A notable feature in this simulation is that the dataset generated by non-EDU preferences (i.e., $\beta = 0.995$ and 1.007) pass the EDU test in many occasions. As in the case of the QHD test, pass rates decrease at the speed depending on the parameter configuration.

This exercise has demonstrated that our revealed preference tests detect irregularities induced by white noise, but we cannot provide a definitive answer to whether the degree of irregularities necessary to violate EDU/QHD rationality is big or small (in other words, how sensitive our tests are) because we do not have a clear benchmark to compare with.

AS provide standard error of NLS error in the aggregate estimate (corresponding to parameter set #4), which is 6.13.

Alternatively, one can use variations observed in the actual experimental data to compare with standard deviations used in this exercise. Let $x_i(p_b, \tau_b, d_b)$ denote subject i's demand for sooner payment in budget b. Then, we calculate the root mean squared error (RMSE)

$$v_i = \sqrt{\frac{1}{45} \sum_{b=1}^{45} \left(x_i(p_b, \tau_b, d_b) - x(p_b, \tau_b, d_b; \alpha, \delta, \beta) \right)^2}$$

[4] Andreoni et al. (2013) note that "truncating is known to bias the frequency of corner solutions downward". An alternative approach is "censoring," which would have a bias in the opposite direction.

[5] We also confirm that predicted choices indeed pass the QHD test in the absence of jittering (4th column in the table).

TABLE A.5: QHD test pass rates.

	Parameters			Standard deviation (σ)							
#	α	δ	β	0	0.001	0.005	0.010	0.050	0.100	0.500	1.000
1	0.897	0.999	0.974	1.00	1.00	0.99	0.83	0.21	0.02	0.00	0.00
2	0.897	0.999	0.995	1.00	1.00	1.00	1.00	0.47	0.16	0.00	0.00
3	0.897	0.999	1.000	1.00	1.00	1.00	1.00	0.46	0.18	0.00	0.00
4	0.897	0.999	1.007	1.00	1.00	1.00	0.98	0.30	0.10	0.00	0.00
5	0.897	0.999	1.050	1.00	1.00	1.00	0.92	0.23	0.05	0.00	0.00

TABLE A.6: EDU test pass rates.

	Parameters			Standard deviation (σ)							
#	α	δ	β	0	0.001	0.005	0.010	0.050	0.100	0.500	1.000
1	0.897	0.999	0.974	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.897	0.999	0.995	1.00	1.00	1.00	1.00	0.47	0.16	0.00	0.00
3	0.897	0.999	1.000	1.00	1.00	1.00	1.00	0.46	0.18	0.00	0.00
4	0.897	0.999	1.007	1.00	1.00	1.00	0.96	0.25	0.09	0.00	0.00
5	0.897	0.999	1.050	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

for each subject i. Table A.7 reports summary statistics for the distribution of v_i's. It is clear that the variation of the observed data measured by RMSE is much higher than the standard deviation of white noise at which we achieve 50% pass rate for the QHD test. This may suggest that about 50% of the subjects are not rationalized by QHD model because of structural irregularities rather than trembling on their choices. However, we emphasize again that we do not have clear guidance for the benchmark: we demonstrate the case of v_i's but this may not be the right one to compare with.

Table A.7: Distributions of v_i's.

	Parameters			Percentile						
#	α	δ	β	5-th	10-th	25-th	50-th	75-th	90-th	95-th
1	0.897	0.999	0.974	3.00	3.76	4.68	5.93	6.33	7.83	10.50
2	0.897	0.999	0.995	2.91	3.66	4.60	5.93	6.17	7.94	10.61
3	0.897	0.999	1.000	2.93	3.68	4.63	5.94	6.15	7.97	10.64
4	0.897	0.999	1.007	2.95	3.71	4.62	5.91	6.18	8.02	10.67
5	0.897	0.999	1.050	3.10	3.58	4.48	5.61	6.13	8.28	10.92

Appendix B

Appendix to Chapter 3

B.1 Gaze Data Processing

Prior to analysis, blinks and saccades were removed using a velocity threshold of 8 pixels/4 ms. That is, any rapid gaze movement whose velocity is greater than 200 pixels/100 ms was considered as blinks or saccades and discarded. It is known that little or no visual processing can be achieved during saccades (Fuchs, 1971).

Any problem related to the calibration was noticed and addressed during the first trial (which was immediately after the calibration procedure). Further, subjects would often ask a question during the first trial. This affected overall trial durations and the integrity of fixations in the first trial. Therefore, for eye tracking data analysis, the first trial of each eyetracking part was discarded unless noted otherwise.

Regions of interests (ROIs) for the product were defined individually within the boundary of each image padded with a 25-pixel-wide band around the edge. The padding was added to accommodate noise of the eyetracker and viewing of the product through peripheral vision. ROIs for the price were defined as a rectangular block containing the displayed price. The rest of the area was considered blank. We defined a fixation as the event when the gaze enters, stays in, and leaves an ROI; in this definition, any continuous eye movement within a given ROI is considered as one fixation. Therefore, a fixation time was defined as the time elapsed from the moment when the gaze entered the ROI to the moment when it left the ROI, and a cumulative fixation time was defined as the sum of fixation times within a trial.

We constructed spatial gaze distribution maps for the hypothetical and real conditions (and further for Yes and No decisions) that reflect the frequency of the gaze at any coordinate on the screen. For the construction of group heat maps, we took two subsets of the screen pixel space: one area of 370×370 (width by height) in size that includes the product image and the other of 370×150 in size that includes the price. We then placed the product image area above the price area to reconcile the counter-balanced display with the product-top/price-bottom display; this resulted in the 370×520 pixel space for gaze distribution. For each subject, a gaze distribution map was computed by summing the duration of gaze at each pixel within a trial and averaging it across all relevant trials, and then smoothing it with a $2D$ Gaussian kernel ($\sigma = 10$ pixels). The smoothed individual maps were averaged across subjects for each condition to create Figure 3.3, panel A. We also performed statistical comparisons of gaze distributions between the hypothetical and real conditions (and further for Yes and No decisions), applying a procedure adapted from (Caldara and Miellet, 2011). Specifically, we took the difference of individual heat maps between conditions of interest, averaged across subjects, and then converted all pixel values into z-scores relative to the mean and standard deviation of the group difference heat map. Significance was established with the statistical threshold provided by a two-tailed Pixel test ($|z| > 4.3125$ for $p < 0.05$; Chauvin et al., 2005), which corrects for multiple comparisons in the heat map pixel space.

B.2 Additional Results

B.2.1 Monotonicity of Preferences

In each condition, participants faced the same item three times with three different prices, low, medium, and high. If a participant has a monotonic preference and she chooses to purchase an item at price P, she should also purchase at any lower prices $P' < P$. We count the number of monotonicity violations at item level for each subject in each condition. In experiment I, the average number of monotonicity violations is 0.857 in the Hypothetical condition and 0.357 in the Real condition. In experiment II, those numbers are 1.471 and 0.294.

When we aggregate responses across items at each price level, we observe monotonically decreasing purchase rates as in Figure B.2.

B.2.2 Estimating the Size of Hypothetical Bias

Following Kang et al. (2011), we define the *adjusted consumer surplus* from item i by $aCS_i = \theta \times WTP_i - P_i$, where θ is a discount factor. We then estimate θ for each condition as follows: (i) let $\Pr(x)$ denote a probability of Yes decision at x; (ii) estimate logistic regression $\Pr(aCS) = 1/(1 + \exp(-(\alpha + \beta \cdot aCS)))$; and (iii) find θ at which $\Pr(0) = 0.5$. Median θ^{Hyp} (0.98 in Experiment I; 0.98 in Experiment II) is not significantly different from 1 (signed-rank tests, p's > 0.05), median θ^{Real} (0.61 in Experiment I; 0.51 in Experiment II) is significantly less than 1 (signed-rank tests, p's < 0.001), and median difference between θ^{Hyp} and θ^{Real} (0.36 in Experiment I; 0.49 in Experiment II) is significantly larger than 0 (signed-rank tests, p's < 0.001) in both experiment. Consistent with the previous finding (Kang et al., 2011), the observed pattern suggests that subjects behaved as if they used their (hypothetically) stated WTPs in the hypothetical condition, but that the values subjects placed on the objects were about 35% to 50% lower in the real condition. It is not surprising that θ^{Hyp} is close to 1, since subjects stated "hypothetical" willingness to pay during the first part of the experiment.

B.3 Supplementary Figures and Tables

TABLE B.1: Summary statistics in Experiment I (mousetracking) and Experiment II (eye-tracking).

	Experiment I		Experiment II	
Average	Hyp	S-Real	Hyp	S-Real
Purchase percentage (%) a, b^{***}	55.99	27.14	55.88	18.69
Response time (sec) a, b^{***}	4.01	3.09	2.30	1.74
Cumulative image viewing time (sec) b^{***}	0.95	0.87	1.32	0.99
Cumulative price viewing time (sec) a, b^{***}	0.70	0.51	0.49	0.40
# of image clicks/fixations b^{***}	1.27	1.23	2.44	2.12
# of price clicks/fixations b^{**}	1.21	1.12	1.54	1.31
Image viewing time per click/fixation (sec) a^{**} b^{***}	0.72	0.60	0.53	0.43
Price viewing time per click/fixation (sec) a^{***} b^{**}	0.60	0.45	0.32	0.29

Notes: Asterisks indicate statistical significance between hypothetical and real condition. $***$: $p < 0.01$, $**$: $p < 0.05$, $*$: $p < 0.10$, for Experiment I (a) and II (b). In Experiment II, average for *purchase percentage* is calculated including first trial of each condition.

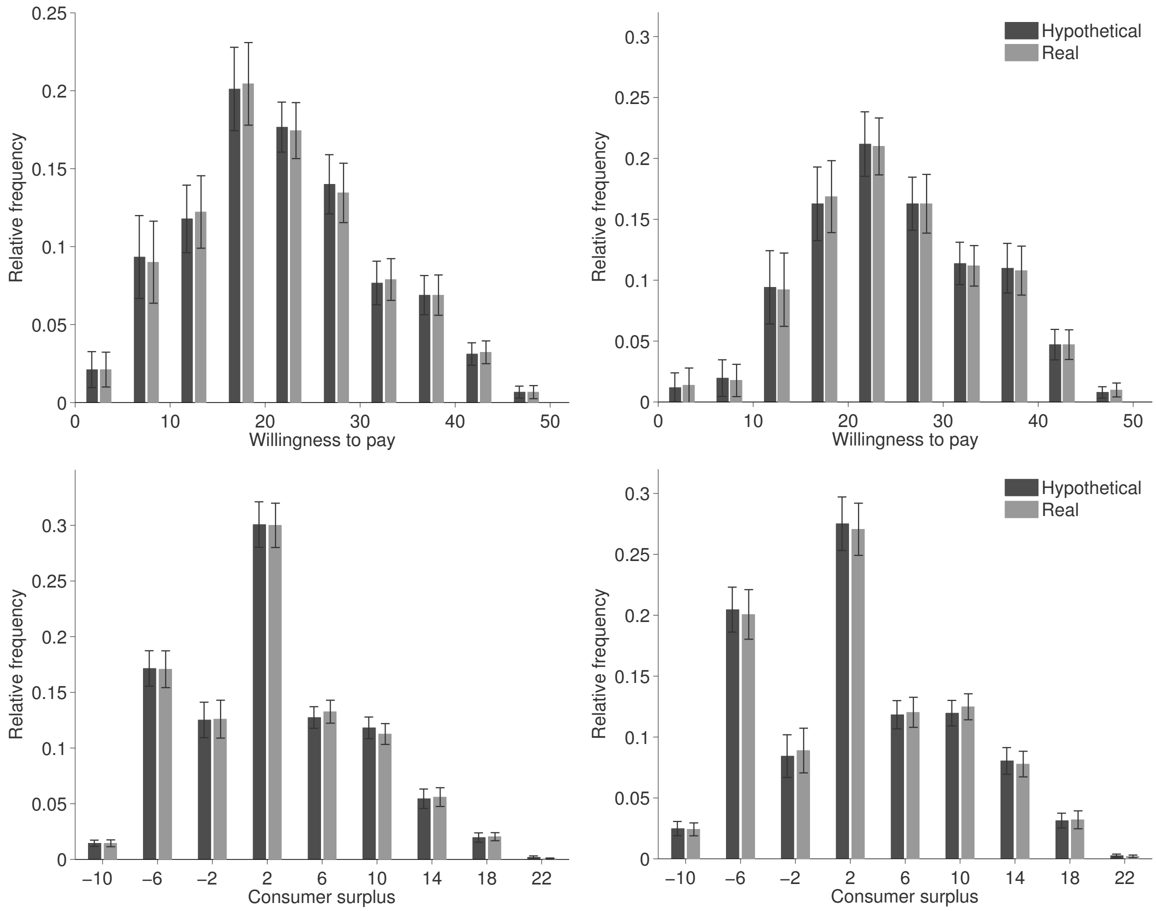


FIGURE B.1: Distribution of WTP and CS (= WTP − price) in the hypothetical and real conditions. (Top) Distribution of WTP (left: Experiment I; right: Experiment II). Bin size = 5, bin center = $[2.5, 7.5, \ldots, 47.5]$. (Bottom) Distribution of in CS (left: Experiment I; right: Experiment II). Bin size = 4, bin center = $[-10, -6, \ldots, 22]$. The first trial of each condition in Experiment II is included.

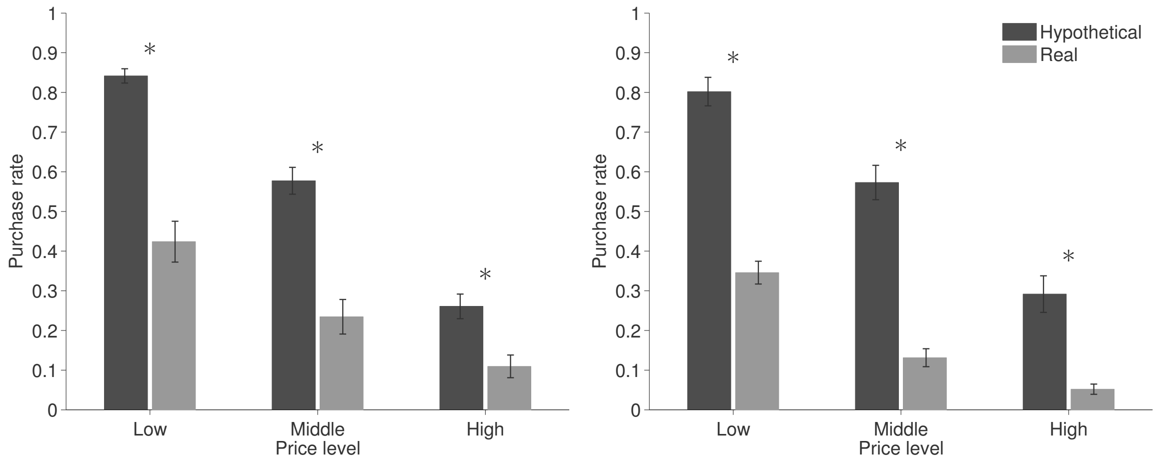


FIGURE B.2: Average purchase rate by condition and price level in Experiment I (left) and Experiment II (right). All comparisons of means between hypothetical and real conditions are significant at $p < 0.0001$, two-sided paired sample t-test. The first trial of each condition in Experiment II is included.

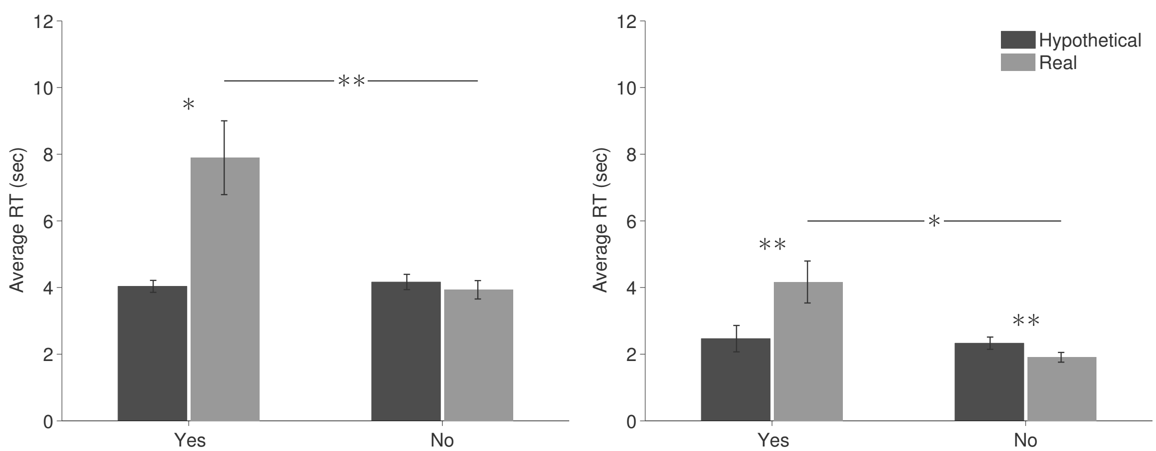


FIGURE B.3: Average RT by condition and decision in Experiment I (left) and Experiment II (right). Error bars indicate standard errors. $** : p < 0.001$, $* : p < 0.01$, two-sided paired sample t-test.

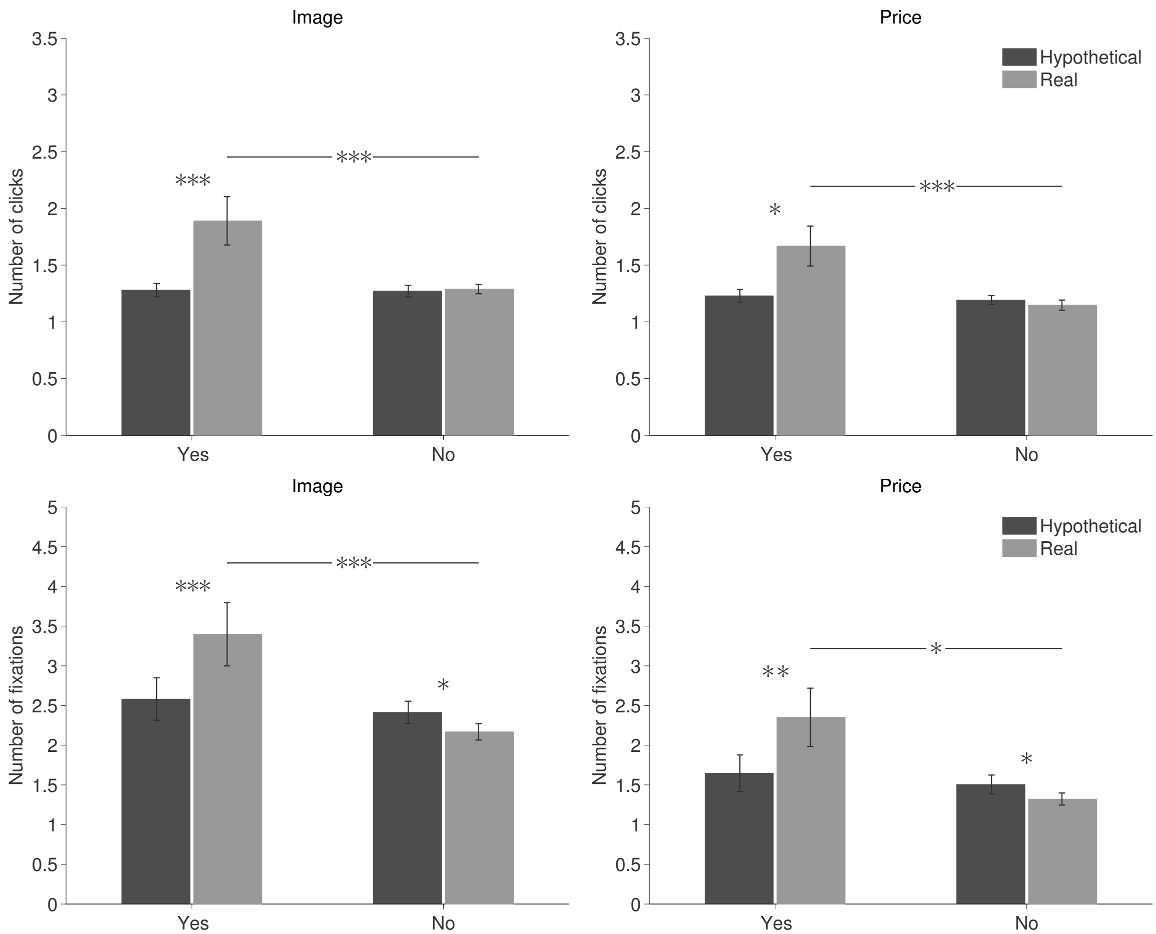


FIGURE B.4: The average number of clicks/fixations on the occluded areas by condition and decision. Clicks on the image and on the price in Experiment I (top panels) and Fixations on the image and the price in Experiment II (bottom panels). Error bars indicate standard errors. $*** : p < 0.005$, $** : p < 0.01$, $* : p < 0.05$, two-sided paired sample t-test.

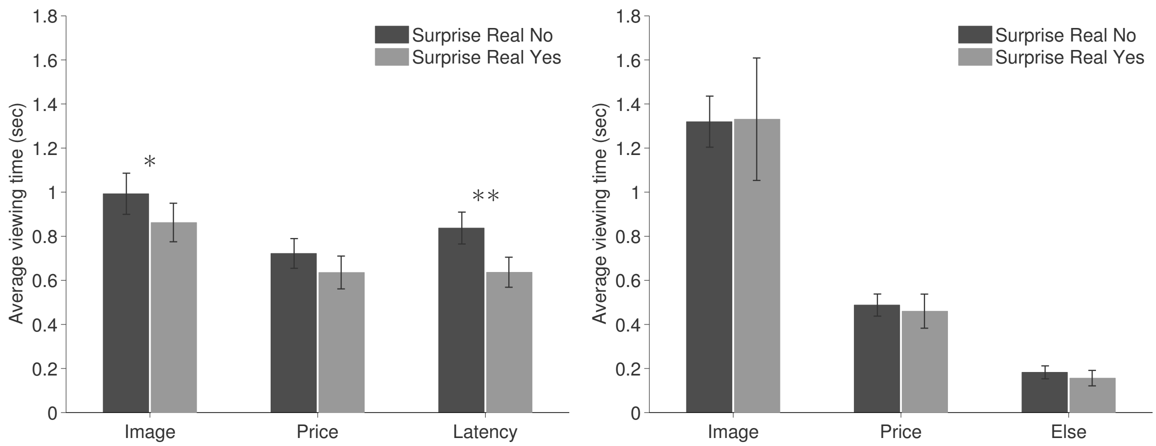


Figure B.5: Average viewing time and latency/else (sec) in hypothetical trials, sorted by decision in the matching surprise real trials. Image and price viewing times in Experiment I (top) and in Experiment II (bottom). Error bars indicate standard errors. $*** : p < 0.001$, $** : p < 0.005$, $* : p < 0.05$, two-sided paired sample t-test. The comparison between hypothetical Yes and No is not significant.

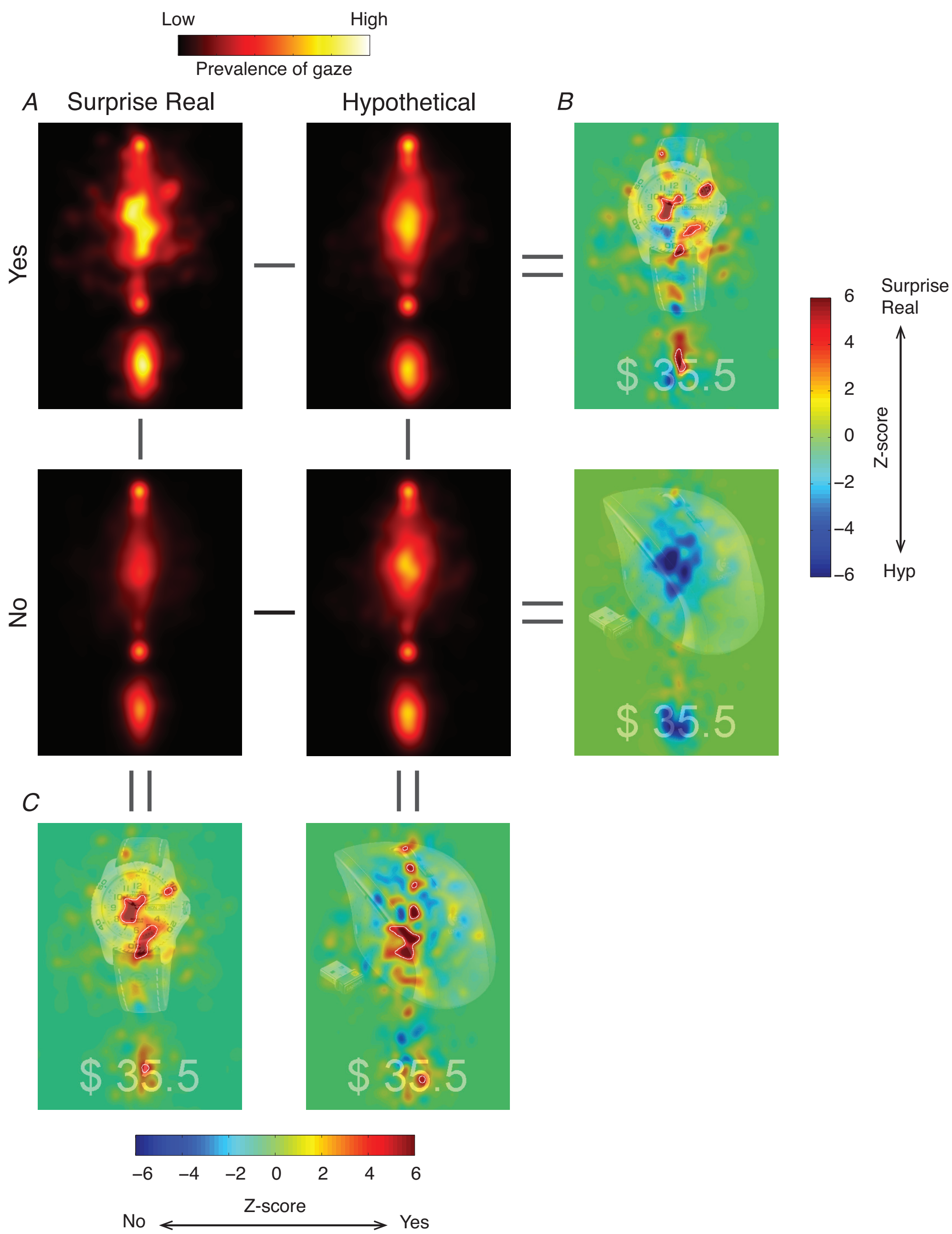


FIGURE B.6: Gaze distribution maps by condition and decision from Experiment II. *A*. Average gaze prevalence. *B*. Statistical significance of the difference between surprise real and hypothetical conditions. Red indicates gaze bias toward surprise real choice and blue indicates gaze bias toward hypothetical choice. *C*. Statistical significance of the difference between Yes and No decisions within each condition. In Panels *B* and *C*, product images are shown in the background for illustration. Red indicates gaze bias toward Yes and blue indicates bias toward No. The threshold for two-tailed Pixel test $p < 0.05$ is $|z| > 4.3125$, corrected for multiple comparisons.

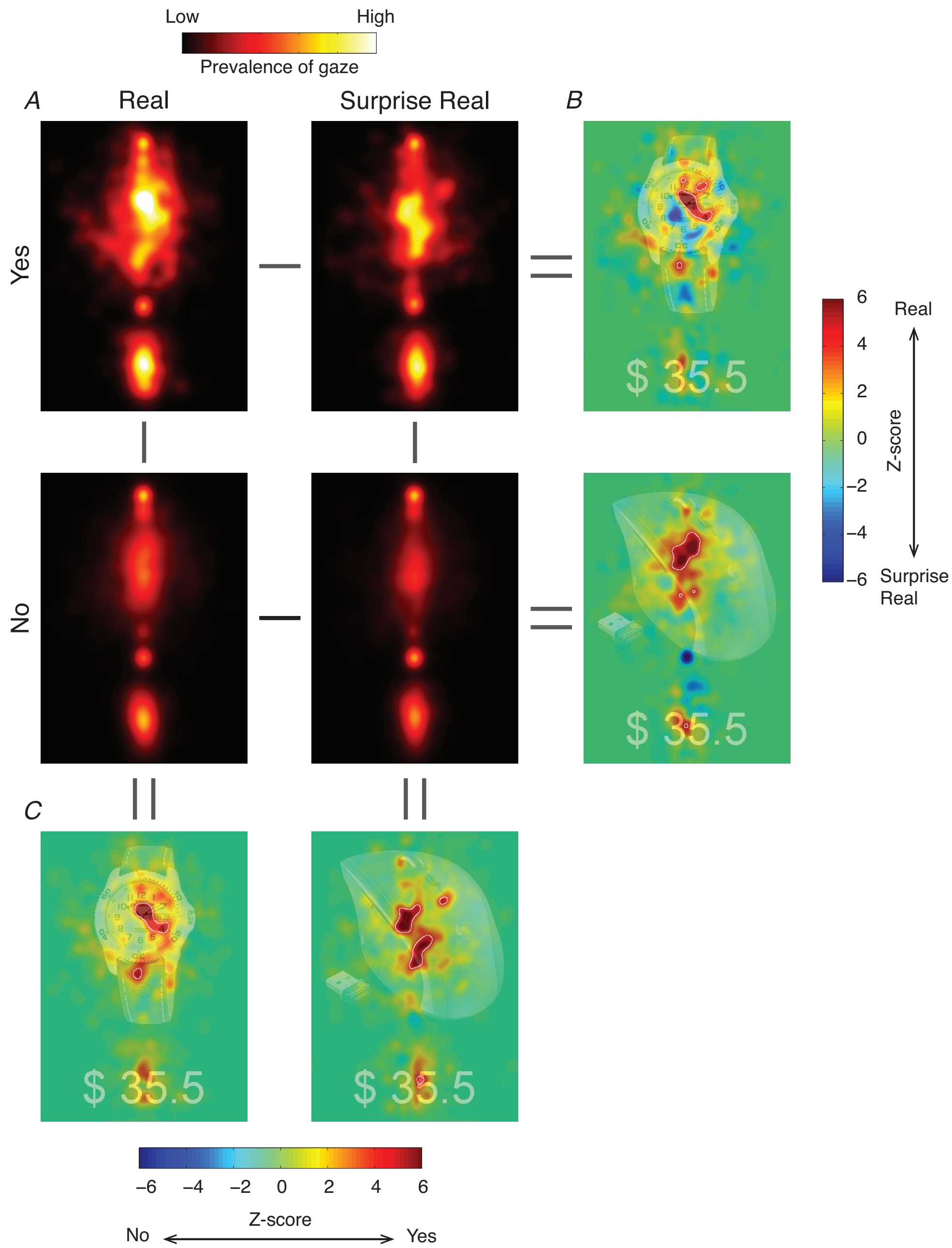


FIGURE B.7: Gaze distribution maps by condition and decision from Experiment II. *A*. Average gaze prevalence. *B*. Statistical significance of the difference between real and surprise real conditions. Red indicates gaze bias toward real choice and blue indicates gaze bias toward surprise real choice. *C*. Statistical significance of the difference between Yes and No decisions within each condition. In Panels *B* and *C*, product images are shown in the background for illustration. Red indicates gaze bias toward Yes and blue indicates bias toward No. The threshold for two-tailed Pixel test $p < 0.05$ is $|z| > 4.3125$, corrected for multiple comparisons.

TABLE B.2: Summary statistics in the hypothetical condition in Experiment I (mousetracking) and Experiment II (eyetracking). First trials are excluded from data in Experiment II. "Stick" indicates hypothetical Yes trial for which participants stick to Yes decision in later surprise real condition, while "switch" indicates hypothetical Yes trials for which they later change their mind to no in the surprise real condition. Two participants who never "switched" are excluded in Experiment I.

	Experiment I		Experiment II	
Average	Stick	Switch	Stick	Switch
WTP ($) *a***	24.55	22.64	26.73	27.96
Price ($) *a*** *b****	17.00	19.10	18.53	23.66
Response time (sec) *a****	3.65	4.51	2.27	2.44
Cumulative image viewing time (sec) *a****	0.87	1.03	1.35	1.39
Cumulative price viewing time (sec) *a**** *b**	0.63	0.77	0.47	0.51
Latency/Else (sec) *a****	0.62	0.86	0.18	0.20
Standardized response time *a**** *b***	-0.06	0.21	0.00	0.18
Standardized image viewing time *a***	-0.10	0.00	-0.00	0.13
Standardized price viewing time *a**** *b****	0.03	0.30	0.04	0.21
Standardized latency/else *a****	-0.02	0.32	0.05	0.18
Number of observations	627	676	253	580

Notes: In the bottom three rows, RT and viewing times are standardized within subject across conditions. Asterisks indicate statistical significance between hypothetical and real condition. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.10$, for Experiment I (*a*) and II (*b*).

Table B.3: List of 120 consumer products used in the study.

#	Product Name
1	Accudart Classic Bristle Dartboard
2	Allied 180-pc. Household Tool Set
3	Angry Birds Speaker
4	Apple 2 GB iPod Shuffle
5	AUDIO TECHNICA Noise-Canceling Headphones
6	Austin Bazaar Full Size Acoustic Guitar with Carrying Bag and Accessories
7	Black & Decker Toast-R-Oven 4-Slice Toaster Oven
8	Black and Decker 12 cup SmartBrew plus Coffeemaker
9	Boss Fabric Deluxe Posture Chair
10	Braun Free Control Men's Shaver
11	Bushnell Perma Focus 10 × 50 Wide Angle Binocular
12	Canon Color Image Scanner
13	Canon PIXMA Wireless Inkjet Photo All-In-One
14	Casio Mens G-Shock Classic Watch
15	Celestron 60mm PowerSeeker Telescope
16	Celestron Handheld Digital Microscope Camera
17	Cisco-Linksys Dual-Band Wireless-N Gigabit Router
18	Classic Pillow by Tempur Pedic
19	Coby 1.8″ Digital TFT LCD Photo Keychain
20	Coca-Cola Personal 6-Can Mini Fridge
21	Coleman Crescent Mummy Bag
22	Columbia Sportswear Men's Northbend Wide Hiking Boot
23	Columbia Sportswear Men's Steens Mountain Sweater
24	Cooler Master Notepal U2 Notebook Cooler with Two Fans
25	Corsair 32 GB USB 3.0 Flash Voyager
26	Crane 2.3 Gallon COOL Mist humidifier
27	Energizer Rechargeable 15 Minute Charger with 4 AA Batteries
28	Eureka Hand-Held Vacuum
29	Febreze Odor Removal Appliance
30	Fossil Men's AM4369 Stainless Steel Analog with Blue Dial Watch

#	Product Name
31	Garmin eTrex Handheld GPS Navigator
32	GODIVA Chocoiste Solid Milk Chocolate Bars (24 pc)
33	GODIVA Thankyou Ballotin (36pc)
34	Green Laser Pointer (5 mW, Class IIIa Laser Product)
35	GSM Quadband Voice Dialing Watch Cell Phone Unlocked
36	Homedics Shiatsu Massage Cushion
37	HoMedics Shiatsu Neck Massager with Vibration and Heat
38	Honeywell Compact Air Purifier with Permanent HEPA Filter
39	Hoover Tempo Widepath Bagged Upright Vacuum
40	iMPROV Electronics 8.5″ Boogie Board Tablet
41	iNeed Lumbar Massage Cushion
42	Intex Raised Downy Queen Airbed with Built-in Electric Pump
43	Kensington DomeHub 7-port USB 2.0 Hub with FlyLight
44	Kodak Easyshare 12 MP Digital Camera with 5x Digital Zoom
45	Kodak PlayFull Waterproof Video Camera
46	Koolatron Kool Fridge-1.7 Cu. Ft.
47	La Crosse Technology Projection Alarm Clock with Outdoor Temperature
48	LAVA 20-oz. Motion Lamp
49	Levi's Relaxed Straight 559 Jeans
50	Logitech 2 MP Portable Webcam
51	Logitech Gaming Keyboard G110
52	Logitech Wireless Marathon Mouse M705 with 3-Year Battery Life
53	Logitech Wireless Presenter R400 with Red Laser Pointer
54	Lumian Design LED Desk Light (built-in night light)
55	M-51 Engineers Field Bag - Military Style
56	M-Audio E-Keys 37 MIDI Keyboard
57	Memory Foam Travel Pillow
58	Microsoft Natural Ergo Keyboard
59	Monopoly The Mega Edition
60	Monoprice 10 $\times$ 6.25 Inches Graphic Drawing Tablet with 8 Hot Key

#	Product Name
61	Mr. Coffee Espresso/Cappuccino Maker
62	New Plantronics Bluetooth Headset- Volume Control/Hands Free/ Wind Noise Reduction
63	Nike Team Training Duffel Bag
64	Nike Tear-away II Mens Pants
65	Nikon Action 8×40 Binocular
66	Northface Jester Backpack
67	Northface Men's Divide Jacket
68	Northface Utility Waist - Sport Hiker
69	Oral B Pulsonic Sonic Electric Rechargeable Power Toothbrush
70	Oregon Scientific Weather Forecaster with Projection Clock
71	Oster Stainless-Steel 1-2/3-Liter Electric Water Kettle
72	Pail of Treats - Mrs. Fields 48 bite-sizes cookies and 36 brownie bites
73	Panasonic Upper Arm Blood Pressure Monitor
74	Perpetual Calendar
75	Philips 7-Inch LCD Portable TV/DVD Player
76	Philips MP3 Portable Speaker Universal
77	Philips Norelco 7340 Cordless Men's Shaving System
78	Planet Earth - The Complete BBC Series (DVD)
79	Post-it Desktop Organizer
80	Premium Diamond Suited Poker Chip Set
81	Ray Ban RB 4115 Sunglasses - Smoky Black/Green
82	Raytek MiniTemp No-Contact Thermometer with Laser Sighting
83	Samsung Blu-Ray Combo Internal 12XReadable and DVD-Writable Drive with Lightscribe
84	Scosche Flexible Bluetooth Mini Keyboard
85	Scrabble Premier Wood Edition
86	SentrySafe Security Safe, 0.5 Cubic Feet
87	Sigma PC 15 Heart Rate Monitor
88	Sketchers Men's Alley Cats
89	Sony CD/Cassette Portable Boombox
90	Sony Digital 2 GB Flash Voice Recorder

#	Product Name
91	Sony Stereo CD Clock Radio with Dual Alarm
92	Sony Walkman 8 GB Video MP3 Player
93	Sony Water-Resistant Weather Band Shower Radio
94	Sony Wireless Headphone System
95	Star Wars: The Original Trilogy (Episodes IV - VI) [Blu-ray] (2011)
96	Stellanova Series 4″ Magnetically Levitating Globe
97	Swiss Gear 7 × 7-Foot 3-Person Sport Dome Cheval Tent
98	Syma S107G 3 Channel RC Radio Remote Control Helicopter with Gyro
99	Tanita Duo Scale with Body Fat/Water Monitor
100	Tempur-Pedic Neck Pillow
101	Texas Instruments TI-83 Plus Graphing Calculator
102	T-Fal Avante Deluxe 4-Slice Toaster
103	The Lord of the Rings - The Motion Picture Trilogy (Blu-ray)
104	The Simpsons - The Complete Tenth Season (1998)
105	Three Classic Games All in One Convenient Box - Deluxe Wooden Chess, Checker & Backgammon
106	Three-Way Use Desktop Organizer, Medium Oak
107	Timberland Earthkeepers Campus Quad Messenger Bag
108	Timberland Men's Full-Zip Thick Stripe N.H. Hoodie
109	Timex Men's Expedition Watch with Leather Strap
110	Toshiba Canvio 500 GB USB 3.0 Portable Hard Drive
111	Transcend 16 GB JetFlash 700 Super Speed USB 3.0 Flash Drive
112	Vantec NexStar 2.5″/3.5″ SATA to USB 2.0 and eSATA Hard Drive Dock
113	Veho MUVI Micro digital camcorder with 2GB Memory
114	Victorinox Hanging Toiletry Kit by Swiss Army
115	Victorinox Swiss Army Credit Card-Size Multi-Tools with LED Light
116	Victorinox Swiss Army Explorer Multi-Tool Knife
117	Viewsonic 8-Inch Digital Photo Frame with 800 × 600 High Resolution
118	VIOlight Ultraviolet Travel Toothbrush Sanitizers
119	WearEver Deluxe Aluminum Hi-Back Backpack Chair
120	X Rocker V-Rocker SE Wireless Game Chair

B.4 Instructions

B.4.1 Experiment I: Mousetracking

Part 1

Thank you for participating in this study of consumer preferences for various products. Please follow these instructions carefully and do not hesitate to ask the experimenter if you have a question. This task will take up to 15 minutes and you will be paid $15 for completion of this task, including a $5 show-up fee. Upon completion of this task, we may invite you to participate in another experiment.

This experiment consists of 120 trials, during which time we will show you 120 images of different consumer products. In each trial, one product is shown and you will be asked to state the maximum amount of money that you would be willing to pay to buy this one item; this amount is referred to as your willingness-to-pay. Please determine this amount under the restriction that, whatever you buy, IT MUST BE FOR YOURSELF (i.e., it cannot be purchased as a gift for someone else or for resale).

Note that there will be no actual purchase involved. Whatever amount you state, it is not binding; that is, you will not actually have to buy any of the items shown to you.

Although this is a purely hypothetical task, please keep in mind the following points when reporting your maximum willingness to pay.

- You should rate the value of each item independently from the others, assuming during each trial that the product shown is the only purchase you would make.

- The products should be evaluated from your perspective, not that of someone else. In other words, your willingness-to-pay should reflect how much you would like to keep an item for yourself, not for your friends or family, etc.

- Your current ownership of a particular item might affect your willingness to pay for the item—this is perfectly appropriate. For example, if you already own item A, your willingness to pay for item A might be high or low depending on whether you want a second one for yourself or not.

In each trial you will be allowed to enter an amount between $0 and $50 using a sliding scale. You can change the dollar amount by pressing the UP, DOWN, LEFT and RIGHT arrows keys on the keyboard. The effect of each key upon the willingness-to-pay value is described below:

RIGHT	+$1
LEFT	−$1
UP	+$5
DOWN	−$5

Amounts are entered by pressing the SPACE BAR. Other keys will not work. If you have any questions, please ask the experimenter.

Part 2

You are invited to take part in an experiment on decision making. This experiment consists of three different parts and will generally take up to 45 minutes. We will describe the details of each part of the experiment as it comes up. Upon finishing the entire experiment, you will be paid $50, including the $15 that you have earned from the previous task.

In the first part, we will show you different consumer products, one at a time, each with its own sale price. Your task is to make a hypothetical purchase decision. Assume that you are being offered the chance to buy the product (only one unit) from us at the end of the experiment with the $50 given to you, and that if you bought the product, you had to keep it only for yourself (i.e., you cannot give or sell it to someone else). This is a hypothetical exercise as in the previous task—you are not actually being offered the chance to buy anything. However, please take every decision seriously—when evaluating products for this hypothetical purchase, please assume that the product is only for yourself and treat every decision as if it were the only one. This last point is important—even though you are going to be presented many different products and making a decision for each, please make each decision as if that product was the only one you were thinking about buying at this time.

How the product image and the offer price are displayed At the beginning of each trial, you will see two gray boxes (Figure 1 on the next page). A product image and the offered price are hidden behind the gray boxes. Click the bigger gray box with the left mouse-button and hold to see the product image (Figure 2 on the next page). If you release the button or the cursor leaves the boxed area, the image will disappear. Click the smaller gray box with the left mouse-button and hold to see the offered price (Figure 3 on the next page). There is no time limit on this task. You can take as much time as you wish to click and make a decision.

Figure 1

Figure 2

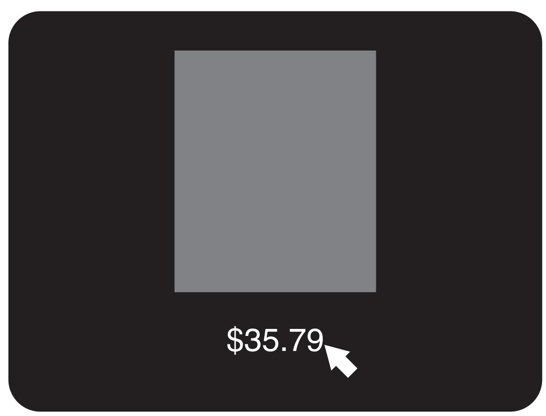


Figure 3

Once you decide whether to buy or not, your decision can be entered by pressing one of the two keys on the keyboard as described below:

No	Yes
z	c

You might find it easy to respond if you place your left ring finger on the ‘z’ key and your left index finger on the ‘c’ key. The key press will terminate the trial and a fixation cross at the center of the screen will appear briefly before a new trial begins. If you are left-handed and use your left-hand to hold a mouse, please let the experimenter know.

There will be 90 trials in this part and the same product might appear more than once with a different price each time.

All decisions are hypothetical and will not be implemented. However, please take each decision seriously.

In order to familiarize you with the software used in this task, we will present you with 5 practice trials.

Part 3

In this part, we will show you different consumer products one at a time with a different offer price. The procedure of this part is almost the same as in the previous decision making part. However, the most significant difference is that in this part we ask you to make a real purchase decision. It is real in the sense that any decision made in this part can count as real at the end of the experiment and you may actually be purchasing the product—more on this follows herein.

During the next 90 trials in this part, you will see various consumer products, which are different from those presented in the earlier hypothetical purchase decision making part of the experiment. In each trial, you are offered the chance to buy a product (only one unit) from us at the end of this experiment at the price listed below the product image. So in each trial, your task is to decide whether or not you want to buy the product from us at the stated price. At the end of the entire experiment, exactly one of the 90 trials will be chosen at random, and whatever decision you made in the chosen trial (to buy or not at the offered price) will be *carried out for real by us at the end of this experiment*!

When you make an actual purchase decision, note the following points:

- Since only one decision will count, you do not have to spread out your funds among the different purchase decisions. Therefore, you should treat each choice as if it is the only one that you are making. Indeed, only one trial will be chosen at the end of the experiment to be carried out for real.
- If in the selected trial you chose to purchase the item, the cost will be deducted from your $50 earnings; you get the item and the remaining cash. If in the selected trial you did not choose to purchase the item, you keep your $50 in earnings in cash and do not receive any product.
- If you buy an item from us, we will ship it to you and pay the shipping costs.

We would like to stress that honesty is the best policy here. Any of the 90 trials has an equal chance of being chosen, whether or not you expressed an interest in purchasing the item—that is, your decision about purchasing DOES NOT affect the chance of a particular trial being chosen. For example, if you were to decline purchasing every item presented to you, except for the one item you really want, you do not increase your chances of getting that item—you have only increased the chance that you will not get any item and you may miss out on other deals you would have liked. In each trial, you should make a purchase decision, independent of anything you have seen in any other trial.

Another important note is that any item you buy here must be for personal use. You should not buy the product in order to resell it or to give it to someone else—only consider whether or not you want to purchase the item for your own personal use. Your participation in this experiment is covered by the Caltech Honor Code, including your agreement to follow these instructions honestly and in particular, to evaluate items only for your personal use. Thinking about the value of the product in terms of its resale or gift value impairs our ability to understand the scientific basis of personal valuation.

As before, at the beginning of each trial you will see two gray boxes, behind which a product image and the offered price are hidden. Click the gray boxes with the left mouse-button to see the product image and the offer price. There is no time limit on this task. You can take as much time as you wish to click and make a decision.

Once you decide whether to buy or not, your decision can be entered by pressing one of the two keys on the keyboard as described below:

No	Yes
z	c

Again, keep in mind that you are asked to make real purchase decisions in this part. One of the decisions you make will be actually implemented.

If you have any questions or if anything is unclear, please read the instructions again or ask the experimenter.

Part 4

This is the last part of the experiment and consists of 90 trials. In this part, we will ask you again to make real purchase decisions on the same items you have already seen in the previous hypothetical purchase decision making part of the experiment. This task is identical to the previous purchase decision making tasks.

Note that at the end of the experiment, exactly one of the 180 real trials (90 from the real purchase decision making part that you have just finished and 90 from this part) will be randomly selected and the decision you made in the chosen trial will be implemented for real (i.e., based on your choice in the selected trial, you might buy that item at the suggested price).

As before, at the beginning of each trial you will see two gray boxes, behind which a product image and the offered price are hidden. Click the gray boxes with the left mouse-button to see the product image and the offer price. There is no time limit on this task. You can take as much time as you wish to click and make a decision.

Once you decide whether to buy or not, your decision can be entered by pressing one of the two keys on the keyboard as described below:

No	Yes
z	c

Keep in mind that you are asked to make real purchase decisions in this part. One of the decisions you make or have already made will be actually implemented.

B.4.2 Experiment II: Eyetracking

Part 1

Thank you for participating in this study of consumer preferences for various products. Please follow these instructions carefully and do not hesitate to ask the experimenter if you have a question. This task will take up to 15 minutes and you will be paid $15 for completion of this task, including a $5 show-up fee. Upon completion of this task, we may invite you to participate in another experiment.

This experiment consists of 120 trials, during which time we will show you 120 images of different consumer products. In each trial, one product is shown and you will be asked to state the maximum amount of money that you would be willing to pay to buy this one item; this amount is referred to as your willingness-to-pay. Please determine this amount under the restriction that, whatever you buy, IT MUST BE FOR YOURSELF (i.e., it cannot be purchased as a gift for someone else or for resale).

Note that there will be no actual purchase involved. Whatever amount you state, it is not binding; that is, you will not actually have to buy any of the items shown to you.

Although this is a purely hypothetical task, please keep in mind the following points when reporting your maximum willingness to pay.

- You should rate the value of each item independently from the others, assuming during each trial that the product shown is the only purchase you would make.
- The products should be evaluated from your perspective, not that of someone else. In other words, your willingness-to-pay should reflect how much you would like to keep an item for yourself, not for your friends or family, etc.
- Your current ownership of a particular item might affect your willingness to pay for the item—this is perfectly appropriate. For example, if you already own item A, your willingness to pay for item A might be high or low depending on whether you want a second one for yourself or not.

In each trial you will be allowed to enter an amount between \$0 and \$50 using a sliding scale. You can change the dollar amount by pressing the UP, DOWN, LEFT and RIGHT arrows keys on the keyboard. The effect of each key upon the willingness-to-pay value is described below:

RIGHT	+\$1
LEFT	−\$1
UP	+\$5
DOWN	−\$5

Amounts are entered by pressing the SPACE BAR. Other keys will not work. If you have any questions, please ask the experimenter.

Part 2

You are invited to take part in an experiment on decision making. This experiment consists of three different parts and will generally take up to 45 minutes. We will describe the details of each part of the experiment as it comes up. Upon finishing the entire experiment, you will be paid $50, including the $15 that you have earned from the previous task.

In the first part, we will show you different consumer products, one at a time, each with its own sale price (see Figure 1 below). Your task is to make a hypothetical purchase decision. Assume that you are being offered the chance to buy the product (only one unit) from us at the end of the experiment with the $50 given to you, and that if you bought the product, you had to keep it only for yourself (i.e., you cannot give or sell it to someone else). This is a hypothetical exercise as in the previous task—you are not actually being offered the chance to buy anything. However, please take every decision seriously—when evaluating products for this hypothetical purchase, please assume that the product is only for yourself and treat every decision as if it were the only one. This last point is important—even though you are going to be presented many different products and making a decision for each, please make each decision as if that product was the only one you were thinking about buying at this time.

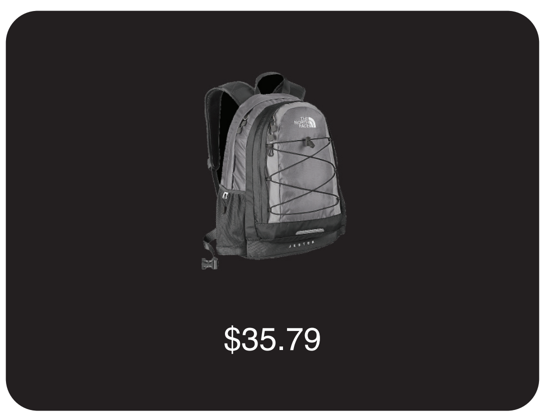


Figure 1

Once you decide whether to buy or not, your decision can be entered by pressing one of the two keys on the keyboard as described below:

No	Yes
z	c

You might find it easy to respond if you place your left ring finger on the ‘z’ key and your left index finger on the ‘c’ key. The key press will terminate the trial and a fixation cross at the center of the screen will appear briefly before a new trial begins. There is no time limit on this task. You can take as much time as you wish to make a decision.

At the beginning of each trial, you will see a blank screen with a fixation marker at the center. Please fixate on this central marker before proceeding—this is performed so the eyetracker can adjust to small head movements. Then press the space bar to begin the task.

There will be 90 trials in this part and the same product might appear more than once with a different price each time.

All decisions are hypothetical and will not be implemented. However, please take each decision seriously.

In order to familiarize you with the software used in this task, we will present you with 5 practice trials.

Part 3

In this part, we will show you different consumer products one at a time with a different offer price. The procedure of this part is almost the same as in the previous decision making part. However, the most significant difference is that in this part we ask you to make a real purchase decision. It is real in the sense that any decision made in this part can count as real at the end of the experiment and you may actually be purchasing the product—more on this follows herein.

During the next 90 trials in this part, you will see various consumer products, which are different from those presented in the earlier hypothetical purchase decision making part of the experiment. In each trial, you are offered the chance to buy a product (only one unit) from us at the end of this experiment at the price listed on the screen. So in each trial, your task is to decide whether or not you want to buy the product from us at the stated price. At the end of the entire experiment, exactly one of the 90 trials will be chosen at random, and whatever decision you made in the chosen trial (to buy or not at the offered price) will be *carried out for real by us at the end of this experiment*!

When you make an actual purchase decision, note the following points:

- Since only one decision will count, you do not have to spread out your funds among the different purchase decisions. Therefore, you should treat each choice as if it is the only one that you are making. Indeed, only one trial will be chosen at the end of the experiment to be carried out for real.

- If in the selected trial you chose to purchase the item, the cost will be deducted from your $50 earnings; you get the item and the remaining cash. If in the selected trial you did not choose to purchase the item, you keep your $50 in earnings in cash and do not receive any product.

- If you buy an item from us, we will ship it to you and pay the shipping costs.

We would like to stress that honesty is the best policy here. Any of the 90 trials has an equal chance of being chosen, whether or not you expressed an interest in purchasing the item—that is, your decision about purchasing DOES NOT affect the chance of a particular trial being chosen. For example, if you were to decline purchasing every item presented to you, except for the one item you really want, you do not increase your chances of getting that item—you have only increased the chance that you will not get any item and you may miss out on other deals you would have liked. In each trial, you should make a purchase decision, independent of anything you have seen in any other trial.

Another important note is that any item you buy here must be for personal use. You should not buy the product in order to resell it or to give it to someone else—only consider whether or not you want to purchase the item for your own personal use. Your participation in this experiment is covered by the Caltech Honor Code, including your agreement to follow these instructions honestly and in particular, to evaluate items only for your personal use. Thinking about the value of the product in terms of its resale or gift value impairs our ability to understand the scientific basis of personal valuation.

As before, at the beginning of each trial you will see a blank screen with a fixation circle at the center. Please fixate on this central marker before proceeding. After fixating on the marker for a moment, you can proceed by pressing the space bar. There is no time limit on this task. You can take as much time as you wish to make a decision.

Once you decide whether to buy or not, your decision can be entered by pressing one of the two keys on the keyboard as described below:

No	Yes
z	c

Again, keep in mind that you are asked to make real purchase decisions in this part. One of the decisions you make will be actually implemented.

If you have any questions or if anything is unclear, please read the instructions again or ask the experimenter.

Part 4

This is the last part of the experiment and consists of 90 trials. In this part, we will ask you again to make real purchase decisions on the same items you have already seen in the previous hypothetical purchase decision making part of the experiment. This task is identical to the previous purchase decision making tasks.

Note that at the end of the experiment, exactly one of the 180 real trials (90 from the real purchase decision making part that you have just finished and 90 from this part) will be randomly selected and the decision you made in the chosen trial will be implemented for real (i.e., based on your choice in the selected trial, you might buy that item at the suggested price).

As before, there is no time limit on this task. You can take as much time as you wish to make a decision.

Once you decide whether to buy or not, your decision can be entered by pressing one of the two keys on the keyboard as described below:

No	Yes
z	c

Keep in mind that you are asked to make real purchase decisions in this part. One of the decisions you make or have already made will be actually implemented.

Appendix C

Appendix to Chapter 4

C.1 Behavioral Responses to a Change in Payoff Structure

As we repeatedly mentioned in the series of analyses, values of the panels during the first two rounds were all doubled on November 26, 2001. Since other features of the game were unchanged, this one-time structural change in payoff structure can be seen as a natural experiment (inside the natural experiment). In this section we investigate how players responded to this structural change.

In order to gain a sufficiently large sample size, we focus on the *Daily Double* situation rather than *Final Jeopardy!* wagering. As we described in Section 4.2, special panels called *Daily Doubles* are hidden on the boards during the first two rounds. Unlike the case with usual panels, only the contestant who selects the panel has an opportunity to respond. Before the clue is revealed, the contestant has to decide how much to bet (the minimum amount is $5 and the maximum is bigger one of her/his current score and the highest dollar clue in the round).

A preference relation $\succsim$ is *homothetic* if an agent is indifferent between two options c_1 and c_2 then she is indifferent between λc_1 and λc_2 for any nonnegative scaling factor λ. In other words, homotheticity requires that ranking of any two payoffs is not reversed if all amounts are scaled by the same constant. It is essentially a scale invariance property and is ubiquitous in models in macroeconomics and finance (e.g., Epstein and Zin, 1991). The class of constant relative risk aversion (CRRA) utility functions is a special case of homothetic preferences. Comparing wagering decisions in *Daily Double* situations before and after the structural change, we aim to test whether the scale invariance property holds. Essentially, this is testing if players have homothetic preferences and the population distribution of risk preferences is stable over time. Note that the hypothesis we are testing is *joint* of homotheticity and stability of preferences. The latter is unavoidable since we do not observe same player making decisions both before and after the structural change.

To get some idea regarding how players wager in general, we estimate the following model:

$$\begin{aligned} W_i = \beta_0 + \beta_1 \mathit{Timing}_i + \beta_2 x_i + \beta_3 \mathit{Diff}_i \\ + \beta_4 \mathit{Position}_i + \beta_5 \mathit{Male}_i + \beta_6 \mathit{Gain}_i + \varepsilon_i, \end{aligned} \tag{C.1}$$

where *Timing* $\in (0, 1]$ codes when this *Daily Double* panel was opened (number of panels opened so far, including the current one, divided by the total number of panels opened in the round), *Diff* codes the absolute difference between i's score and the highest score of the rest of the two players (i.e., $|x_i - \max_{j \neq i} x_j|$), *Position* $\in \{1, \ldots, 5\}$ codes on which row the panel was hidden (possibly reflecting difficulty level of the question), *Male* is a dummy variable for male players, and *Gain* is the value added to the player's score immediately before picking a *Daily Double* panel. With slight abuse of notation, here we represent score and wagering amounts at the time of *Daily Double* by x_i and y_i.

We estimate equation (C.1) for round 1 and 2, leading players and trailing players, separately. The first four columns in Table C.1 present estimation results where $W_i = y_i$, i.e., raw wagering amounts, as the dependent variable. Similarly, the first four columns in Table C.2 present estimation results where $W_i = y_i / x_i$, i.e., share of wager.

Table C.1: Determinants of wager in *Daily Doubles*. Dependent variable is raw wager amount.

Dependent variable *Wager* (y_i)		(1) Round 1 Leading	(2) Round 1 Trailing	(3) Round 2 Leading	(4) Round 2 Trailing	(5) Round 1 Leading	(6) Round 1 Trailing	(7) Round 2 Leading	(8) Round 2 Trailing
Constant	β_0	704.1***	587.1***	1397.6***	613.4***	345.1***	329.9***	765.0***	281.3**
		(59.77)	(53.54)	(105.9)	(91.67)	(59.59)	(50.58)	(116.0)	(91.01)
Double	γ_0					495.2***	391.9***	991.3***	539.5***
						(38.04)	(41.90)	(119.9)	(79.63)
Timing	β_1	−743.1***	−537.0***	−1899.3***	−885.7***	−211.3*	−190.0***	−1199.8***	−475.6***
		(93.38)	(55.98)	(94.07)	(75.50)	(103.9)	(56.25)	(116.4)	(92.63)
Timing × *Double*	γ_1					−86.79	−49.62	−1192.6***	−578.7***
						(200.0)	(93.08)	(185.0)	(139.6)
Score (*K*)	β_2	316.2***	348.3***	160.9***	251.3***	217.9***	312.5***	191.7***	263.2***
		(24.63)	(16.44)	(10.94)	(9.136)	(40.18)	(28.24)	(20.69)	(17.85)
Score × *Double*	γ_2					−15.18	−43.25	−32.80	−21.18
						(36.37)	(31.15)	(20.21)	(19.70)
Diff (*K*)	β_3	−41.21	115.7***	−7.577	153.4***	14.77	65.61***	−9.564	149.1***
		(25.98)	(9.833)	(12.35)	(7.312)	(30.08)	(13.10)	(13.50)	(9.324)
Position	β_4	−21.40	−17.27	−48.14	−52.62*	5.245	−4.529	−46.34	−45.40
		(16.82)	(12.66)	(27.73)	(23.31)	(16.89)	(12.51)	(28.01)	(23.51)
Male	β_5	119.3***	110.9***	221.6***	460.4***	138.8***	120.7***	234.6***	461.4***
		(28.81)	(23.14)	(42.38)	(37.71)	(28.15)	(22.68)	(41.99)	(37.78)
Gain (*K*)	β_6	273.9***	226.9**	317.1***	169.4**	74.07	95.60	286.4***	141.6**
		(72.72)	(53.05)	(80.66)	(47.76)	(77.08)	(55.17)	(82.47)	(48.60)
N		2,623	1,917	4,354	4,653	2,623	1,917	4,354	4,653
Adjusted R^2		0.327	0.408	0.259	0.342	0.357	0.436	0.272	0.346

Notes: Robust standard errors are in parentheses. Stars indicate significance level. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

TABLE C.2: Determinants of wager in *Daily Doubles*. Dependent variable is wager amount divided by maximum possible wager.

Dependent variable *Share* (y_i/x_i)		(1) Round 1 Leading	(2) Round 1 Trailing	(3) Round 2 Leading	(4) Round 2 Trailing	(5) Round 1 Leading	(6) Round 1 Trailing	(7) Round 2 Leading	(8) Round 2 Trailing
Constant	β_0	0.931***	0.952***	0.390***	0.622***	0.912***	0.978***	0.409***	0.673***
		(0.021)	(0.026)	(0.011)	(0.017)	(0.023)	(0.031)	(0.015)	(0.020)
Double	γ_0					−0.000	−0.054	−0.019	−0.055**
						(0.017)	(0.028)	(0.015)	(0.018)
Timing	β_1	−0.617***	−0.341***	−0.222***	−0.172***	−0.368***	−0.218***	−0.207***	−0.059*
		(0.026)	(0.030)	(0.010)	(0.015)	(0.044)	(0.043)	(0.016)	(0.025)
Timing × *Double*	γ_1					−0.016	−0.003	0.002	−0.112***
						(0.060)	(0.057)	(0.019)	(0.031)
Score (*K*)	β_2	−0.014*	−0.086***	−0.005***	−0.033***	−0.119***	−0.199***	−0.010***	−0.073***
		(0.005)	(0.006)	(0.001)	(0.001)	(0.014)	(0.016)	(0.002)	(0.004)
Score × *Double*	γ_2					0.066***	0.105***	0.004*	0.042***
						(0.013)	(0.017)	(0.002)	(0.004)
Diff	β_3	−0.022***	0.044***	0.000	0.026***	0.002	0.030***	0.001	0.024***
		(0.006)	(0.004)	(0.001)	(0.001)	(0.007)	(0.006)	(0.001)	(0.001)
Position	β_4	−0.003	−0.006	−0.003	−0.007	0.002	−0.003	−0.003	−0.005
		(0.005)	(0.006)	(0.003)	(0.004)	(0.005)	(0.006)	(0.003)	(0.004)
Male	β_5	0.049***	0.066***	0.024***	0.090***	0.056***	0.067***	0.025***	0.094***
		(0.010)	(0.011)	(0.004)	(0.007)	(0.009)	(0.011)	(0.004)	(0.007)
Gain (*K*)	β_6	0.010	0.072**	0.021***	0.021**	−0.012	0.054*	0.020***	0.017*
		(0.020)	(0.024)	(0.005)	(0.007)	(0.021)	(0.025)	(0.005)	(0.007)
N		2,623	1,917	4,354	4,653	2,623	1,917	4,354	4,653
Adjusted R^2		0.477	0.355	0.229	0.358	0.496	0.376	0.230	0.378

Notes: Robust standard errors are in parentheses. Stars indicate significance level. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

Since main messages are the same we focus on Table C.1. Positive and significant β_1 and β_6 are as expected. The difference between own score and the maximum of others' scores ($|x_i - \max_{j\neq 1} x_j|$) is significant only for trailing players. Furthermore, positive β_3 implies that those trailing players wager more as the difference from the leading player becomes larger, in order to fill this gap. The coefficient on *Timing* is negative and significant, implying that players become conservative in making wagering decisions as the timing of opening *Daily Double* panels gets later. The effect of position of *Daily Double* panels is negative, but not significant. There is also a "gain effect": players who earned more immediately prior to *Daily Double* panel tended to wager more. The finding is similar to what is commonly known as the "house money effect," where prior losses increase risk aversion and prior gains increase risk seeking (Thaler and Johnson, 1990). Finally, there is a strong gender effect as discussed in Section 4.7: male players wager more aggressively compared to female players.

In order to test the scale invariance property, we extend equation (C.1) by incorporating several dummy variables:

$$\begin{aligned} W_i = \beta_0 + \gamma_0 Double + \beta_1 Timing_i + \gamma_1 Timing \times Double \\ + \beta_2 x_i + \gamma_2 x_i \times Double \\ + \beta_3 Diff_i + \beta_4 Position_i + \beta_5 Male_i + \beta_6 Gain_i + \varepsilon_i, \end{aligned} \tag{C.2}$$

where *Double* takes 1 if the game is played after panel values were doubled. If investors have the same homothetic preference but different wealth levels, they will allocate to risky assets the same fraction of their respective wealth. This observation provides several constraints on coefficients. First, when we take $W_i = y_i$ as the dependent variable, the set of coefficients $(\gamma_0, \gamma_1, \gamma_2)$ which satisfies

$$\gamma_0 = \beta_0 \tag{SI-w-1}$$

$$\gamma_1 = \beta_1 \tag{SI-w-2}$$

$$\gamma_2 = 0 \tag{SI-w-3}$$

is consistent with predictions of scale invariance property. Second, when we take $W_i = y_i / x_i$ as the dependent variable, the set of coefficients $(\gamma_0, \gamma_1, \gamma_2)$ which

Table C.3: p-values from scale-invariance tests

	Round 1 Leading	Round 1 Trailing	Round 2 Leading	Round 2 Trailing
$W_i = y_i$				
H_0: $\gamma_0 = \beta_0$	0.0519	0.3904	0.2526	0.0502
H_0: $\gamma_1 = \beta_1$	0.6026	0.2085	0.9784	0.5989
H_0: $\gamma_2 = 0$	0.6765	0.1652	0.1047	0.2824
$W_i = y_i/x_i$				
H_0: $\gamma_0 = 0$	0.9976	0.0502	0.2027	0.0019
H_0: $\gamma_1 = 0$	0.7983	0.9561	0.9344	0.0002
H_0: $\gamma_2 = -\beta_2/2$	0.4705	0.5823	0.5200	0.0146

satisfies

$$\gamma_0 = 0 \tag{SI-s-1}$$

$$\gamma_1 = 0 \tag{SI-s-2}$$

$$\gamma_2 = -\beta_2/2 \tag{SI-s-3}$$

is consistent with predictions of scale invariance property. The idea behind those sets of constraints is worth discussing. Constraints (SI-w-1) and (SI-s-1) require that after panel values were doubled players' baseline wager amounts were also doubled while the wager shares were kept constant. Similarly, constraints (SI-w-3) and (SI-s-3) require that doubling panel values should influence the effect of timing on wagering solely through the wealth effect.

The results of estimating equation (C.2) are presented in columns (5) to (8) in Tables C.1 and C.2. The signs and significance of baseline coefficients (β's) are in line with the first four columns in each Table. We now examine implications of scale invariance. Table C.3 presents p-values from a series of Wald tests for conditions (SI-w) and (SI-s). Those values imply that wagering patterns before and after the structural change are consistent with predictions from the joint assumption of homotheticity and stability, with a notable exception of the trailing players in *Double Jeopardy!* round where players were in general more conservative and became significantly more conservative as the timing gets later.

Appendix D

Appendix to Chapter 5

D.1 The Favorite-Longshot Bias in the Literature

TABLE D.1: Evidence on the favorite-longshot bias.

Study	Market	Event	Country	Bias
Dowie (1976)	Fixed-odds	Racetrack	U.K.	FLB
Henery (1985)	Fixed-odds	Racetrack	U.K.	FLB
Bird et al. (1987)	Fixed-odds	Racetrack	Australia	FLB
Vaughan Williams and Paton (1997)	Fixed-odds	Racetrack	U.K.	FLB
Jullien and Salanié (2000)	Fixed-odds	Racetrack	U.K.	FLB
Direr (2013)	Fixed-odds	Soccer	U.K.	FLB
Lahvička (2014)	Fixed-odds	Tennis	U.K.	FLB
Griffith (1949)	Parimutuel	Racetrack	U.S.	FLB
McGlothlin (1956)	Parimutuel	Racetrack	U.S.	FLB
Ali (1977)	Parimutuel	Racetrack	U.S.	FLB
Snyder (1978)	Parimutuel	Racetrack	U.S.	FLB
Hausch et al. (1981)	Parimutuel	Racetrack	U.S.	FLB
Asch et al. (1982)	Parimutuel	Racetrack	U.S.	FLB
Busche and Hall (1988)	Parimutuel	Racetrack	Hong Kong	No bias
Busche (1994)	Parimutuel	Racetrack	Hong Kong, Japan	No bias
Golec and Tamarkin (1998)	Parimutuel	Racetrack	U.S.	FLB
Gandar et al. (2001)	Parimutuel	Racetrack	New Zealand	Reverse FLB
Walls and Busche (2003)	Parimutuel	Racetrack	Hong Kong, Japan	No bias
Gramm and Owens (2005)	Parimutuel	Racetrack	U.S.	FLB
Snowberg and Wolfers (2010)	Parimutuel	Racetrack	U.S.	FLB
Gandhi and Serrano-Padial (2015)	Parimutuel	Racetrack	U.S.	FLB
Zuber et al. (1985)	Point spread	NFL	U.S.	FLB
Gandar et al. (1988)	Point spread	NFL	U.S.	No bias
Woodland and Woodland (1994)	Point spread	MLB	U.S.	Reverse FLB
Sauer (1998)	Point spread	NFL	U.S.	FLB
Woodland and Woodland (2001)	Point spread	NHL	U.S.	Reverse FLB
Woodland and Woodland (2003)	Point spread	MLB	U.S.	Reverse FLB
Tetlock (2008)	Betting exchange	Sports, financial	U.S.	FLB
Hartzmark and Solomon (2012)	Betting exchange	NFL	U.S.	FLB
Page and Clemen (2013)	Betting exchange	Politics, sports	U.S.	FLB

D.2 Additional Figures

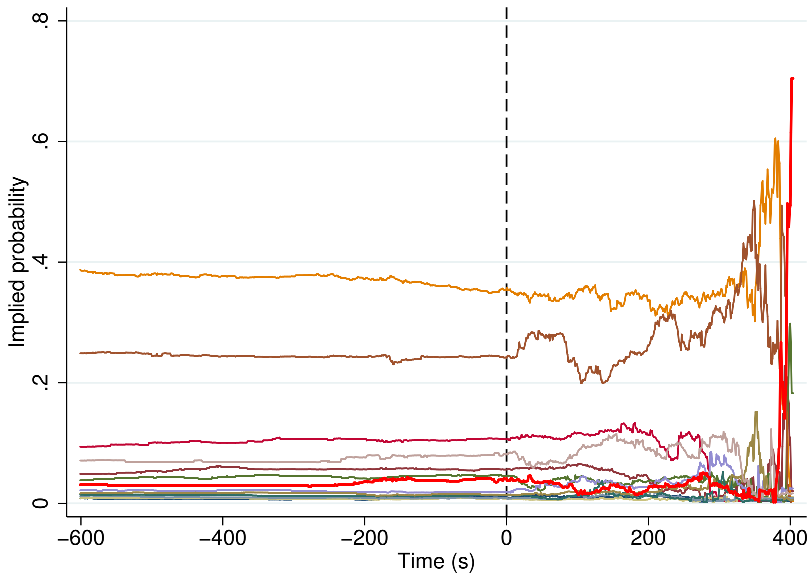


Figure D.1: Pre- and in-race dynamics of implied probabilities.

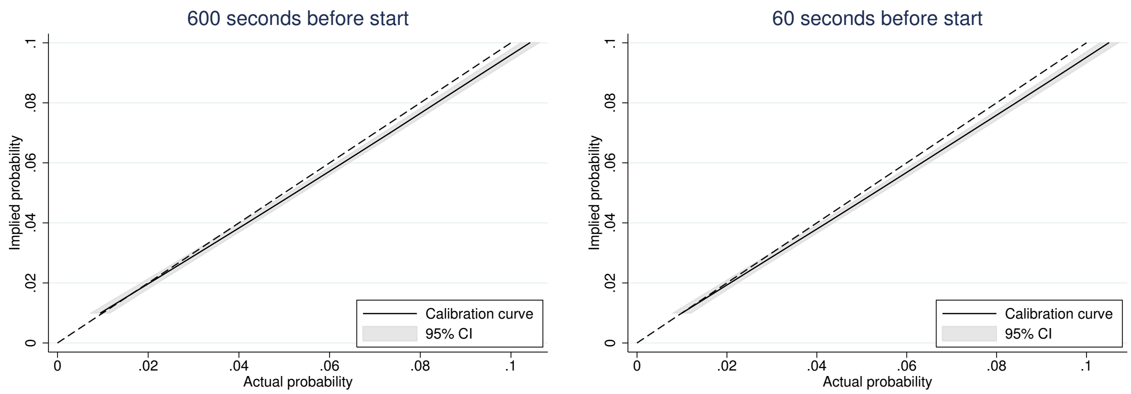


FIGURE D.2: Nonparametric estimation of calibration curves ϕ_t^s, $t \in \{600, 60\}$ seconds before races start for $p \leq 0.1$.

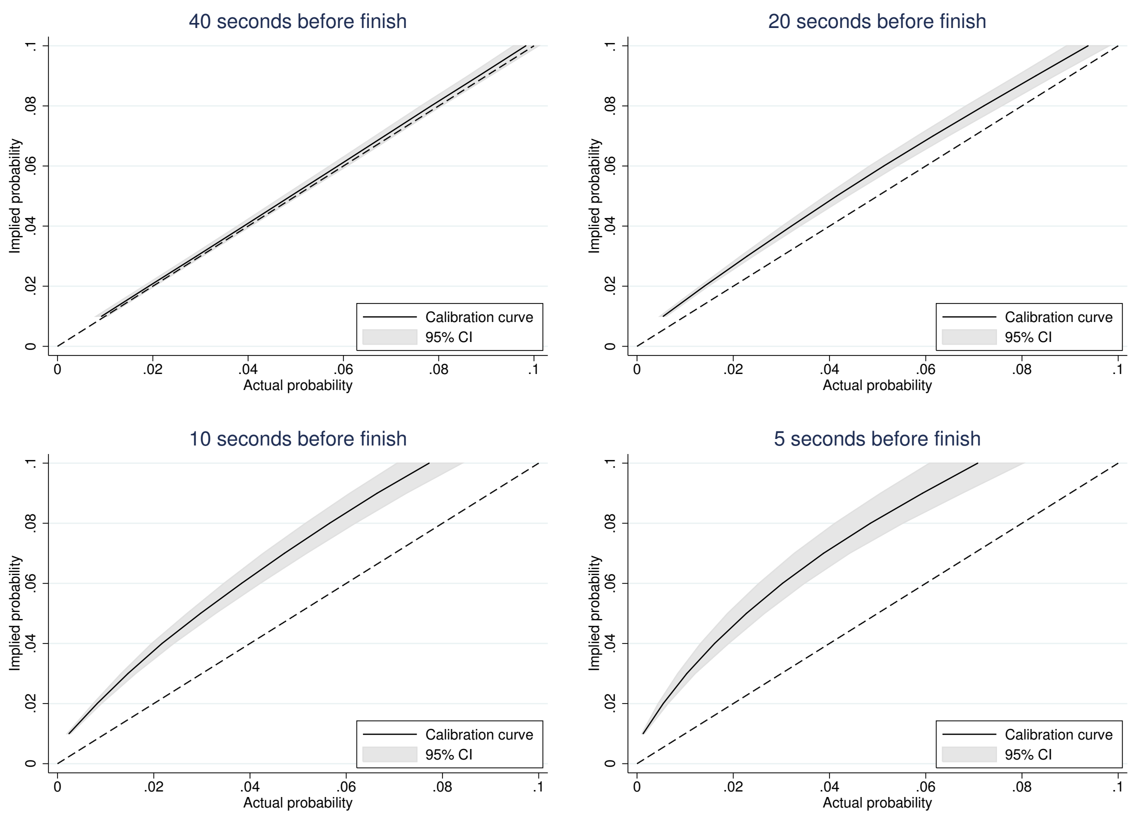


FIGURE D.3: Nonparametric estimation of calibration curves ϕ_t^f, $t \in \{40, 20, 10, 5\}$ seconds before races finish for $p \leq 0.1$.

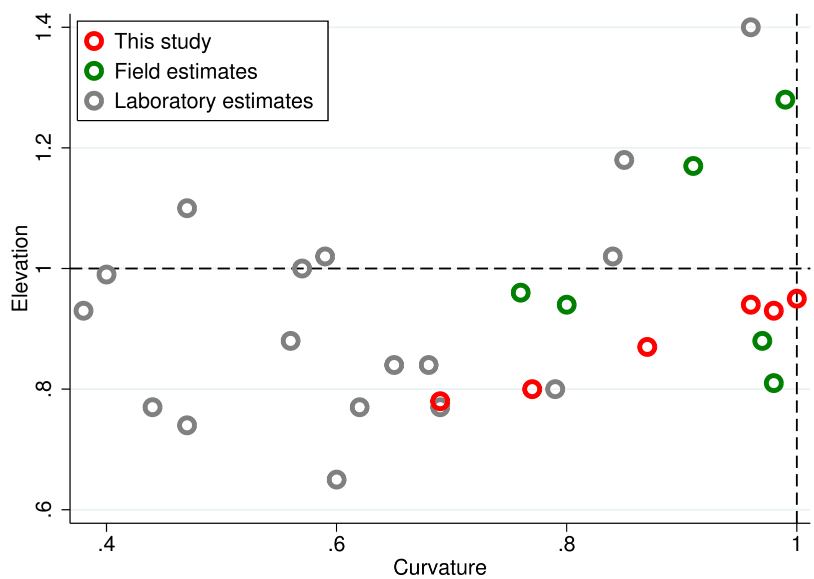


FIGURE D.4: Comparing estimated curvature and elevation across studies.

Bibliography

Abdellaoui, M. (2000): "Parameter-Free Elicitation of Utility and Probability Weighting Functions," *Management Science*, 46, 1497–1512.

Abdellaoui, M., O. l'Haridon, and H. Zank (2010): "Separating Curvature and Elevation: A Parametric Probability Weighting Function," *Journal of Risk and Uncertainty*, 41, 39–65.

Adams, A., L. Cherchye, B. De Rock, and E. Verriest (2014): "Consume Now or Later? Time Inconsistency, Collective Choice, and Revealed Preference," *American Economic Review*, 104, 4147–4183.

Ali, M. M. (1977): "Probability and Utility Estimates for Racetrack Bettors," *Journal of Political Economy*, 85, 803–815.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008a): "Eliciting Risk and Time Preferences," *Econometrica*, 76, 583–618.

——— (2008b): "Risk Aversion in Game Shows," in *Risk Aversion in Experiments*, ed. by G. W. Harrison and J. C. Cox, Bingley: Emerald, vol. 12 of *Research in Experimental Economics*, 361–406.

Andreoni, J., W. T. Harbaugh, and B. J. Gillen (2013): "The Power of Revealed Preference Tests: Ex-Post Evaluation of Experimental Design," Unpublished manuscript.

Andreoni, J., M. A. Kuhn, and C. Sprenger (2015): "On Measuring Time Preferences," *Journal of Economic Behavior & Organization*, 116, 451–464.

Andreoni, J. and J. Miller (2002): "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 70, 737–753.

Andreoni, J. and C. Sprenger (2012a): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–3356.

——— (2012b): "Uncertainty Equivalents: Linear Tests of the Independence Axiom," Unpublished manuscript.

Andrikogiannopoulou, A. and F. Papakonstantinou (2011): "Market Efficiency and Behavioral Biases in the Sports Betting Market," Unpbulished manuscript.

——— (2013): "Estimating Heterogeneous Risk Preferences from a Panel of Real-World Betting Choices," Unpbulished manuscript.

——— (2015): "History-Dependent Risk Preferences: Evidence from Individual Choices and Implications for the Disposition Effect," Unpbulished manuscript.

Anscombe, F. J. and R. J. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199–205.

Arlot, S. and A. Celisse (2010): "A Survey of Cross-Validation Procedures for Model Selection," *Statistics Surveys*, 4, 40–79.

Armel, K. C., A. Beaumel, and A. Rangel (2008): "Biasing Simple Choices by Manipulating Relative Visual Attention," *Judgment and Decision Making*, 3, 396–403.

Arrow, K. J., R. Forsythe, M. Gorham, R. Hahn, R. Hanson, J. O. Ledyard, S. Levmore, R. Litan, P. Milgrom, F. D. Nelson, G. R. Neumann, M. Ottaviani, T. C. Schelling, R. J. Shiller, V. L. Smith, E. Snowberg, C. R. Sunstein, P. C. Tetlock, P. E. Tetlock, H. R. Varian, J. Wolfers, and E. Zitzewitz (2008): "The Promise of Prediction Markets," *Science*, 320, 877–878.

Asch, P., B. G. Malkiel, and R. E. Quandt (1982): "Racetrack Betting and Informed Behavior," *Journal of Financial Economics*, 10, 187–194.

Ashton, L. (2014): "Hunger Games: Does Hunger Affect Time Preferences?" Unpublished manuscript.

Attanasio, O. P. and G. Weber (1989): "Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption," *Economic Journal*, 99, 59–73.

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*, 130, 1067–1115.

BAJARI, P., D. NEKIPELOV, S. P. RYAN, AND M. YANG (2015): "Demand Estimation with Machine Learning and Model Combination," NBER Working Paper No. 20955.

BAKER, M. D. AND J. K. MANER (2009): "Male Risk-Taking as a Context-Sensitive Signaling Device," *Journal of Experimental Social Psychology*, 45, 1136–1139.

BALAKRISHNAN, U., J. HAUSHOFER, AND P. JAKIELA (2015): "How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments," Unpublished manuscript.

BARBERIS, N. C. (2013): "Thirty Years of Prospect Theory in Economics: A Review and Assessment," *Journal of Economic Perspectives*, 27, 173–196.

BARCELLOS, S. H. AND L. CARVALHO (2014): "Information about Self-Control and Intertemporal Choices," Unpublished manuscript.

BARSEGHYAN, L., F. MOLINARI, T. O'DONOGHUE, AND J. C. TEITELBAUM (2013): "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 103, 2499–2529.

——— (2015): "Estimating Risk Preferences in the Field," Unpublished manuscript.

BEATTY, T. K. M. AND I. A. CRAWFORD (2011): "How Demanding is the Revealed Preference Approach to Demand?" *American Economic Review*, 101, 2782–2795.

BEETSMA, R. M. W. J. AND P. C. SCHOTMAN (2001): "Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo," *Economic Journal*, 111, 821–848.

BELLONI, A., D. CHEN, V. CHERNOZHUKOV, AND C. HANSEN (2012): "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain," *Econometrica*, 80, 2369–2429.

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "High-Dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economic Perspectives*, 28, 29–50.

BERG, J., R. FORSYTHE, F. NELSON, AND T. RIETZ (2008): "Results from a Dozen Years of Election Futures Markets Research," in *Handbook of Experimental Economics Results*, ed. by C. Plott and V. Smith, New York: Elsevier, vol. 1, 742–751.

BERNHEIM, B. D., D. BJÖRKEGREN, J. NAECKER, AND A. RANGEL (2015): "Non-Choice Evaluations Predict Behavioral Responses to Change in Economic Conditions," NBER Working Paper No. 19269.

BERNS, G. S., E. BELL, C. M. CAPRA, M. J. PRIETULA, S. MOORE, B. ANDERSON, J. GINGES, AND S. ATRAN (2012): "The Price of Your Soul: Neural Evidence for the Non-Utilitarian Representation of Sacred Values," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 754–762.

BIRD, R., M. MCCRAE, AND J. BEGGS (1987): "Are Gamblers Really Risk Takers?" *Australian Economic Papers*, 26, 237–253.

BLACKBURN, M., G. W. HARRISON, AND E. E. RUTSTRÖM (1994): "Statistical Bias Functions and Informative Hypothetical Surveys," *American Journal of Agricultural Economics*, 76, 1084–1088.

BLAMEY, R. K., J. W. BENNETT, AND M. D. MORRISON (1999): "Yea-Saying in Contingent Valuation Surveys," *Land Economics*, 75, 126–141.

BLAVATSKYY, P. AND G. POGREBNA (2008): "Risk Aversion When Gains are Likely and Unlikely: Evidence from a Natural Experiment with Large Stakes," *Theory and Decision*, 64, 395–420.

BLEICHRODT, H. (2001): "Probability Weighting in Choice under Risk: An Empirical Test," *Journal of Risk and Uncertainty*, 23, 185–198.

BLEICHRODT, H. AND J. L. PINTO (2000): "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis," *Management Science*, 46, 1485–1496.

BLISS, R. T., M. E. POTTER, AND C. SCHWARZ (2012): "Decision Making and Risk Aversion in the Cash Cab," *Journal of Economic Behavior & Organization*, 84, 163–173.

BLUMENSCHEIN, K., G. C. BLOMQUIST, M. JOHANNESSON, N. HORN, AND P. FREEMAN (2008): "Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment," *Economic Journal*, 118, 114–137.

BOHM, P. (1972): "Estimating Demand for Public Goods: An Experiment," *European Economic Review*, 3, 111–130.

BOMBARDINI, M. AND F. TREBBI (2012): "Risk Aversion and Expected Utility Theory: An Experiment with Large and Small Stakes," *Journal of the European Economic Association*, 10, 1348–1399.

BOOIJ, A. S., B. M. S. VAN PRAAG, AND G. VAN DE KUILEN (2010): "A Parametric Analysis of Prospect Theory's Functionals for the General Population," *Theory and Decision*, 68, 115–148.

BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2012): "Salience Theory of Choice Under Risk," *Quarterly Journal of Economics*, 127, 1243–1285.

BORDER, K. C. (2013): "Alternative Linear Inequalities," Unpublished manuscript.

BOTTI, F., A. CONTE, D. T. DI CAGNO, AND C. D'IPPOLITI (2008): "Risk Attitude in Real Decision Problems," *B.E. Journal of Economic Analysis & Policy*, 8.

BRAINARD, D. H. (1997): "The Psychophysics Toolbox," *Spatial Vision*, 10, 433–436.

BROCAS, I., J. D. CARRILLO, AND A. SACHDEVA (2014a): "The Path to Equilibrium in Sequential and Simultaneous Games," Unpublished manuscript.

BROCAS, I., J. D. CARRILLO, AND J. TARRASÓ (2015): "How Long is a Minute?" Unpublished manuscript.

BROCAS, I., J. D. CARRILLO, S. W. WANG, AND C. F. CAMERER (2014b): "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games," *Review of Economic Studies*, 81, 944–970.

BRONARS, S. G. (1987): "The Power of Nonparametric Tests of Preference Maximization," *Econometrica*, 55, 693–698.

BROOKS, R., R. FAFF, D. MULINO, AND R. SCHEELINGS (2009): "Deal or No Deal, That is the Question: The Impact of Increasing Stakes and Framing Effects on Decision-Making under Risk," *International Review of Finance*, 9, 27–50.

BROWN, A. (2014): "Information Processing Constraints and Asset Mispricing," *Economic Journal*, 124, 245–268.

BROWNING, M. (1989): "A Nonparametric Test of the Life-Cycle Rational Expectations Hypothesis," *International Economic Review*, 30, 979–992.

BRUHIN, A., H. FEHR-DUDA, AND T. EPPER (2010): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Econometrica*, 78, 1375–1412.

Busche, K. (1994): "Efficient Market Results in an Asian Setting," in *Efficiency of Racetrack Betting Markets*, ed. by D. B. Hausch, S. Y. Lo, and W. T. Ziemba, London: World Scientific, 183–198.

Busche, K. and C. D. Hall (1988): "An Exception to the Risk Preference Anomaly," *Journal of Business*, 61, 337–346.

Busche, K. and W. D. Walls (2000): "Decision Costs and Betting Market Efficiency," *Rationality and Society*, 12, 477–492.

Byrnes, J. P., D. C. Miller, and W. D. Schafer (1999): "Gender Differences in Risk Taking: A Meta-Analysis," *Psychological Bulletin*, 125, 367.

Caldara, R. and S. Miellet (2011): "*i*Map: A Novel Method for Statistical Fixation Mapping of Eye Movement Data," *Behavior Research Methods*, 43, 864–878.

Callen, M., M. Isaqzadeh, J. D. Long, and C. Sprenger (2014): "Violence and Risk Preference: Experimental Evidence from Afghanistan," *American Economic Review*, 104, 123–148.

Camerer, C. (1995): "Individual Decision Making," in *Handbook of Experimental Economics*, ed. by J. H. Kagel and A. E. Roth, Princeton: Princeton University Press, 587–703.

Camerer, C. F. (1989): "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty*, 2, 61–104.

——— (1998): "Can Asset Markets be Manipulated? A Field Experiment with Racetrack Betting," *Journal of Political Economy*, 106, 457–482.

——— (2004): "Prospect Theory in the Wild: Evidence from the Field," in *Advances in Behavioral Economics*, ed. by C. F. Camerer, G. Loewenstein, and M. Rabin, Princeton: Princeton University Press, 148–161.

Camerer, C. F., A. Dreber, E. Forsell, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, E. Heikensten, F. Holzmeister, T. Imai, S. Isaksson, G. Nave, T. Pfeiffer, M. Razen, and H. Wu (2015): "Do Lab Experiments in Economics Replicate?" Unpublished manuscript.

Camerer, C. F. and T.-H. Ho (1994): "Violations of the Betweenness Axiom and Nonlinearity in Probability," *Journal of Risk and Uncertainty*, 8, 167–196.

Camerer, C. F., E. Johnson, T. Rymon, and S. Sen (1993): "Cognition and Framing in Sequential Bargaining for Gains and Losses," in *Frontiers of Game Theory*, ed. by K. Binmore, A. Kirman, and P. Tani, Cambridge: MIT Press, 27–47.

Camerer, C. F., G. Nave, and A. Smith (2014): "Dynamic Unstructured Bargaining with Private Information and Deadline: Theory and Experiment," Unpublished manuscript.

Carson, R. T. (2012): "Contingent Valuation: A Practical Alternative when Prices Aren't Available," *Journal of Economic Perspectives*, 26, 27–42.

Carson, R. T., N. E. Flores, K. M. Martin, and J. L. Wright (1996): "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods," *Land Economics*, 72, 80–99.

Carson, R. T. and W. M. Hanemann (2005): "Contingent Valuation," in *Handbook of Environmental Economics*, ed. by K.-G. Mäler and J. Vincent, Amsterdam: North-Holland, vol. 2, 821–936.

Carvalho, L., S. Meier, and S. W. Wang (forthcoming): "Poverty and Economic Decision Making: Evidence from Changes in Financial Resources at Payday," *American Economic Review*.

Carvalho, L., S. Prina, and J. Sydnor (2013): "The Effects of Savings on Risk Attitudes and Intertemporal Choices," Unpublished manuscript.

Cavagnaro, D. R., M. A. Pitt, R. Gonzalez, and J. I. Myung (2013): "Discriminating among Probability Weighting Functions Using Adaptive Design Optimization," *Journal of Risk and Uncertainty*, 47, 255–289.

Chambers, C. P. and F. Echenique (2014): "On the Consistency of Data with Bargaining Theories," *Theoretical Economics*, 9, 137–162.

Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum (1997): "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods," *Journal of Environmental Economics and Management*, 33, 151–162.

Chandon, P., V. G. Morwitz, and W. J. Reinartz (2004): "The Short- and Long-Term Effects of Measuring Intent to Repurchase," *Journal of Consumer Research*, 31, 566–572.

Charness, G. and U. Gneezy (2012): "Strong Evidence for Gender Differences in Risk Taking," *Journal of Economic Behavior & Organization*, 83, 50–58.

Chauvin, A., K. J. Worsley, P. G. Schyns, M. Arguin, and F. Gosselin (2005): "Accurate Statistical Tests for Smooth Classification Images," *Journal of Vision*, 5, 659–667.

Chen, K.-Y. and C. R. Plott (2002): "Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem," Unpublished manuscript.

Chetty, R. (2006): "A New Method of Estimating Risk Aversion," *American Economic Review*, 96, 1821–1834.

Chiappori, P.-A., B. Salanié, F. Salanié, and A. Gandhi (2012): "From Aggregate Betting Data to Individual Risk Preferences," Unpublished manuscript.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007): "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review*, 97, 1921–1938.

Cicchetti, C. J. and J. A. Dubin (1994): "A Microeconometric Analysis of Risk Aversion and the Decision to Self-Insure," *Journal of Political Economy*, 102, 169–186.

Cleveland, W. S. (1979): "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Cohen, A. and L. Einav (2007): "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97, 745–788.

Cohen, J. (2015): "What is *Jeopardy!*?" *Science*, 348, 493–495.

Coleman, L. (2004): "New Light on the Longshot Bias," *Applied Economics*, 36, 315–326.

Cornelissen, F. W., E. M. Peters, and J. Palmer (2002): "The Eyelink Toolbox: Eye Tracking with MATLAB and the Psychophysics Toolbox," *Behavior Research Methods, Instruments, & Computers*, 34, 613–617.

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001): "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69, 1193–1235.

Coventry, K. R. and A. C. Norman (1997): "Arousal, Sensation Seeking and Frequency of Gambling In Off-course Horse Racing Bettors," *British Journal of Psychology*, 88, 671–682.

Cowgill, B. and E. Zitzewitz (2015): "Corporate Prediction Markets: Evidence from Google, Ford, and Firm X," *Review of Economic Studies*, 82, 1309–1341.

Cramer, J. S. (1999): "Predictive Performance of the Binary Logit Model in Unbalanced Samples," *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48, 85–94.

Crawford, I. (2010): "Habits Revealed," *Review of Economic Studies*, 77, 1382–1402.

Crawford, I. and B. De Rock (2014): "Empirical Revealed Preference," *Annual Review of Economics*, 6, 503–524.

Crawford, I. and M. Polisson (2014): "Testing for Intertemporal Nonseparability," *Journal of Mathematical Economics*, 52, 46–49.

Croson, R. and U. Gneezy (2009): "Gender Differences in Preferences," *Journal of Economic Literature*, 47, 448–474.

Croxson, K. and J. J. Reade (2014): "Information and Efficiency: Goal Arrival in Soccer Betting," *Economic Journal*, 124, 62–91.

Cummings, R. G., G. W. Harrison, and E. E. Rutström (1995): "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review*, 85, 260–266.

Cummings, R. G. and L. O. Taylor (1999): "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method," *American Economic Review*, 89, 649–665.

Cutler, D. M., E. Richardson, T. E. Keeler, and D. Staiger (1997): "Measuring the Health of the U.S. Population," *Brookings Papers on Economic Activity: Microeconomics*, 1997, 217–271.

Daly, M. and M. Wilson (2001): "Risk-Taking, Intrasexual Competition, and Homicide," in *Symposium on Motivation*, vol. 47, 1–36.

Dawid, A. P. (1982): "The Well-Calibrated Bayesian," *Journal of the American Statistical Association*, 77, 605–610.

de Roos, N. and Y. Sarafidis (2010): "Decision Making under Risk in *Deal or No Deal*," *Journal of Applied Econometrics*, 25, 987–1027.

Dean, M. and D. Martin (forthcoming): "Measuring Rationality with the Minimum Cost of Revealed Preference Violations," *Review of Economics and Statistics*.

Deck, C., J. Lee, and J. Reyes (2008): "Risk Attitudes in Large Stake Gambles: Evidence from a Game Show," *Applied Economics*, 40, 41–52.

Demuynck, T. and E. Verriest (2013): "I'll Never Forget My First Cigarette: A Revealed Preference Analysis of the "Habits as Durables" Model," *International Economic Review*, 54, 717–738.

Devetag, G., S. Di Guida, and L. Polonio (forthcoming): "An Eye-Tracking Study of Feature-Based Choice in One-Shot Games," *Experimental Economics*.

Diecidue, E. and P. P. Wakker (2001): "On the Intuition of Rank-Dependent Utility," *Journal of Risk and Uncertainty*, 23, 281–298.

Direr, A. (2013): "Are Betting Markets Efficient? Evidence from European Football Championships," *Applied Economics*, 45, 343–356.

Donkers, B., B. Melenberg, and A. Van Soest (2001): "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach," *Journal of Risk and Uncertainty*, 22, 165–195.

Dowie, J. (1976): "On the Efficiency and Equity of Betting Markets," *Economica*, 43, 139–150.

Dreber, A., T. Pfeiffer, J. Almenberg, S. Isaksson, B. Wilson, Y. Chen, B. A. Nosek, and M. Johannesson (2015): "Using Prediction Markets to Estimate the Reproducibility of Scientific Research," *Proceedings of the National Academy of Sciences*, 112, 15343–15347.

Dziewulski, P. (2015): "Revealed Time-Preference," Unpublished manuscript.

Echenique, F., S. Lee, and M. Shum (2011): "The Money Pump as a Measure of Revealed Preference Violations," *Journal of Political Economy*, 119, 1201–1223.

Echenique, F. and K. Saito (2015): "Savage in the Market," *Econometrica*, 83, 1467–1495.

Eckel, C. C. and P. J. Grossman (2002): “Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk,” *Evolution and Human Behavior*, 23, 281–295.

——— (2008): “Differences in the Economic Decisions of Men and Women: Experimental Evidence,” in *Handbook of Experimental Economics Results*, ed. by C. Plott and V. Smith, New York: Elsevier, vol. 1, 509–519.

Efron, B. and R. Tibshirani (1993): *An Introduction to the Bootstrap*, London: Chapman and Hall.

Einav, L., A. Finkelstein, S. P. Ryan, P. Schrimpf, and M. R. Cullen (2013): “Selection on Moral Hazard in Health Insurance,” *American Economic Review*, 103, 178–219.

Einav, L. and J. Levin (2014): “Economics in the Age of Big Data,” *Science*, 346, 1243089.

Ely, J., A. Frankel, and E. Kamenica (2015): “Suspense and Surprise,” *Journal of Political Economy*, 123, 215–260.

Epstein, L. G. and S. E. Zin (1991): “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis,” *Journal of Political Economy*, 99, 263–286.

Etchart-Vincent, N. (2004): “Is Probability Weighting Sensitive to the Magnitude of Consequences? An Experimental Investigation on Losses,” *Journal of Risk and Uncertainty*, 28, 217–235.

Fama, E. F. (1965): “The Behavior of Stock-Market Prices,” *Journal of Business*, 38, 34–105.

——— (1970): “Efficient Capital Markets: A Review of Theory and Empirical Work,” *Journal of Finance*, 25, 383–417.

Fan, J. (1992): “Design-Adaptive Nonparametric Regression,” *Journal of the American Statistical Association*, 87, 998–1004.

——— (1993): “Local Linear Regression Smoothers and Their Minimax Efficiencies,” *Annals of Statistics*, 21, 196–216.

Fan, J. and I. Gijbels (1992): “Variable Bandwidth and Local Linear Regression Smoothers,” *Annals of Statistics*, 20, 2008–2036.

FEESS, E., H. MÜLLER, AND C. SCHUMACHER (2014): "Estimating Risk Preferences of Bettors with Different Bet Sizes," Unpublished manuscript.

FEHR-DUDA, H., A. BRUHIN, T. EPPER, AND R. SCHUBERT (2010): "Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size," *Journal of Risk and Uncertainty*, 40, 147–180.

FEHR-DUDA, H., M. DE GENNARO, AND R. SCHUBERT (2006): "Gender, Financial Risk, and Probability Weights," *Theory and Decision*, 60, 283–313.

FEHR-DUDA, H. AND T. EPPER (2012): "Probability and Risk: Foundations and Economic Implications of Probability-Dependent Risk Preferences," *Annual Review of Econonomics*, 4, 567–593.

FELDMANHALL, O., T. DALGLEISH, R. THOMPSON, D. EVANS, S. SCHWEIZER, AND D. MOBBS (2012a): "Differential Neural Circuitry and Self-Interest in Real vs Hypothetical Moral Decisions," *Social Cognitive and Affective Neuroscience*, 7, 743–751.

FELDMANHALL, O., D. MOBBS, D. EVANS, L. HISCOX, L. NAVRADY, AND T. DALGLEISH (2012b): "What We Say and What We Do: The Relationship between Real and Hypothetical Moral Choices," *Cognition*, 123, 434–441.

FISHBURN, P. AND W. EDWARDS (1997): "Discount-Neutral Utility Models for Denumerable Time Streams," *Theory and Decision*, 43, 139–166.

FISHBURN, P. C. AND A. RUBINSTEIN (1982): "Time Preference," *International Economic Review*, 23, 677–694.

FISMAN, R., S. KARIV, AND D. MARKOVITS (2007): "Individual Preferences for Giving," *American Economic Review*, 97, 1858–1876.

FLEISSIG, A. R. AND G. A. WHITNEY (2005): "Testing for the Significance of Violations of Afriat's Inequalities," *Journal of Business & Economic Statistics*, 23, 355–362.

FORSYTHE, R., F. NELSON, G. R. NEUMANN, AND J. WRIGHT (1992): "Anatomy of an Experimental Political Stock Market," *American Economic Review*, 82, 1142–1161.

FOX, J. A., J. F. SHOGREN, D. J. HAYES, AND J. B. KLIEBENSTEIN (1998): "CVM-X: Calibrating Contingent Values with Experimental Auction Markets," *American Journal of Agricultural Economics*, 80, 455–465.

Fuchs, A. F. (1971): "The Saccadic System," in *The Control of Eye Movements*, ed. by P. Bach-Y-Rita, C. C. Collins, and J. E. Hyde, London: Academic Press, 343–362.

Fullenkamp, C., R. Tenorio, and R. Battalio (2003): "Assessing Individual Risk Attitudes Using Field Data From Lottery Games," *Review of Economics and Statistics*, 85, 218–226.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006): "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 96, 1043–1068.

Galperti, S. and B. Strulovici (2014): "From Anticipations to Present Bias: A Theory of Forward-Looking Preferences," Unpublished manuscript.

Gandar, J., R. Zuber, T. O'Brien, and B. Russo (1988): "Testing Rationality in the Point Spread Betting Market," *Journal of Finance*, 43, 995–1008.

Gandar, J. M., R. A. Zuber, and R. S. Johnson (2001): "Searching for the Favourite-Longshot Bias Down Under: An Examination of the New Zealand Pari-mutuel Betting Market," *Applied Economics*, 33, 1621–1629.

Gandhi, A. and R. Serrano-Padial (2015): "Does Belief Heterogeneity Explain Asset Prices: The Case of the Longshot Bias," *Review of Economic Studies*, 82, 156–186.

Garber, A. M. and C. E. Phelps (1997): "Economic Foundations of Cost-Effectiveness Analysis," *Journal of Health Economics*, 16, 1–31.

Geisser, S. (1975): "The Predictive Sample Reuse Method with Applications," *Journal of the American Statistical Association*, 70, 320–328.

Gertner, R. (1993): "Game Shows and Economic Behavior: Risk-Taking on 'Card Sharks'," *Quarterly Journal of Economics*, 108, 507–521.

Gilboa, I. (1989): "Expectation and Variation in Multi-Period Decisions," *Econometrica*, 57, 1153–1169.

Gillen, B. J., C. R. Plott, and M. Shum (2015): "A Pari-mutuel like Mechanism for Information Aggregation: A Field Test Inside Intel," Unpublished manuscript.

Giné, X., J. Goldberg, D. Silverman, and D. Yang (2013): "Revising Commitments: Time Preference and Time-Inconsistency in the Field," Unpublished manuscript.

Gjerstad, S. (2005): "Risk Aversion, Beliefs, and Prediction Market Equilibrium," Unpublished manuscript.

Gold, M. R., J. E. Seigel, L. B. Russell, and M. C. Weinstein, eds. (1996): *Cost-Effectiveness in Health and Medicine*, Oxford: Oxford University Press.

Goldberg, J. H., M. J. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky (2002): "Eye Tracking in Web Search Tasks: Design Implications," in *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ACM, 51–58.

Goldstein, W. M. and H. J. Einhorn (1987): "Expression Theory and the Preference Reversal Phenomena." *Psychological Review*, 94, 236–254.

Golec, J. and M. Tamarkin (1998): "Bettors Love Skewness, Not Risk, at the Horse Track," *Journal of Political Economy*, 106, 205–225.

Gonzalez, R. and G. Wu (1999): "On the Shape of the Probability Weighting Function," *Cognitive Psychology*, 38, 129–166.

Gramm, M. and D. H. Owens (2005): "Determinants of Betting Market Efficiency," *Applied Economics Letters*, 12, 181–185.

Green, P. E. and V. Srinivasan (1990): "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *Journal of Marketing*, 54, 3–19.

Greene, J. D., L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen (2004): "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron*, 44, 389–400.

Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen (2001): "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science*, 293, 2105–2108.

Griffith, R. M. (1949): "Odds Adjustments by American Horse-Race Bettors," *American Journal of Psychology*, 62, 290–294.

Handel, B. R. (2013): "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103, 2643–2682.

HANNULA, D. E. AND C. RANGANATH (2009): "The Eyes Have It: Hippocampal Activity Predicts Expression of Memory in Eye Movements," *Neuron*, 63, 592–599.

HÄRDLE, W. (1990): *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.

HÄRDLE, W. AND A. W. BOWMAN (1988): "Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands," *Journal of the American Statistical Association*, 83, 102–110.

HÄRDLE, W. AND J. S. MARRON (1991): "Bootstrap Simultaneous Error Bars for Nonparametric Regression," *Annals of Statistics*, 19, 778–796.

HARIRI, A. R., S. M. BROWN, D. E. WILLIAMSON, J. D. FLORY, H. DE WIT, AND S. B. MANUCK (2006): "Preference for Immediate over Delayed Rewards is Associated with Magnitude of Ventral Striatal Activity," *Journal of Neuroscience*, 26, 13213–13217.

HARLESS, D. W. AND C. F. CAMERER (1994): "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62, 1251–1289.

HARRISON, G. W. AND E. E. RUTSTRÖM (2008): "Risk Aversion in the Laboratory," in *Risk Aversion in Experiments*, ed. by J. Cox and G. W. Harrison, Bingley: Emerald, vol. 12 of *Research in Experimental Economics*, 41–196.

——— (2009): "Expected Utility Theory and Prospect Theory: One Wedding and a Decent Funeral," *Experimental Economics*, 12, 133–158.

HARTLEY, R., G. LANOT, AND I. WALKER (2014): "Who Really Wants to be a Millionaire? Estimates of Risk Aversion from Gameshow Data," *Journal of Applied Econometrics*, 29, 861–879.

HARTZMARK, S. M. AND D. H. SOLOMON (2012): "Efficiency and the Disposition Effect in NFL Prediction Markets," *Quarterly Journal of Finance*, 2, 1250013.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The Elements of Statistical Learning*, New York: Springer.

HAUSCH, D. B., W. T. ZIEMBA, AND M. RUBINSTEIN (1981): "Efficiency of the Market for Racetrack Betting," *Management Science*, 27, 1435–1452.

Hausman, J. A., J. Abrevaya, and F. M. Scott-Morton (1998): "Misclassification of the Dependent Variable in a Discrete-Response Setting," *Journal of Econometrics*, 87, 239–269.

Hayashi, T. (2003): "Quasi-Stationary Cardinal Utility and Present Bias," *Journal of Economic Theory*, 112, 343–352.

Henery, R. J. (1985): "On the Average Probability of Losing Bets on Horses with Given Starting Price Odds," *Journal of the Royal Statistical Society. Series A (General)*, 148, 342–349.

Hey, J. D. (2014): "Choice Under Uncertainty: Empirical Methods and Experimental Results," in *Handbook of the Economics of Risk and Uncertainty*, ed. by M. J. Machina and W. K. Viscusi, Amsterdam: North-Holland, vol. 1, 809–850.

Hey, J. D. and C. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326.

Holt, C. A. and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Hopkins, D. J. (2009): "No More Wilder Effect, Never a Whitman Effect: When and Why Polls Mislead about Black and Female Candidates," *Journal of Politics*, 71, 769–781.

Houtman, M. and J. A. H. Maks (1985): "Determining All Maximal Data Subsets Consistent with Revealed Preference," *Kwantitatieve Methoden*, 19, 89–104.

Hsu, M., I. Krajbich, C. Zhao, and C. F. Camerer (2009): "Neural Response to Reward Anticipation under Risk is Nonlinear in Probabilities," *Journal of Neuroscience*, 29, 2231–2237.

Infosino, W. J. (1986): "Forecasting New Product Sales from Likelihood of Purchase Ratings," *Marketing Science*, 5, 372–384.

Jacquemet, N., R.-V. Joule, S. Luchini, and J. F. Shogren (2013): "Preference Elicitation under Oath," *Journal of Environmental Economics and Management*, 65, 110–132.

Jamieson, L. F. and F. M. Bass (1989): "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods," *Journal of Marketing Research*, 26, 336–345.

Janssens, W., B. Kramer, and L. Swart (2013): "Time Inconsistent Behavior Under Incomplete Markets," Unpublished manuscript.

Johannesson, M., B. Liljas, and P.-O. Johansson (1998): "An Experimental Comparison of Dichotomous Choice Contingent Valuation Questions and Real Purchase Decisions," *Applied Economics*, 30, 643–647.

Johnson, E. J., C. F. Camerer, S. Sen, and T. Rymon (2002): "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory*, 104, 16–47.

Jullien, B. and B. Salanié (2000): "Estimating Preferences under Risk: The Case of Racetrack Bettors," *Journal of Political Economy*, 108, 503–530.

——— (2008): "Empirical Evidence on the Preferences of Racetrack Bettors," in *Handbook of Sports and Lottery Markets*, ed. by W. T. Ziemba and D. B. Hausch, Amsterdam: North-Holland, 83–101.

Just, M. A. and P. A. Carpenter (1980): "A Theory of Reading: From Eye Fixations to Comprehension," *Psychological Review*, 87, 329–354.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

Kang, M. J. and C. F. Camerer (2013): "fMRI Evidence of a Hot-Cold Empathy Gap in Hypothetical and Real Aversive Choices," *Frontiers in Neuroscience*, 7.

Kang, M. J., A. Rangel, M. Camus, and C. F. Camerer (2011): "Hypothetical and Real Choice Differentially Activate Common Valuation Areas," *Journal of Neuroscience*, 31, 461–468.

Kang, M. J., D. Ray, and C. Camerer (2012): "Measured Anxiety and Choices in Experimental Timing Games," Unpublished manuscript.

Karmarkar, U. S. (1978): "Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model," *Organizational Behavior and Human Performance*, 21, 61–72.

——— (1979): "Subjectively Weighted Utility and the Allais Paradox," *Organizational Behavior and Human Performance*, 24, 67–72.

KEETER, S. AND N. SAMARANAYAKE (2007): "Can You Trust What Polls Say about Obama's Electoral Prospects?" *Pew Research Center Report, available online at: http://pewresearch.org*.

KELDENICH, K. AND M. KLEMM (2014): "Double or Nothing?! Small Groups Making Decisions under Risk in "Quiz Taxi"," *Theory and Decision*, 77, 243–274.

KELLEY, M. R. AND R. J. LEMKE (2015): "Gender Differences When Subjective Probabilities Affect Risky Decisions: An Analysis from the Television Game Show *Cash Cab*," *Theory and Decision*, 78, 153–170.

KLEINER, M., D. BRAINARD, D. PELLI, A. INGLING, R. MURRAY, AND C. BROUSSARD (2007): "What's New in Psychtoolbox-3," *Perception*, 36, ECVP Abstract Supplement.

KNOEPFLE, D. T., J. T.-Y. WANG, AND C. F. CAMERER (2009): "Studying Learning in Games Using Eye-tracking," *Journal of the European Economic Association*, 7, 388–398.

KOOPMANS, T. C. (1960): "Stationary Ordinal Utility and Impatience," *Econometrica*, 28, 287–309.

KOPYLOV, I. (2010): "Simple Axioms for Countably Additive Subjective Probability," *Journal of Mathematical Economics*, 46, 867–876.

KRAJBICH, I., C. CAMERER, J. LEDYARD, AND A. RANGEL (2009): "Using Neural Measures of Economic Value to Solve the Public Goods Free-Rider Problem," *Science*, 326, 596–599.

KÜHBERGER, A., M. SCHULTE-MECKLENBECK, AND J. PERNER (2002): "Framing Decisions: Hypothetical and Real," *Organizational Behavior and Human Decision Processes*, 89, 1162–1175.

KUHN, M., P. KUHN, AND M. C. VILLEVAL (2014): "Self Control and Intertemporal Choice: Evidence from Glucose and Depletion Interventions," Unpublished manuscript.

KURZ, M. (1974): "Experimental Approach to the Determination of the Demand for Public Goods," *Journal of Public Economics*, 3, 329–348.

LACETERA, N., D. POPE, AND J. SYDNOR (2012): "Heuristic Thinking and Limited Attention in the Car Market," *American Economic Review*, 102, 2206–2236.

LAHVIČKA, J. (2014): "What Causes the Favourite-Longshot Bias? Further Evidence from Tennis," *Applied Economics Letters*, 21, 90–92.

LAIBSON, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–478.

LATTIMORE, P. K., J. R. BAKER, AND A. D. WITTE (1992): "The Influence of Probability on Risky Choice: A Parametric Examination," *Journal of Economic Behavior & Organization*, 17, 377–400.

LEVY, I., S. C. LAZZARO, R. B. RUTLEDGE, AND P. W. GLIMCHER (2011): "Choice from Non-Choice: Predicting Consumer Preferences from Blood Oxygenation Level-Dependent Signals Obtained during Passive Viewing," *Journal of Neuroscience*, 31, 118–125.

LICHTENSTEIN, S., B. FISCHHOFF, AND L. D. PHILLIPS (1977): "Calibration of Probabilities: The State of the Art," in *Decision Making and Change in Human Affairs*, ed. by H. Jungermann and G. De Zeeuw, Netherlands: Springer, 275–324.

LINDQUIST, G. S. AND J. SÄVE-SÖDERBERGH (2011): ""Girls will be Girls", especially among Boys: Risk-taking in the "Daily Double" on *Jeopardy*," *Economics Letters*, 112, 158–160.

LIST, J. A. AND C. A. GALLET (2001): "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics*, 20, 241–254.

LIST, J. A. AND J. F. SHOGREN (1998): "Calibration of the Difference between Actual and Hypothetical Valuations in a Field Experiment," *Journal of Economic Behavior & Organization*, 37, 193–205.

——— (2002): "Calibration of Willingness-to-Accept," *Journal of Environmental Economics and Management*, 43, 219–233.

LITTLE, J. AND R. BERRENS (2004): "Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis," *Economics Bulletin*, 3, 1–13.

LIU, E. M., J. MENG, AND J. T.-Y. WANG (2014): "Confucianism and Preferences: Evidence from Lab Experiments in Taiwan and China," *Journal of Economic Behavior & Organization*, 104, 106–122.

LOHSE, G. L. AND E. J. JOHNSON (1996): "A Comparison of Two Process Tracing Methods for Choice Tasks," *Organizational Behavior and Human Decision Processes*, 68, 28–43.

LOOMIS, J., A. GONZALEZ-CABAN, AND R. GREGORY (1994): "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" *Land Economics*, 70, 499–506.

LOOMIS, J., K. TRAYNOR, AND T. BROWN (1999): "Trichotomous Choice: A Possible Solution to Dual Response Objectives in Dichotomous Choice Contingent Valuation Questions," *Journal of Agricultural and Resource Economics*, 24, 572–583.

LÜHRMANN, M., M. SERRA-GARCIA, AND J. WINTER (2014): "The Impact of Financial Education on Adolescents' Intertemporal Choices," Unpublished manuscript.

MANSKI, C. F. (2006): "Interpreting the Predictions of Prediction Markets," *Economics Letters*, 91, 425–429.

MCGLOTHLIN, W. H. (1956): "Stability of Choices among Uncertain Alternatives," *American Journal of Psychology*, 69, 604–615.

METRICK, A. (1995): "A Natural Experiment in "Jeopardy!"," *American Economic Review*, 85, 240–253.

MONTEROSSO, J. R., G. AINSLIE, J. XU, X. CORDOVA, C. P. DOMIER, AND E. D. LONDON (2007): "Frontoparietal Cortical Activity of Methamphetamine-Dependent and Comparison Subjects Performing a Delay Discounting Task," *Human Brain Mapping*, 28, 383–393.

MONTIEL OLEA, J. L. AND T. STRZALECKI (2014): "Axiomatization and Measurement of Quasi-Hyperbolic Discounting," *Quarterly Journal of Economics*, 129, 1449–1499.

MULINO, D., R. SCHEELINGS, R. BROOKS, AND R. FAFF (2009): "Does Risk Aversion Vary with Decision-Frame? An Empirical Test Using Recent Game Show Data," *Review of Behavioral Finance*, 1, 44–61.

MURPHY, J. J., P. G. ALLEN, T. H. STEVENS, AND D. WEATHERHEAD (2005): "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation," *Environmental and Resource Economics*, 30, 313–325.

Nalebuff, B. (1990): "Puzzles: Slot Machines, Zomepirac, Squash, and More," *Journal of Economic Perspectives*, 4, 179–187.

Neill, H. R. (1995): "The Context for Substitutes in CVM Studies: Some Empirical Observations," *Journal of Environmental Economics and Management*, 29, 393–397.

Odean, T. (1998): "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53, 1775–1798.

Ottaviani, M. and P. N. Sørensen (2008): "The Favorite-Longshot Bias: An Overview of the Main Explanations," in *Handbook of Sports and Lottery Markets*, ed. by W. T. Ziemba and D. B. Hausch, Amsterdam: North-Holland, 83–101.

——— (2009): "Surprised by the Parimutuel Odds?" *American Economic Review*, 99, 2129–2134.

——— (2010): "Noise, Information, and the Favorite-Longshot Bias in Parimutuel Predictions," *American Economic Journal: Microeconomics*, 2, 58–85.

——— (2015): "Price Reaction to Information with Heterogeneous Beliefs and Wealth Effects: Underreaction, Momentum, and Reversal," *American Economic Review*, 105, 1–34.

Page, L. and R. T. Clemen (2013): "Do Prediction Markets Produce Well-Calibrated Probability Forecasts?" *Economic Journal*, 123, 491–513.

Pelli, D. G. (1997): "The VideoToolbox Software for Visual Psychophysics: Transforming Numbers into Movies," *Spatial Vision*, 10, 437–442.

Phelps, E. S. and R. A. Pollak (1968): "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, 35, 185–199.

Polkovnichenko, V. and F. Zhao (2013): "Probability Weighting Functions Implied in Options Prices," *Journal of Financial Economics*, 107, 580–609.

Polonio, L., S. Di Guida, and G. Coricelli (2015): "Strategic Sophistication and Attention in Games: An Eye-Tracking Study," *Games and Economic Behavior*, 94, 80–96.

Post, T., M. J. van den Assem, G. Baltussen, and R. H. Thaler (2008): "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review*, 98, 38–71.

PRELEC, D. (1998): "The Probability Weighting Function," *Econometrica*, 66, 497–528.

QUANDT, R. E. (1986): "Betting and Equilibrium," *Quarterly Journal of Economics*, 101, 201–208.

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, 3, 323–343.

RAGHUBIR, P. AND E. A. GREENLEAF (2006): "Ratios in Proportion: What should the Shape of the Package Be?" *Journal of Marketing*, 70, 95–107.

REUTSKAJA, E., R. NAGEL, C. F. CAMERER, AND A. RANGEL (2011): "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study," *American Economic Review*, 101, 900–926.

ROCKAFELLAR, R. T. (1970): *Convex Analysis*, Princeton: Princeton University Press.

ROTTENSTREICH, Y. AND C. K. HSEE (2001): "Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk," *Psychological Science*, 12, 185–190.

SAUER, R. (1998): "The Economics of Wagering Markets," *Journal of Economic Literature*, 36, 2021–2064.

SAWADA, Y. AND Y. KUROISHI (2015): "How Does a Natural Disaster Affect People's Preferencec? An Investigation of a Large Scale Flood in the Philippines using Convex Time Budget Experiments," Unpublished manuscript.

SCHLOSSER, A. E., T. B. WHITE, AND S. M. LLOYD (2006): "Converting Web Site Visitors into Buyers: How Web Site Investment Increases Consumer Trusting Beliefs and Online Purchase Intentions," *Journal of Marketing*, 70, 133–148.

SHAW, W. D., T. GRIJALVA, J. L. LUSK, AND R. RONG (2014): "Does the Convex Time Budget Approach Lead to Evidence of Hyperbolic Discounting when the Time Horizon is Very Long?" Unpublished manuscript.

SHEFRIN, H. AND M. STATMAN (1985): "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence," *Journal of Finance*, 40, 777–790.

SHIMOJO, S., C. SIMION, E. SHIMOJO, AND C. SCHEIER (2003): "Gaze Bias both Reflects and Influences Preference," *Nature Neuroscience*, 6, 1317–1322.

Shin, H. S. (1991): "Optimal Betting Odds against Insider Traders," *Economic Journal*, 101, 1179–1185.

——— (1992): "Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias," *Economic Journal*, 102, 426–435.

Shogren, J. F. (1993): "Experimental Markets and Environmental Policy," *Agricultural and Resource Economics Review*, 22, 117–129.

——— (2005): "Experimental Methods and Valuation," in *Handbook of Environmental Economics*, ed. by K.-G. Mäler and J. Vincent, Amsterdam: North-Holland, vol. 2, 969–1027.

——— (2006): "Valuation in the Lab," *Environmental and Resource Economics*, 34, 163–172.

Silk, A. J. and G. L. Urban (1978): "Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," *Journal of Marketing Research*, 15, 171–191.

Smith, A., B. D. Bernheim, C. F. Camerer, and A. Rangel (2014): "Neural Activity Reveals Preferences without Choices," *American Economic Journal: Microeconomics*, 6, 1–36.

Snowberg, E. and J. Wolfers (2010): "Explaining the Favorite-Longshot Bias: Is it Risk-Love or Misperceptions?" *Journal of Political Economy*, 118, 723–746.

Snowberg, E., J. Wolfers, and E. Zitzewitz (2011): "How Prediction Markets can Save Event Studies," in *Prediction Markets*, ed. by L. Vaughn Williams, London: Routledge, 18–34.

——— (2013): "Prediction Markets for Economic Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, New York: Elsevier, vol. 2, 657–687.

Snyder, W. W. (1978): "Horse Racing: Testing the Efficient Markets Model," *Journal of Finance*, 33, 1109–1118.

Sonnemann, U., C. F. Camerer, C. R. Fox, and T. Langer (2013): "How Psychological Framing Affects Economic Market Prices in the Lab and Field," *Proceedings of the National Academy of Sciences*, 110, 11779–11784.

Starmer, C. (2000): "Developments in Non-expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature*, 38, 332–382.

Stewart, N., S. Gächter, T. Noguchi, and T. L. Mullett (2015): "Eye Movements in Strategic Choice," Unpublished manuscript.

Stoer, J. and C. Witzgall (1970): *Convexity and Optimization in Finite Dimensions*, Berlin: Springer-Verlag.

Stone, C. J. (1977): "Consistent Nonparametric Regression," *Annals of Statistics*, 5, 595–620.

Stone, M. (1974): "Cross-Validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society. Series B (Methodological)*, 36, 111–147.

Stott, H. P. (2006): "Cumulative Prospect Theory's Functional Menagerie," *Journal of Risk and Uncertainty*, 32, 101–130.

Takahashi, H., H. Matsui, C. Camerer, H. Takano, F. Kodaka, T. Ideno, S. Okubo, K. Takemura, R. Arakawa, Y. Eguchi, et al. (2010): "Dopamine D1 Receptors and Nonlinear Probability Weighting in Risky Choice," *Journal of Neuroscience*, 30, 16567–16572.

Tanaka, T., C. F. Camerer, and Q. Nguyen (2010): "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam," *American Economic Review*, 100, 557–571.

Tetlock, P. C. (2008): "Liquidity and Prediction Market Efficiency," Unpublished manuscript.

Thaler, R. H. and E. J. Johnson (1990): "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice," *Management Science*, 36, 643–660.

Thaler, R. H. and W. T. Ziemba (1988): "Parimutuel Betting Markets: Racetracks and Lotteries," *Journal of Economic Perspectives*, 2, 161–174.

Tompkins, R. G., W. T. Ziemba, and S. D. Hodges (2008): "The Favorite-Longshot Bias in S&P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance," in *Handbook of Sports and Lottery Markets*, ed. by W. T. Ziemba and D. B. Hausch, Amsterdam: North-Holland, 161–178.

Trivers, R. (1972): "Parental Investment and Sexual Selection," in *Sexual Selection and the Descent of Man*, ed. by B. Campbell, New York: Aldine, 136–179.

Tusche, A., S. Bode, and J.-D. Haynes (2010): "Neural Responses to Unattended Products Predict Later Consumer Choices," *Journal of Neuroscience*, 30, 8024–8031.

Tversky, A. and C. R. Fox (1995): "Weighing Risk and Uncertainty." *Psychological Review*, 102, 269–283.

Tversky, A. and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Urban, G. L., G. M. Katz, T. E. Hatch, and A. J. Silk (1983): "The ASSESSOR Pre-Test Market Evaluation System," *Interfaces*, 13, 38–59.

van de Kuilen, G. and P. P. Wakker (2011): "The Midweight Method to Measure Attitudes Toward Risk and Ambiguity," *Management Science*, 57, 582–598.

Varian, H. R. (1985): "Non-Parametric Analysis of Optimizing Behavior with Measurement Error," *Journal of Econometrics*, 30, 445–458.

——— (2014): "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 28, 3–28.

Vaughan Williams, L. and D. Paton (1997): "Why Is There a Favourite-Longshot Bias in British Racetrack Betting Markets?" *Economic Journal*, 107, 150–158.

von Gaudecker, H.-M., A. van Soest, and E. Wengstrom (2011): "Heterogeneity in Risky Choice Behavior in a Broad Population," *American Economic Review*, 101, 664–694.

Wakker, P. and A. Tversky (1993): "An Axiomatization of Cumulative Prospect Theory," *Journal of Risk and Uncertainty*, 7, 147–175.

Wakker, P. P. (2010): *Prospect Theory: For Risk and Ambiguity*, Cambridge: Cambridge University Press.

Walls, W. D. and K. Busche (2003): "Broken Odds and the Favourite-Longshot Bias in Parimutuel Betting: A Direct Test," *Applied Economics Letters*, 10, 311–314.

Wang, J. T.-Y., M. Spezio, and C. F. Camerer (2010): "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games," *American Economic Review*, 100, 984–1007.

Weibull, J. W. (1985): "Discounted-Value Representations of Temporal Preferences," *Mathematics of Operations Research*, 10, 244–250.

Weitzman, M. (1965): "Utility Analysis and Group Behavior: An Empirical Study," *Journal of Political Economy*, 73, 18–26.

Wilson, M. and M. Daly (1985): "Competitiveness, Risk Taking, and Violence: The Young Male Syndrome," *Ethology and Sociobiology*, 6, 59–73.

Wolfers, J. and E. Zitzewitz (2004): "Prediction Markets," *Journal of Economic Perspectives*, 18, 107–126.

——— (2007): "Interpreting Prediction Market Prices as Probabilities," Unpublished manuscript.

Woodland, L. M. and B. M. Woodland (1994): "Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market," *Journal of Finance*, 49, 269–279.

——— (2001): "Market Efficiency and Profitable Wagering in the National Hockey League: Can Bettors Score on Longshots?" *Southern Economic Journal*, 67, 983–995.

——— (2003): "The Reverse Favourite–longshot Bias and Market Efficiency in Major League Baseball: An Update," *Bulletin of Economic Research*, 55, 113–123.

Wooldridge, J. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Wu, G. and R. Gonzalez (1996): "Curvature of the Probability Weighting Function," *Management Science*, 42, 1676–1690.

Wu, G., J. Zhang, and R. Gonzalez (2004): "Decision Under Risk," in *Blackwell Handbook of Judgment and Decision Making*, ed. by D. J. Koehler and N. Harvey, Oxford: Oxford University Press, 399–423.

Wulfert, E., B. D. Roland, J. Hartley, N. Wang, and C. Franco (2005): "Heart Rate Arousal and Excitement in Gambling: Winners Versus Losers." *Psychology of Addictive Behaviors*, 19, 311–316.

Zeckhauser, R. and D. Shepard (1976): "Where Now for Saving Lives," *Law & Contemporary Problems*, 40, 5–45.

Zhang, H. and L. T. Maloney (2012): "Ubiquitous Log Odds: A Common Representation of Probability and Frequency Distortion in Perception, Action, and Cognition," *Frontiers in Neuroscience*, 6.

Zuber, R. A., J. M. Gandar, and B. D. Bowers (1985): "Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games," *Journal of Political Economy*, 93, 800–806.