Information Markets and Aggregation

by

Narahari Mohan Phatak

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Christine Parlour, Chair
Professor William Fuchs
Professor Terrence Hendershott
Professor Ulrike Malmendier

Spring 2012

Information Markets and Aggregation

Abstract

Information Markets and Aggregation

by

Narahari Mohan Phatak

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Christine Parlour, Chair

Markets serve a price discovery function. In commodity markets, this supports efficient trade between agents with different preferences and endowments. Prices equilibrate so that aggregate supply of a good matches aggregate demand for that good. In financial markets, agents form demands on the basis of both preferences and information about cash flows and discount rates. Here, prices take on additional significance. In the chapters that follow, I explore how markets aggregate private information in experimental and natural settings. Each chapter considers information and aggregation from a different perspective.

Chapter one considers bounded rationality and information aggregation in an experimental forecasting game. I examine the effects of different types of complexity on decision making and information transmission. I focus on two types of complexity: the number of alternatives and the organization of information about risks and rewards. Each of these can prevent subjects from making appropriate choices. My results suggest that simplification of financial decisions, within limits, may improve information transmission while helping individuals make better choices. Though individuals make poor choices in very complex environments, constraining their choices too much also makes it difficult for them to choose well. My experimental setup also enables me to construct forecasts by aggregating information from agents' choices. I show how these complex choice environments can lead to inefficient forecasts.

Chapter two explores the use of markets to elicit and aggregate information in places where these mechanisms do not normally exist. I assemble a novel data set gathered from a corporate prediction market in which managers at a software firm allowed employees to place bets on key variables. I use these data to examine the static and dynamic properties of information within the firm. I find that employees are privately informed about project outcomes. However, information is not evenly distributed across the firm - some groups appear to know more than other about products and sales. To examine the flow of information within the firm, I focus on a subset of bets that were later revised by employees. Revised bets preform well relative to bets pre-revision, suggesting that employees acquire private information over

time. Again, the quality of information flow appears related to job function. This study lends insight into the source of managers' private information in a corporate finance setting.

Finally, in chapter three, I focus on the motivations of market participants. I examine a betting market which experienced an exogenous shock to incentives. I use data on bets placed by participants to assess whether tournament prizes elicit exaggerated forecasts. I also study whether changing the value of monetary incentives has bearing on the participants' willingness to make forecasts. I find that, in line with predictions, reducing the proportion of prize winners, appears to increase the riskiness of their bets with no measurable increase in information content. However, contrary to expectations, I do not find that smaller prize values lead to lower participation. This study complements the theoretical literature on forecasting contests that suggests professional analysts have incentives to exaggerate their claims.

To my parents, Mohan and Rekha Phatak

I have so much to live up to.

Contents

List of Figures

List of Tables

Acknowledgments

I owe the successful completion of this dissertation to a very long list of people. Terry Hendershott and Willie Fuchs shepherded me through my oral exams and through the dissertation-writing process. Ulrike Malmendier provided excellent feedback and offered plenty of opportunities to present my work in her student seminars. I also appreciate suggestions from Terry Odean and Stefano DellaVigna on chapter one of my dissertation. My committee chair, Christine Parlour, deserves special thanks for her constant support, patience, good humor, and for seeing me through many false starts in research.

I also thank a number of my classmates at the Haas School of Business. James McLoughlin, Boris Albul, Sean Wilkoff, Bradyn Breon-Drish, Javed Ahmed, Andres Donangelo, Barney Hartman-Glaser, Nishanth Rajan, Sara Holland, Adam Yonce, Pratish Patel, Orie Shelef, Sharon Poczter and Jenny Chu provided advice, lent support, and made graduate school enjoyable.

I spent the summer of 2009 getting practical experience in prediction market design with Crowdcast. This work motivated much of my interest in information markets. For this, I owe thanks to Buckley White and Leslie Fine. Moreover, none of this work would have been possible without technical support from Jay Margolis and Danny Olson.

Chapter 1

Menu-Based Complexity: Experiments on Choice Over Lotteries

1.1 Introduction

Policy makers who want to improve the financial decision making of people with varying degrees of financial literacy often cite simplicity as a goal. Financial decisions, they claim, are too complicated. The proliferation of instruments, features and rules discourage participation by less-sophisticated agents and can cause even sophisticated players to make unwise choices. Yet, a few key points are conspicuously absent from the discussion. What makes financial decisions complex? What distortions result from complicated decisions? Do distortions at the individual level appear as distortions in aggregates?

I design a series of experiments to investigate how the complexity of financial decisions distorts choices and makes it difficult to make inferences from agents' behavior. I invite workers on Amazon Mechanical Turk to play a forecasting game that requires them to bet on a random outcome. Bets take the form of lotteries in which subjects receive a fixed payoff if a random variable lands within a given range of values. The task I require of subjects bears resemblance to choosing between alternative financial investments. Subjects must examine the universe of available alternatives that pay off in different states of the world. For each alternative, they must attach the probability that a state is realized to compute expected payoffs. I manipulate subjects' decision problems to investigate two types of complexity:

1. Alternative-based complexity: Do subjects make poorer decisions when confronted with more options?

2. Disordered alternatives: Will better organization of alternatives result in better decisions?

I find that subjects respond to increasing complexity in both of the forms listed above. More complicated decision environments introduce errors in decision making. However, not all reductions in complexity improve decisions. The evidence I present suggests a more nuanced view of the effects of simplifying choice problems. Confronted with too many alternatives, subjects make poor choices but too few alternatives can also lead subjects to choose poorly.

Studies in marketing acknowledge that alternative-based complexity may result in less-efficient choices. A number of papers examine the decisions of agents confronted with menus of different lengths. Iyengar and Lepper (2000) show that consumers are less likely to buy products when confronted with a dizzying array of choices. Multiple studies have document a "choice overload" effect, where individuals confronted with many alternatives tend to defer choice or feel less satisfaction with their choices. Iyengar and Kamenica (2010) observe, in the context of lotteries, that increasing the number of alternatives drives subjects to select lotteries with simpler payoff profiles.

Marketing research also describes results standing in opposition to the choice-overload hypothesis. Agents with very exacting tastes are less likely to encounter satisfying products from shorter menus than from larger ones. Chernev (2003) examines the behavior of subjects allowed to choose chocolates from assortments of differing size and then permitted to exchange their selections for a default bundle. Using the propensity of exchange as a measure of satisfaction, he finds that subjects with clear preferences over chocolate are less satisfied with selections from long menus than from shorter menus, relative to their less-discriminating counterparts. Less clear from the literature is whether agents in extremely simple environments continue to make efficient choices, conditional on the menus they observe.

Finance papers on portfolio choice have also taken a marketing approach to alternative-based complexity. Huberman, Iyengar and Jiang (2004) find that participation rates in 401(k) plans are sensitive to the number of funds offered to employees. They find the highest participation rate (75%) obtains in a plan of two funds. Increasing the number of alternatives to 59 funds, reduces participation to 60% of eligible employees.

My study deals less with measures of participation and satisfaction, focusing instead on how well agents optimize when menus change. Benartzi and Thaler (2001) pursue a similar goal when they present survey evidence indicating that the assortment of funds offered to subjects strongly influences asset allocations. In their study, participants who observed menus composed mostly of stock funds allocated significantly more to equities than counterparts who observed menus dominated by bond funds.

To this, I add an additional point: removing alternatives can also influence the likelihood with which agents choose optimally. Traditional theories of utility maximization suggest that agents are less likely to be satisfied with choices made from a constrained set of alternatives. The data I collect indicates that agents who choose from too limited an assortment of lotteries may also choose alternatives that are dominated by others available to them.

Similar to the strong effects of ordering found in online listings (Malmendier and Lee (2011)) and the below-the-fold effect noted in newspapers, I also ask whether the placement of alternatives within a menu has any bearing on the efficiency of choices made by subjects. Specifically, I test how subjects respond to menus that are disorganized. I demonstrate that simple reorganization of menus has consequences for behavior. Policies that promote clear presentation of relevant information about products can help consumers decide between alternatives.

In a similar vein, several recent papers in finance have examined menus and fee schedules in financial markets. Carlin (2009) and Carlin and Manso (2011) model obfuscation in financial product offerings. The former paper considers complexity as the extent to which firms control the ability of agents to optimize. In the latter, suppliers of financial products make it difficult for consumers to choose by periodically refreshing menus, forcing buyers to re-learn about financial products.

In product markets and financial markets alike, consumer choice communicates information about preferences to firms. Financial choices differ from other decisions in a couple of key respects. Besides communicating preferences, financial choices can also convey private information about future payoffs. Further, markets aggregate the information conveyed by financial decisions in prices that are inputs to production and policy decisions across the economy. This study considers ways of extending the choice literature which focuses on products to an environment where the information content of choices has broader impact.

Without anything to infer from purchase decisions, earlier studies did not consider how complexity might impair information transmission. I construct my experiments to enable me to examine how the proliferation of choice prevents agents from faithfully communicating their private signals. Through this channel, too much choice imposes costs on all market participants.

I consider the problem of a principal who observes the choices made by a group of agents over lotteries. He uses these data to form beliefs about the population of agents and uses these beliefs to construct lotteries for agents in subsequent rounds. This process of forecasting might describe a firm trying to choose the variety of products to offer consumers or the problem of an employer trying to construct a retirement savings plan for employees.

I assume from the start that information revelation enhances welfare. I take a view going back to Hayek (1945), echoed by Allen (1995) and Bond and Goldstein (2009): the price system collects private information from market participants and communicates it to decision makers. Implicitly, the market provides incentives for acquisition and transmission of information relevant to production decisions and the net benefits to revelation are positive.[*]

An existing strand of finance research considers how asset prices respond to investor psychology. Hirshleifer (2001) surveys the ways in which behavioral biases impact asset prices. Odean (1998) shows how overconfident traders can generate under-reaction of prices to information from rational traders. To the extent that inefficient choice impairs or biases information revelation in markets, the costs of complexity extend beyond its effect on those faced with difficult financial decisions.

Recent work in finance has also begun to analyze the externalities associated with complexity and simplification. Carlin, Gervais and Manso (2010) show how default options impose an externality on the production of information. In their economy, information is valuable to agents, costly to produce, and easily shared. Availability of a default option

[*]Hirshleifer (1971) considers cases where information collection can be inefficient.

reduces individuals agents' incentives to learn, reducing information sharing and, ultimately, social welfare. Carlin and Kogan (2010) present experimental evidence detailing how making securities complex changes bidding strategies, reduces liquidity and decreases trade efficiency. I add to these papers by looking at how inefficient choice results in incomplete transmission of private information.

For each experimental treatment, I begin by analyzing the behavior of subjects to expose the impact of different types of complexity on decision making. Next, I consider the implications for aggregation and quantify the effect of different types of complexity on forecast efficiency. Section 1.2 lays out the general experimental conditions common to all trials I conducted. Section 1.3 describes the different experiments and discusses the data obtained from them. Section 1.4 looks at the robustness of results while Section 1.5 concludes. The appendices contain further details on the experimental setting and calibration exercises mentioned in the paper.

1.2 Experimental Conditions

I conducted my tests using Amazon Mechanical Turk (MTurk). I supply a description of MTurk in Appendix A.2. I chose this tool over a traditional laboratory for a number of reasons. Foremost among these, MTurk allows for fast development and implementation of asynchronous experiments. MTurk also allows me to modify conditions and re-deploy without having to formally schedule trials.

The MTurk subject pool is diverse. Ross, et al (2010) aggregate data from multiple surveys conducted in 2008 and 2009. As of November 2009, they find 56% of the population comes from the United States and 36% comes from India, with the Indian share growing through time. Females constitute 52% of the worker base. Buhrmester, Kwang and Gosling (2011) find that a higher percentage (36%) of MTurk workers self-reported as non-white, relative to a separate sample of Internet users.

Mason and Suri (2011) find the population has an average age of 32 and the majority of workers make about $30,000 per year. These authors also surveyed workers about their motivation to work on MTurk. While a small number of workers reported using MTurk as their primary source of income, the majority responded that they found MTurk was "a fruitful way to spend free time and get some cash". These studies suggest a sample more diverse that the pool of undergraduates usually available to university researchers.

As an MTurk requester, I posted assignments for workers to complete. The responses from these assignments form the observations in my sample. I allowed agents to participate once in any given trial. I checked IP addresses so that individuals could not participate more than once using multiple MTurk accounts. An assignment contained all of the information necessary to make a forecast of a random variable. The information set provided to each worker was slightly different within each treatment, but I calibrated signals to ensure all workers could achieve the same expected payoff.

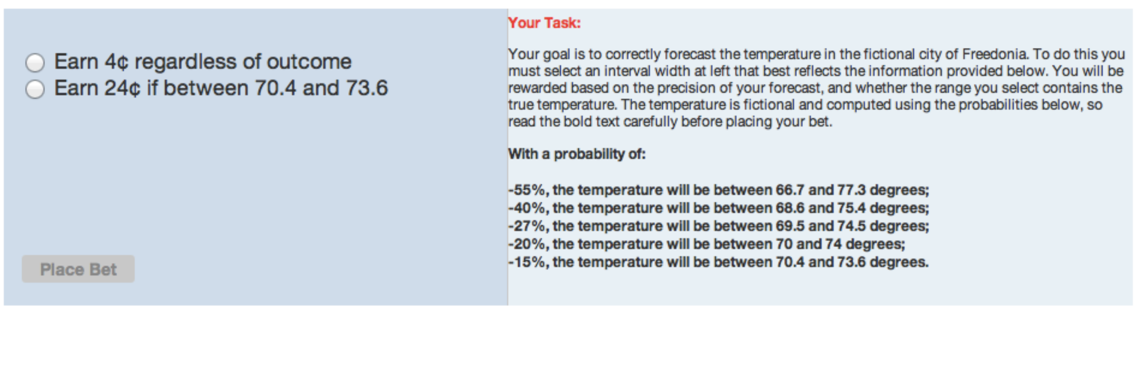


Figure 1.1: Sample Game Screen

The software allowed subjects to place interval bets on an underlying random variable taking values on the real line. The decision problem faced by each subject consists of choosing an interval from a menu. I promised subjects a payoff if the random variable's realization fell inside their interval. Narrower intervals yielded higher payoffs.

Figure 1.1 reproduces what an agent might see before placing a bet. On the right side of the page are instructions and a section of bold text containing the prompt assigned to a particular user. On the left is a set of lotteries from which subjects must choose. Workers select their preferred lottery and submit their work for approval. This concludes a worker's participation in the study. Section A.2 in the appendix details the steps a worker takes in the course of the study.

In these experiments, I endowed individuals with information about the likelihood of the random variable falling in each interval, with smaller probabilities associated with narrower intervals. I constructed the forecasting game so that each agent optimally chooses a bet interval that reveals the precision of his signal. Agents with more precise information about the underlying should choose narrower intervals than those who are less certain. I chose this setup to serve as an analogue for financial innovation, in the spirit of Axelson (2007). In some treatments, when I lengthen menus, I offer subjects riskier bets on the same random variable just as financial intermediaries might partition cash flows into claims of increasing risk.

All else equal, trade in riskier instruments suggests that agents possess higher-quality information about fundamentals. Section A.3 in the appendix presents a detailed numerical example illustrating how I adapt this concept to construct bet menus and demonstrates how the contracts separate expected payoff maximizing agents with signals of differing precision so that a higher-precision agent prefers a narrower, riskier bet.

By examining inefficient choices among the intervals offered to subjects I can quantify the effects on information transmission in a way that prior studies confined to product markets cannot. These failures of good decision making are not confined to choices between different

Version	N	Short Description
A	615	Two-, Four- and Six-alternative menus
B	171	Risky optimal in SHORT
C	170	Full specification of probabilities
D	106	Random removal of alternatives
E	82	MEDIUM optimal available in SHORT

Table I: Trial versions for alternative-based complexity

financial securities and their consequences are not confined to price efficiency. Other everyday mechanisms (bankruptcy proceedings, the tax code,) rely on self-selection to sort individuals into groups. As these become more complex, understanding the consequences of complexity on individuals and in aggregate data becomes more important.

1.3 Results

The following section describes experiments tied of each form of complexity and discusses results.

Alternative-Based Complexity

Choice overload suggests that agents who observe longer menus are confused and demotivated by too many alternatives and this makes it difficult for them to choose optimally. The evidence I collect suggests too little choice also comes with disadvantages. In particular, agents observing only two alternatives may believe they are less likely to find their best option and choose poorly as a result. Is there a level of alternative-based complexity that maximizes the efficiency of choices?

Conjecture 1. *Subjects choose more efficiently from menus that present an intermediate level of choice than from menus that offer too much or too little choice.*

I collected data to answer this question by randomly assigning workers into treatment groups that each encountered lottery menus of different lengths. The different trial versions I discuss here are enumerated in Table I. For example, in version A, I sorted subjects into one of three treatments. SHORT offered agents two alternatives, a five-cent bet and a 30-cent bet. MEDIUM added two intermediate intervals, 15-cents and 20-cents. LONG added a further two options, 10-cents and 25-cents.

All workers received signals of identical quality, meaning that the same probabilities were attached to intervals in all treatment conditions falling into a particular trial. For the

Option #	SHORT	MEDIUM	LONG	Probability
1	5*	5	5	1.0
2			10*	
3		15*	15*	> 0.40
4		20	20	< 0.20
5			25	
6	30	30	30	< 0.10

Table II: Treatments for alternative-based complexity, version A. Asterisks denote the best-responses of an expected payoff-maximizer in each treatment.

	Lottery Payoffs						
Treatment	5	10	15	20	25	30	Total
SHORT	58					147	205
MEDIUM	21		73	39		62	195
LONG	19	28	50	54	19	45	215
Total	98	28	123	93	19	254	615

Table III: Payoffs selected by subjects in version A

treatments presented in Table II, an expected payoff maximizer observing LONG would rank a 15-cent bet above all other options, except for the 10-cent bet, where there remained some ambiguity. Since the other treatments do not include the 10-cent option, subjects should strictly prefer the 15-cent lottery to all other lotteries in MEDIUM and a five-cent lottery to a 30-cent lottery in SHORT.

Results

To support Conjecture 1, I partition responses in each treatment into those that chose correctly, given the prompt, and those that did not. Performing Barnard's exact test on pairs of rows of contingency Table IV, I reject the null that choice in SHORT is better than choice in MEDIUM ($p = 0.04$) as well as the null that choice in SHORT is better than choice in LONG ($p = 0.06$). However, I cannot strongly reject the null that choice in LONG is better than choice in MEDIUM ($p = 0.42$).

The observation that subjects perform better in MEDIUM than in SHORT deserves special attention as it suggests oversimplification comes with accompanying efficiency costs. Despite only having to select between two lotteries with obviously different expected payoffs, subjects consistently choose poorly relative to counterparts allowed to select between more

Treatment	Dominating	Dominated	Total
SHORT	58	147	205
MEDIUM	73	122	195
LONG	78	137	215

Table IV: Contingency table for Barnard's exact test, version A

Option #	SHORT	MEDIUM	Probability	Expected Payoff
1	5	5	<0.50	<0.025
2		10		<0.05
3		15*	>0.40	>0.06
4	20	20	0.20	0.04
5	30*	30	0.15	0.045

Table V: Treatments for alternative-based complexity, version B. Asterisks denote the best-response of an expected payoff-maximizer in each treatment.

options.

The nature of private information in my experiment could underlie this result. When agents encounter SHORT, they cannot find the interval I suggest is highly likely. They have a choice between a risky, high-payoff lottery and a risk-free bonus. Given they do not see the option they really want, they may disregard the prompt entirely and choose the high-payoff lottery out of frustration or confusion.

To explore this phenomenon further, I ran versions of this trial varying the information I provided to subjects and the options available to them as I reduced the number of alternatives. Version B of this trial addresses the question of whether the poor performance in the smaller menu was simply due to the fact that the optimal payoff was the low-risk lottery. Table V contains the payoffs, probabilities and expected payoffs to lotteries for version B. While the 15-cent lottery dominates the alternatives offered in the MEDIUM group, the 30-cent lottery dominates the alternatives in the SHORT group.

As before, I partition the set of responses into those that chose the highest expected payoff lottery and those that did not. This contingency table appears as Table VI. I reject the null that subjects perform better in SHORT than in MEDIUM ($p = 0.03$). Even when the optimal bet is the riskiest alternative, subjects perform worse when selecting from the shorter menu.

This effect also appears robust to more precise specification of probabilities. In version C of this trial, I provided subjects with prompts that exactly specified probabilities at different

Treatment	Optimal	Other	Total
MEDIUM	34	48	82
SHORT	24	65	89

Table VI: Contingency table for Barnard's exact test, version B

Option #	SHORT	MEDIUM	LONG	Probability
1	4*	4	4	1.0
2			8	0.55
3		12*	12*	0.40
4		16	16	0.27
5			20	0.20
6	24	24	24	0.15

Table VII: Treatments for alternative-based complexity, version C. Asterisks denote the best-responses of an expected payoff-maximizer in each treatment.

Treatment	Optimal	Other	Total
MEDIUM	24	24	48
SHORT	13	29	42

Table VIII: Contingency table for Barnard's exact test, version C

levels. I present details of payoffs and probabilities from this trial in Table VII

With a more complete specification of probabilities, I expected the choice problem to become easier to solve for agents and this appears so for subjects in MEDIUM and SHORT. Indeed, in the case of MEDIUM, performance improved relative to version A. However, performance in SHORT still lags behind MEDIUM in terms of optimality (Table VIII). In particular, Barnard's test rejects a null that subjects in SHORT outperform subjects in MEDIUM ($p = 0.04$).

To check whether this result is robust to different ways of removing alternatives when shortening menus, I ran version D of this trial. Here, the longest menu still had six lotteries from which to choose with payoffs identical to version A (Table VII). I attached to these payoffs exactly the same probabilities. The key difference between this and versions A and C was that subjects encountering a four-option menu, saw four options drawn randomly from LONG and subjects encountering a two-option menu saw two options drawn randomly from LONG.

Treatment	Optimal	Other	Total
MEDIUM	13	15	28
SHORT	13	21	34

Table IX: Contingency table for Barnard's exact test, version D

Option #	SHORT	MEDIUM	Probability	Expected Payoff
1	5	5	1.0	0.05
2	15*	15*	0.40	0.06
3		20	0.20	0.04
4		30	0.10	0.03

Table X: Treatments for alternative-based complexity, version E. Asterisks denote the best-response of an expected payoff-maximizer in each treatment.

Treatment	Optimal	Other	Total
MEDIUM	10	30	40
SHORT	34	8	42

Table XI: Contingency table for Barnard's exact test, version E

Under random removal of alternatives, I fail to reject a null that subjects in SHORT choose optimally more often than subjects in MEDIUM ($p = 0.34$). Table IX contains the observations, partitioned into those that chose optimally, given the alternatives they observed, and those that did not. These data suggest that the effects of simplification in this context depend critically on the alternatives that I exclude for shorter menus.

I illustrate this by examining a further modification of this trial. In version E, I altered lottery payoffs offered in the SHORT menu so that the contract optimal in MEDIUM was still available to subjects (Table X). This change produces a reversal in the result, and performance in SHORT dominates performance in MEDIUM ($p < 0.01$) as depicted in Table XI.

Cases in which subjects performed worse with fewer options appear similar in that they require subjects to choose between high-probability, low-payoff lotteries and low-probability, high-payoff lotteries. To investigate this further, I examine the set of small, two-alternative menus, formed randomly in version C of this trial. To assess whether subjects do a poorer job of choosing between alternative payoffs when they carry very different probabilities, I

Coefficient	Estimates
$\hat{\beta}$	-1.039*
$\hat{\gamma}$	-2.523*
$\hat{\delta}$	0.649

Table XII: Conditional logit results for two-option menus (N = 59)

estimate the following conditional logit specification:

$$choice_{ij} = \beta * ev_j + \gamma * I_{j=1} * (p_{i2} - p_{i1}) + \delta * I_{j=1} + \varepsilon_{ij} \tag{1.1}$$

Here, β measures the sensitivity of subjects to the expected value of alternative j. The coefficient γ measures the additional probability of choosing alternative one for a larger difference in the probabilities $(p_2 - p_1)$ associated with two alternatives available to the agent. The final term controls for position on the menu. I present the results of this estimation in Table XII. The larger is the gap between probabilities associated with lotteries on a short menu, more more likely are subjects to choose the second, higher-payoff option, regardless of its expected value. Another interpretation of this result is that, for subjects who choose between lotteries with very different risk-return profiles, high payoffs are salient.

This result suggests that simplification by removing alternatives may come with the unexpected consequence of exacerbating other behavioral biases. When removing alternatives leaves agents with stark choices, choices become driven by other concerns. Subjects who encounter a menu with a low-probability, high-payoff lottery paired with a high-probability, low-payoff lottery systematically opt for the former. These subjects' behavior appears more consistent with payoff salience or risk-seeking than with expected-payoff maximization.

The behavior of subjects in this experiment adds nuance to the debates over the value of financial innovation from the standpoint of information transmission. Adding information-sensitive securities may promote learning and revelation of private signals. However, if giving error-prone investors access to these securities allows them more rope with which to hang themselves, this may render the additional information valueless. At the same time, these results suggest that removing alternatives in an effort to simplify choices might actually cause agents to make worse choices.

Information Transmission

These results on the accuracy of choices and biases have important implications for efficiency. An efficient price includes workers' private signals and weighs them based on signal precision. The analogue in my experiment is that a market maker or bookmaker observing selections should weigh narrower bets more heavily. Conjecture 1 shows that the availability

Treatment	Benchmark	Observed	Underestimate
SHORT	1.18	0.65	45%
MEDIUM	1.10	0.91	17%
LONG	1.10	1.01	8%

Table XIII: How alternative-based-complexity affects forecasts though inefficient decisions (version C)

of alternatives in the market affects the bets workers place. In turn, this implies that the structure of menus affects how well prices or odds reflect private information.

Table XIII presents the results of a thought experiment. Suppose 100 agents approached each of the three treatments and chose lotteries with the frequencies suggested by my subjects. I compute conditional variance as a bookmaker might, if he were presented with bet data and had a mapping between lotteries and signal precision. I compare this to an alternative where all agents truthfully report their signal precision to the bookmaker. Details about how I compute these measures are available in the appendix, Section A.3.

Oversimplication carries another cost. In my experiment, all subjects received signals of equal precision. In a more realistic setting, offering a wide variety of risk/return profiles would allow agents to sort themselves based on the strength of their signals. Too few options can impair this sorting. In the context of my study, any agent given a choice between a five-cent and a 30-cent lottery, with information not precise enough to make the 30-cent lottery, would pool in the five-cent lottery, reducing the information about precision communicated to the market.

Disorder

Decisions can become difficult for agents when presented with information about alternatives in an unintuitive way. Advocates of standardizing mortgage disclosure forms suggest that borrowers will find alternatives easier to interpret and evaluate when forms highlight key mortgage terms such as maximum rate adjustments and total closing costs. The third trial attempts to test whether better organization of alternatives can affect the efficiency of information transmission. Following Carlin (2009), this might reflect the efforts of a firm wishing to increase customers' cost of search. Alternatively, along the lines of Novemsky et al. (2007), shuffling the menu of alternatives reduces fluency, thereby influencing choice.

Conjecture 2. *Workers find decision-making more difficult when menus are not ordered. Additional difficulty results in poorer choices.*

To test whether subjects respond differently to organized versus disorganized menus, I exposed workers to treatments in which I varied the order of alternatives. Each treatment

Version	N	Short Description
F	303	Static GARBLE for all subjects
G	409	Randomized GARBLE for each subject

Table XIV: Trial versions for order-based complexity

Option #	ASCEND	DESCEND	GARBLE
1	5	25	15*
2	10*	20	25
3	15*	15*	10*
4	20	10*	5
5	25	5	20

Table XV: Menus offered to subjects for a study of menu order and choice, version F. Asterisks denote the best-responses of an expected payoff-maximizer in each treatment.

gave workers a choice between five different lotteries. ASCEND and DESCEND are self-explanatory. In version F, GARBLE presented all workers in the treatment with the same random shuffling of lotteries. All workers receive identical prompts that indicates a 15-cent lottery dominates all options except for the 10-cent lottery. If disorganization makes decisions more difficult for workers and this results in error, then I expect poorer choices in GARBLE than in the ordered menus.

Results

Conjecture 2 finds incomplete support in my data. Workers perform worse in GARBLE than in ASCEND ($p = 0.09$), but their performance is indistinguishable from DESCEND. Strikingly, the GARBLE menu places the optimal 15-cent bet right at the top of the list and the ambiguous 10-cent bet closer to its position in ASCEND. If a "primacy effect", whereby workers are generally more likely to choose options listed first, drove differences between treatments, I expect better performance from GARBLE.

I appeal to the ideas of fluency contained in Novemsky, et al. (2007) to help explain this observation. Shuffling around alternatives in the menu makes it more difficult for subjects to choose because it forces them to rank the lotteries in order of riskiness or payoff before making a selection. The additional difficulty causes them to deviate from maximizing expected payoffs to other behaviors.

In version G, I reshuffled the the lottery menu for each subject in the GARBLE treatment.

	Lottery Payoffs					
Treatment	5	10	15	20	25	Total
ASCEND	9	15	31	18	22	95
DESCEND	13	13	23	26	24	99
GARBLE	9	14	28	15	43	109
Total	31	42	82	59	89	303

Table XVI: Payoffs selected by subjects in version F

Payoff (cents)	Probability	Expected Payoff
5	1.0	5
10*	0.55	5.5
15	0.30	4.5
20	0.20	4
25	0.25	3.75

Table XVII: Payoffs and probabilities for a study of menu order, version G

	Optimal	Other
ASCEND	72	124
GARBLE	59	147

Table XVIII: Contingency table for a study of menu order, version G

I restricted attention to comparing an ascending menu against a shuffled menu. Table XVII contains payoffs and associated probabilities for this trial.

Table XVIII contains counts of optimal and suboptimal choices in the two treatments. Barnard's test rejects a null that subjects did no worse under GARBLE ($p = 0.02$).

Conditional Logit Estimation

Version G of this trial admits a conditional logit analysis of subjects' choices in each treatment. In this case, I construct an indicator variable, $shuffle_i$, that takes a value of 1 if individual i observes a shuffled menu and zero otherwise. Similar to my analysis of alternative-based complexity, I interact this variable with the expected payoff of each alternative.

To control for the fact that the ascending menu included high-expected payoff alternatives

VARIABLES	choice
$\hat{\beta}$	0.512***
	(0.146)
$\hat{\gamma}$	-0.488**
	(0.174)
$\hat{\delta}_2$	0.158
	(0.137)
$\hat{\delta}_3$	0.250*
	(0.139)
$\hat{\delta}_4$	0.002
	(0.161)
$\hat{\delta}_5$	-0.324*
	(0.181)
Observations	2,045
df	6
χ^2	46.149

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table XIX: Conditional logit results for disordered menus (N = 409)

near the top of the menu, I add a set of indicator variables $position_j(\tilde{j}), \tilde{j} \in 1...5$. Each of these take a value of 1 if $j = \tilde{j}$ and zero otherwise. Including these in the estimation allows me to control for "primacy effects" that bias subjects to choose more frequently from the top of a list. I estimate the equation:

$$U_{ij} = \beta ev_j + \gamma * shuffle_i * ev_j + \delta_j position_j(\tilde{j}) + \varepsilon_{ij} \tag{1.2}$$

My null hypothesis is that there is no difference in sensitivity that comes as a result of shuffling the order of lotteries on the page. This should be reflected by $\hat{\beta} > 0$ and $\hat{\gamma} = 0$. Table XIX contains brief results of this estimation. An estimate of γ that is both negative and significant suggests that subjects who observe a shuffled menu make choices that are less consistent with expected payoff maximization. Further, the coefficients on the controls, though not statistically significant, have signs consistent with a higher propensity to select from the top of the menu.

How much does shuffling the menu change subjects' sensitivity to expected payoffs? In version G, the average expected value of the lotteries offered to subjects was 0.0455 and the average position on the menu was 3. At this "average" alternative, for a subject who encountered an ordered menu, the marginal effect of a change in expected payoffs is 3.36%.

Treatment	Benchmark	Observed	Underestimate
ASCEND	1.32	1.08	18%
GARBLE	1.32	0.86	35%

Table XX: How disorder results in inefficient aggregation (version G)

This implies an elasticity of selection probability with respect to expected payoffs of 16.42%. By contrast, if a subject encountered the same average alternative in a shuffled menu, I estimate a marginal effect of expected payoffs of only 0.59% and an elasticity of 4.53%.

Appendix A.4 contains an extension of my application of conditional logit models that considers the degree of disorganization of menus.

Information Transmission

Table XX presents the results of the same thought experiment I carried in section 1.3 using the data produced by the second trial of disordered menu. My goal is to understand the externalities on information transmission imposed by disorganization of alternatives. Here, I compare the conditional variance of the underlying random variable for 100 bets in each treatment against a benchmark where agents correctly select the 10-cent bet. A simple reorganization of alternatives into ascending or descending order reduces the underestimate approximately 17 percentage points.

1.4 Robustness

Workers earn money based on the number of tasks of they complete, not the amount of time spent on their tasks. Despite offering workers richer rewards for carefully choosing from the menu of lotteries they observe, I expect some workers to randomly click through the experiment to earn their base fee. In Figure 1.2 I plot the amount of time recorded between login and bet placement for subjects in who encountered the menus in version C (Table VII). Figure 1.3 takes the same data and plots bet times separately for each treatment group.

To gauge the impact of "fast" workers, I apply Barnard's exact test to contingency tables formed from subsamples of the data that exclude bets made quickly. Generally, the results described so far are robust to removing these users for a variety of minimum bet times. Table XXI duplicates Table VIII, but for those subjects that took at least 40 seconds to respond.

I repeat the same exercises with data on menu ordering and choice. Figures 1.4 and 1.5 present bet times for subjects in version G (Table XVII). Table XXII is the contingency

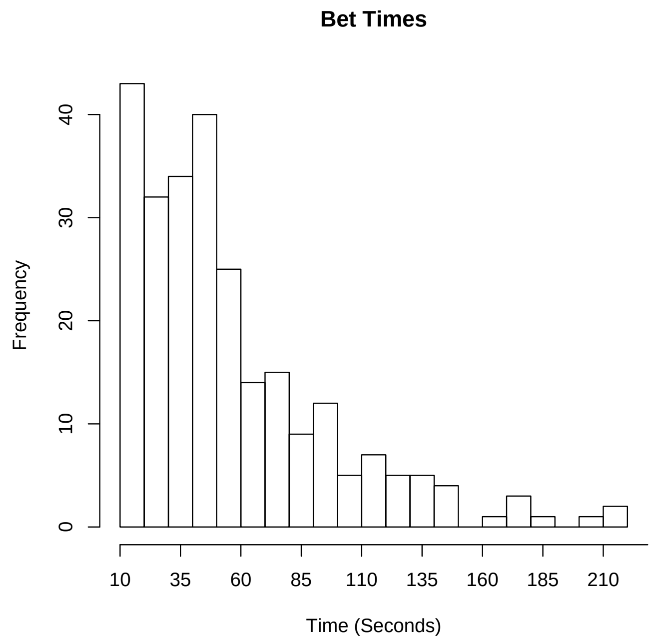


Figure 1.2: Number of seconds taken to place each bet, alternative-based complexity

Treatment	Optimal	Other
MEDIUM	18	15
SHORT	7	15

Table XXI: Contingency table for subjects taking at least 40 seconds to respond (version C)

	Optimal	Other
ASCEND	41	57
GARBLE	29	63

Table XXII: Contingency table for a second test of disordered menus (version G)

table for subjects who took more than 40 seconds to place a bet. Rejection here is weaker ($p = 0.10$).

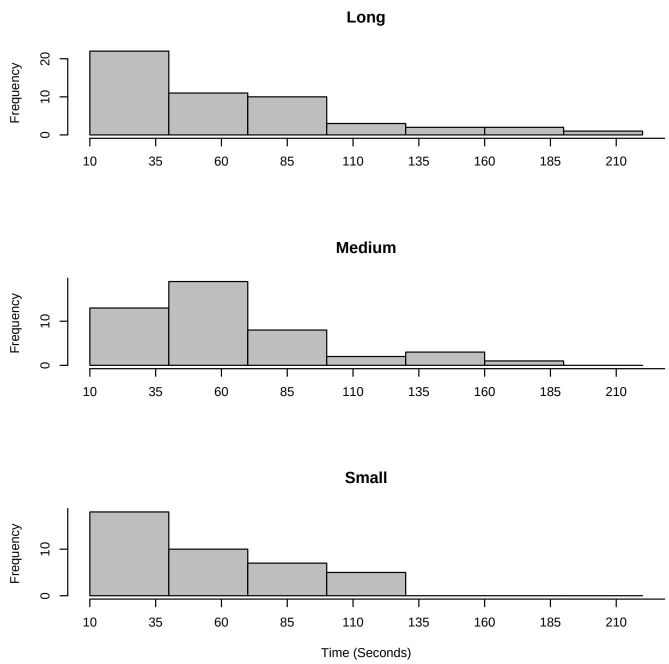


Figure 1.3: Number of seconds taken to place each bet, by treatment

1.5 Conclusion

This paper examines the effects of complex environments on individual decisions and considers how these effects may aggregate. I focus on two forms of complexity: alternative-based complexity, and the organization of alternatives. I observe how subjects recruited to play a forecasting game respond to different levels of complexity and I use their responses to quantify the effect of these forms of complexity on forecast efficiency.

While I find that subjects tend to perform poorly when choosing from more alternatives, I also find that removing too many alternatives can also reduce the efficiency of individual choices. When I examine how subjects respond to disorganized menus, I find that without an intuitive ranking of alternatives, subjects tend to choose dominated options.

Broadly, my results suggest that some degree of simplification can improve the welfare of individual agents. Moreover, in settings such as financial markets, where prices convey information about fundamentals, my environment allows me to show how these effects of complexity on individual decision making reduces the efficiency of information aggregation.

A precise understanding of the effects of complexity is important. Ever greater diversity in financial products burdens individuals. At the same time, circumstances force households to make more financial decisions for themselves. Where and how much should I save for retirement? What type of mortgage should I choose? Should I declare bankruptcy? Cru-

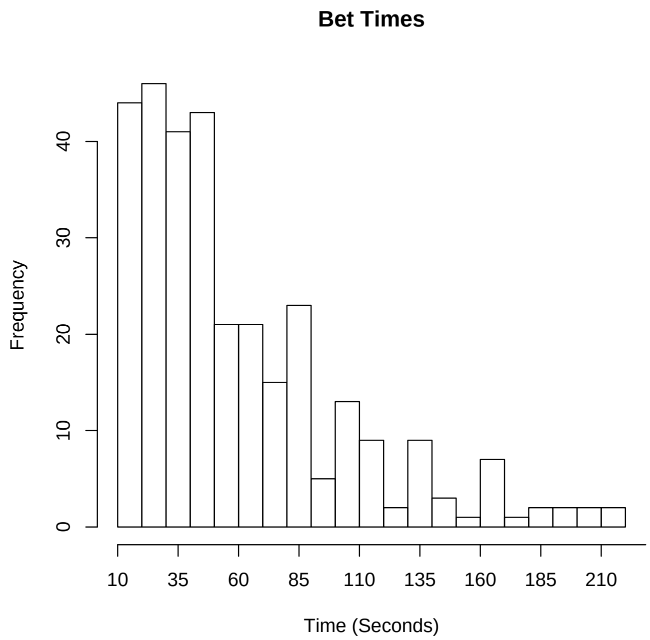


Figure 1.4: Number of seconds taken to place each bet, menu order

cially, if the decision making environment discourages or confuses participants, then prices are unable to fully incorporate the information they possess.

My findings contribute to current policy debates surrounding reform of consumer finance, suggesting potential avenues for simplification. One key step is better organization of information disseminated to agents; constraints on the number of available investment alternatives could be another. While these small changes clearly impact the lives of individuals, they also enable them to effectively communicate their preferences and beliefs to the market.

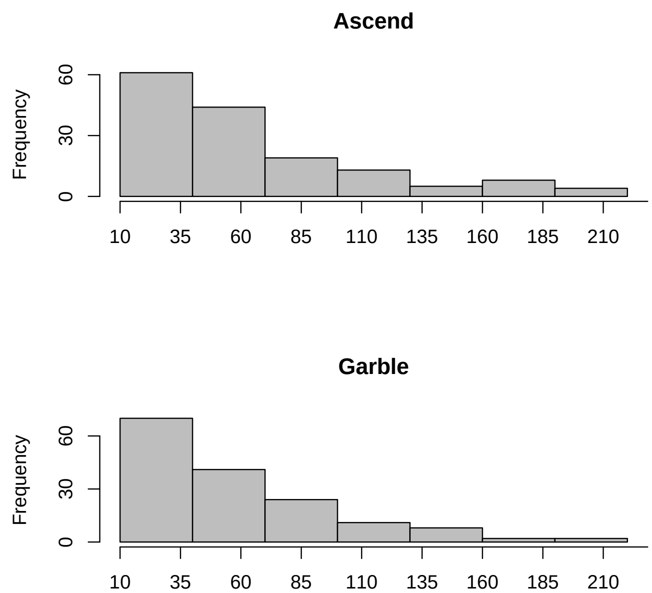


Figure 1.5: Number of seconds taken to place each bet, by treatment

Chapter 2

Information Inside the Firm - Evidence From a Prediction Market

2.1 Introduction

Who is informed about firms' prospects? A lot of corporate finance theory is predicated on asymmetric information between managers and outside financiers; manager- and investor-side private information has implications for capital structure and security design. In this study, I examine the quality of managers' information. More generally, I use bet data from a corporate prediction market to evaluate how informed employees are about firm prospects. Who within the firm possesses private information that is relevant to managerial decisions? Which sets of employees uncover private signals about project outcomes and are these employees able to articulate their beliefs in a market setting?

For a period of over two years, a cross-section of employees at a software firm were allowed to place bets on variables relevant to management such as product quality scores, completion dates and unit sales. This unique dataset allows me to infer the degree to which employees possess private information about the firm by looking at the payoffs they receive from their bets as well as other measures of informedness. I find some evidence that groups differ in their ability to choose underpriced bets and in how early they place bets relative to their peers. One group in particular, Publishing, appears to perform better than others. I attribute this to increased contact with the firms products and customers.

The dataset also allows me to consider the dynamics of information within the firm. I find that employees communicate private information to the betting market when they modify existing bets. They systematically revise to bets that are more underpriced than the bets they initially held. I also find that the informedness of these revisions varies by job function. Employees responsible for publishing and distributing software title do a particularly good job of updating their positions. These employees tend to have a high degree of contact with products and customers I show how the accuracy of bets and the frequency with which employees revise bets varies by job function. This analysis reveals differences in the quality of information and the frequency of signals employees within the firm receive.

In the context of this paper, information arrival and information acquisition are not static. Employees participating in the prediction market learn information about outcomes over time. Without a liquidity motive for trade, employees initiate or rebalance their positions in the betting market when they receive information that presents opportunities for profit. I use the frequency and performance of these bets to infer the rate at which employees in different groups learn. Moreover, profits in the prediction market reflect the quality of information possessed by employees who participate.

A long line of research supports the idea that markets aggregate information. Roll (1984) presents evidence that orange juice futures contracts provide information beyond official weather forecasts. Earlier work on race-track betting by Figlewski (1979) shows that competitive bettors fully discount the forecasts of professional handicappers. Moreover, other authors have evaluated prediction markets as a means to aggregate information in places where markets do not already exist. Wolfers and Zitzewitz (2004) provide a useful survey of market design and application from the Oscars to presidential elections. Manski (2006) and consider the problem of inference from prediction market price data. Berg and Rietz (2003) evaluate the usefulness of prediction markets for decision support.

A preliminary question is why prediction markets are useful in a corporate setting. Ortner (1998) provides an early example of a prediction market used to support project management in the telecommunications industry. Chen and Plott (2002) report results from a trial at Hewlett Packard; Cowgill, et al (2008) present a prediction market experiment at Google. Absent strategic motives, a simple survey could produce a truthful forecast of outcomes. If managers suspect information arrives over time, they might repeat the same survey. A number of assumptions underlie the informational efficiency of such a scheme. The survey administrator must have a sense of when information has changed and must know which employees are informed. If information acquisition requires even minimal effort on the part of employees, a survey might not provide sufficient incentives for information production.

A prediction market such as the one I examine allows for the weakening of some of these requirements. Allowing employees to revise bets over time, a continuous market mechanism elicits new information as beliefs change. A betting market where employees can alter the riskiness of their bets helps managers assess how confident employees are in their beliefs. Finally, the incentives promised to top performers might encourage employees to put some effort into acquiring information about outcomes.

Using data from a prediction markets implementation I will identify rebalancing behavior and show that this rebalancing behavior is informed. Users profit from revising their positions and new positions more accurately reflect observed outcomes. This evidence shows that the market mechanism's flexibility in incorporating new information represents an advantage over surveys for eliciting information.

2.2 Market Design

Employees access the betting market using a computer interface with software designed by Crowdcast. This startup sells prediction market software to managers who hope to elicit information from employees. Managers, in turn, use data produced by employee gameplay to inform decisions about budgeting and supply chains. For example, if management learns that employees forecast a product to be of low quality, they might choose to reduce the size of the marketing outlay for that particular product or consider remedial measures to improve the product.

Besides constructing a software tool for the firm that I study, Crowdcast also managed basic data reporting and the distribution of incentives (prizes) to employees who performed well. Crowdcast's position as an intermediary between employees and management yielded a degree of anonymity to employees. Conceivably, this made it easier for employees to bet truthfully. At the same time, anonymity was not totally guaranteed - Crowdcast could report behavior that resembled manipulation or collusion.

The market is composed of a set of *forecasts*, chosen at the discretion of management. Employees log onto the application and enter with an endowment of *credits*, or currency within the system. Credits have three important features. First, the site administrator chooses the initial endowment of credits allocated to employees who participate. Second, employees may not transfer credits between one another. Finally, employees may not redeem credits for cash; the site administrator redeems credits for prizes according to a *prizing schedule*. Figure 2.1 contains a menu of variables on which employees in a fictional betting

Figure 2.1: A market concerning new product introduction

market might place bets. On entering the market, an employee may bet in any of four forecasts. Each of these forecasts corresponds to a random variable whose realization the firm wishes to predict. An employee may believe she has information about the number of new product units ordered by customers for July delivery. The computer system allows her to click a link and place a bet on this particular variable.

An employee who clicks to bet on the number of units ordered for July delivery sees a screen depicting the aggregate beliefs of the *crowd* of other participants in the form of a distribution function. These beliefs are formed from (1) a prior chosen by the market administrator; and (2) signals inferred from all preceding bets. Figure 2.2 shows the crowd density for units ordered for July delivery.

In this prediction market game, participants place *bets* on continuous variables by specifying a bet *interval* and a number of credits to wager. In Figure 2.2 an employee has selected an interval between 4.2 million units and 5.5 million units. If the firm experiences unit sales within the interval specified by the agent, she will receive a *promised payoff* in credits.

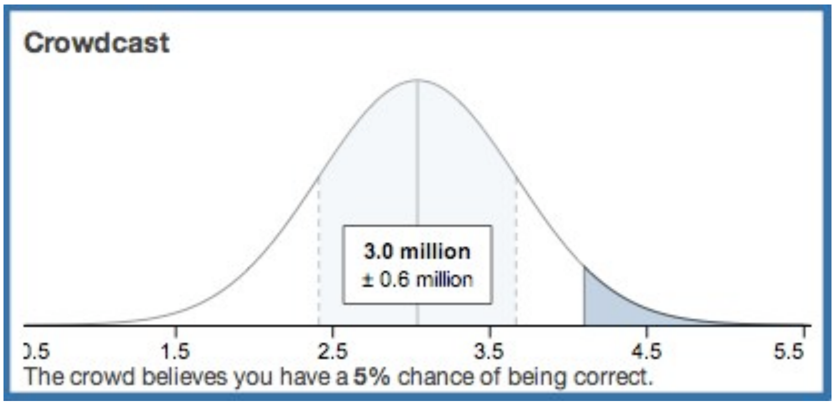


Figure 2.2: Bet placement in a forecast

In this example, the promised payoff associated with the employee's bet is approximately 20 times the bet amount (Figure 2.3), corresponding to the crowd's belief that the likelihood of realized unit sales within the bet interval is about 5%. Had the employee instead chosen an interval carrying 10% probability in the crowd's estimate, then the promised payoff would decrease to 10 times the wager.

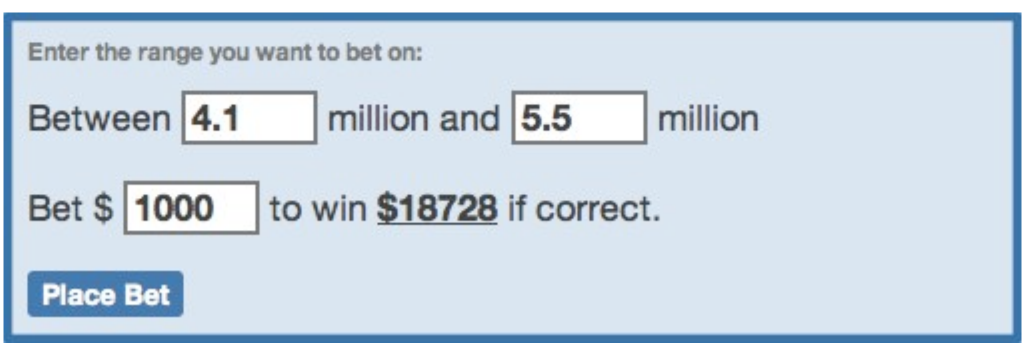


Figure 2.3: An employee bets on unit sales

Once she submits a bet to the market the employee receives a *contract* specifying a bet interval, a wager amount and the promised payoff. The system distributes payoffs once forecasts close and a measurement of the underlying variables becomes available. In this example, if the realization of unit sales is outside of the user's bet interval, she loses her stake. If her bet pays off, then she receives her promised payment of 18728 credits.

This market design achieved a number of goals, foremost of which was to appeal to employees without much training. For continuous random variables like quality scores or ship dates, eliciting intervals seemed natural. Allowing employees to choose their own bet amounts gave them an additional way to communicate their confidence in their forecast. Odds based on crowd beliefs were easy for employees to interpret and had the additional benefit of rewarding early activity. Intuitively, if signals are conditionally-independent, employees have an incentive to wait for others to place bets and observe their activity. When odds are based on the crowd's beliefs, an agent who waits risks losing information rents to competitors.

Management determines the length of time that forecasts remain open. So long as a forecast is open and not yet suspended, employees may withdraw bets or revise them. Users may hold one bet in a forecast at a time.

Between trades, prices reset as a result of activity according to an updating algorithm. Only transactions move prices. When an employee submits a bet through the interface, an

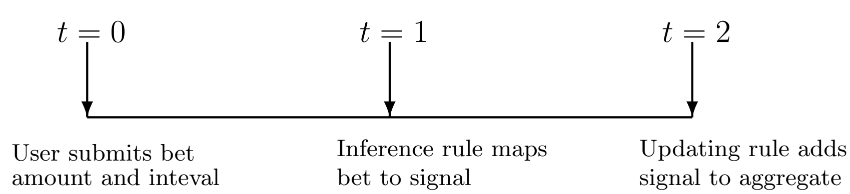


Figure 2.4: How bets lead to updates

algorithm estimates parameters of the employee's signal from from the interval and wager size. The algorithm infers the location of an employees beliefs based on the location of the bet interval and measures signal precision using the size of the wager and the width of the bet interval. This conforms to intuition. High signal precision implies less dispersion in conditional beliefs and a more peaked posterior distribution. This is consistent with a contraction in bet intervals. A more peaked distribution results in a larger degree of perceived underpricing and hence a higher quantity demanded. I will return to this in Section 2.4 when I examine how information leads employees to place and revise bets. Once the software has imputed the signal that produced an employee's bet, her signal is added to those of other bettors in the system.

The market provides incentives by awarding prizes. Employees receive a ranking based on the number of credits they earn during each month. These rankings map to Amazon gift cards of different denominations. I provide more detail about prizes in Section 2.3. At any point in time participants may call up a real-time *leaderboard* showing their performance relative to other participants. In Figure 2.5, Brooke is ranked the fourth highest earner in the period beginning 1 May, 2010.

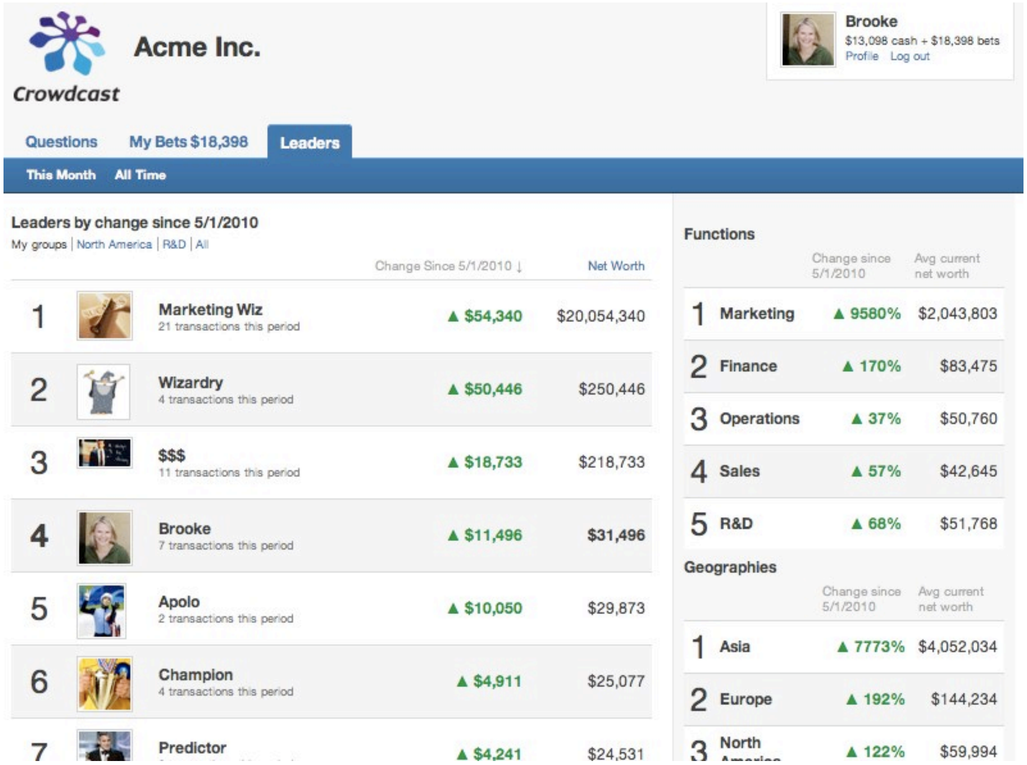


Figure 2.5: A typical leaderboard

2.3 Data

The data on which I base my analysis comes from a single, ongoing implementation within a software firm. In opening a prediction market, management hoped to elicit private information about a number of different variables belonging to the categories listed in Table I. My sample contains a total of 264 different variables about which management tried to elicit information between 20 March, 2009 and 9 March, 2011. During this period, employees placed 27414 bets. Of these bets in my sample, employees cashed out 11474 prior to close and held 15940 to close.

Table I: Variable Types

Type	# of Forecasts
Product Quality	130
Unit Sales	51
Dollar Sales	51
Ship Dates	10
Growth Rate	3
Other	19

The plurality (130) of questions asked by management concerned product quality scores. In this particular industry, an independent body produces quality scores by aggregating product reviews from multiple reviewers. Scores lie on the interval $[0, 100]$ and signal overall product quality to wholesale buyers who use them as a basis for orders. Since these questions made up a bulk of my sample, I focus attention here. Management also asked employees to help forecast unit sales (51) and dollar sales (25). The remaining forecasts included ship dates (9), growth rates (3) and market share (1).

The shortest forecast lasted one week from inception to close while the longest lasted for one year. The average forecast lasted for 114 days from inception to close. Forecasts suspended prior to close. The period between forecast suspension and realization of the underlying random variable was typically between one week and one month.

868 employees registered to participate in the prediction market. Of these registrants, 598 chose to place at least one bet in the system. Conditional on placing a bet in the system, employees placed 46 bets, including revisions, on average. Participation was highly variable, with a standard deviation of 77. Figure 2.6 is a histogram of the number of bets, including revisions, placed by employees in my sample.

Besides having unique identifiers, employees in the betting market carry location and function tags. Figure 2.7 plots of employee counts by group. Product developers and quality control represent the largest share of market participants. Other participants come from the head office, marketing and publishing. A fairly large number of employees remain unclassified. Crucially, trading behavior varies by group, as shown in Figure 2.8. Functions "Corporate" and "Marketing" are more managerial in nature. By contrast, employees in quality control and software development ("Quality" and "Development", respectively,) work closer to products.

Management chose how to reward employees for their participation. With the help of site administrators, they ranked employees based on their earnings from month to month and mapped rankings to prizes according to a prizing schedule. Table II contains the prizing

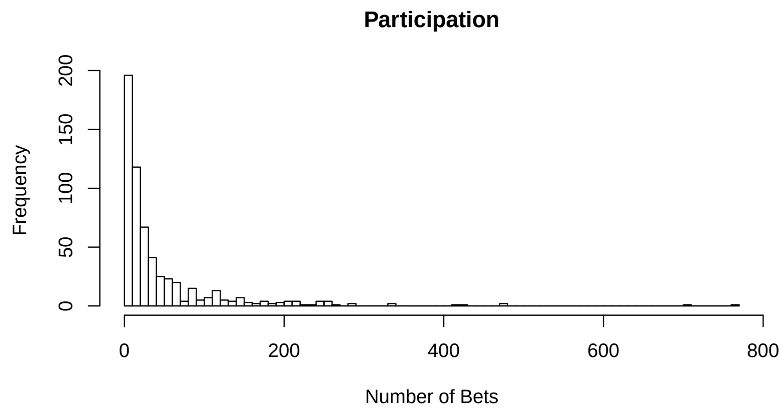


Figure 2.6: Bets and revision counts for each employee

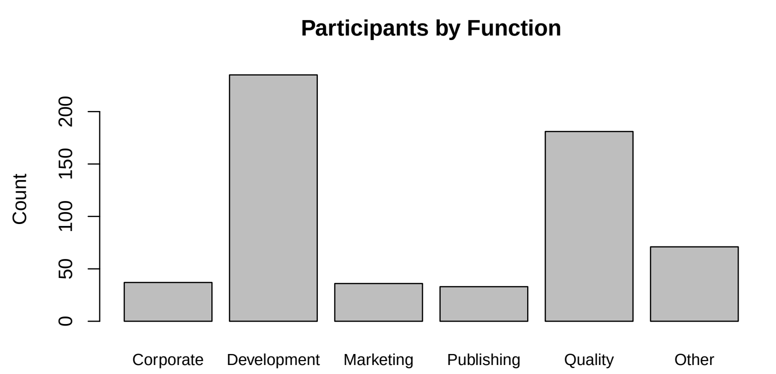


Figure 2.7: Distribution of participants across functional groups.

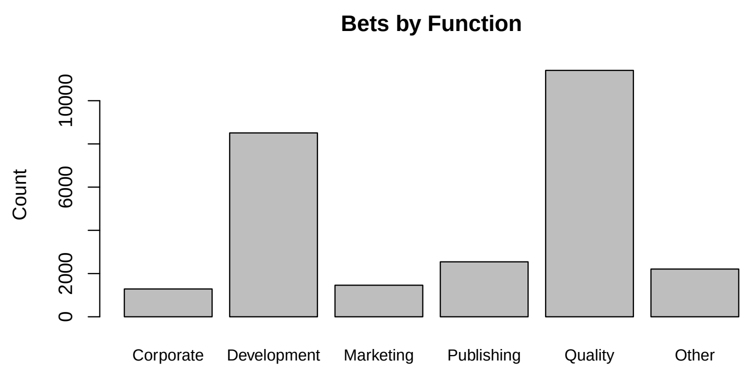


Figure 2.8: Bets placed by employees, by group.

schedule in effect during my period of observation. As shown in the table, the shape of the prizing schedule changed midway through my sample. The change was exogenous and is the subject of a related study.

Rank	Value
1-10	\$100
11-30	\$25
31-100	\$10
Total Payout	\$2200

(a) Until 5/2010

Rank	Value
1-10	\$75
11-20	\$20
21-50	\$10
Total Payout	\$1250

(b) 5/2010 onwards

Table II: Prizing schedules

During the period of study, the prediction markets provider implemented two constraints on bets mid-way through the sample. Between 8 July, 2009 and 3 May, 2010, employees could wager at most 10000 credits on any variable, equal to the initial endowment in the system. Further, between A leverage constraint imposed a 5%-limit on bet intervals so that a contract's maximum promised payoff is 20 times the bet amount. Further, for much of the sample, the system set the maximum acceptable bet amount at 10000 credits, equal to the initial endowment. Figures 2.9 and 2.10 present the distribution of odds and wager amounts, respectively and show how often these constraints bind in my sample.

Employees do not appear constrained by the 5% lower bound on crowd probabilities. Only 1.4% of my sample of bets carried probabilities this low. By contrast, the maximum wager amount binds often. The system offered employees inefficient odds and provided an implicit subsidy in credits. As employees became richer they became more likely to strike this upper bound. Bet interval widths likely provide more information about the precision of private information than wager amounts in my sample.

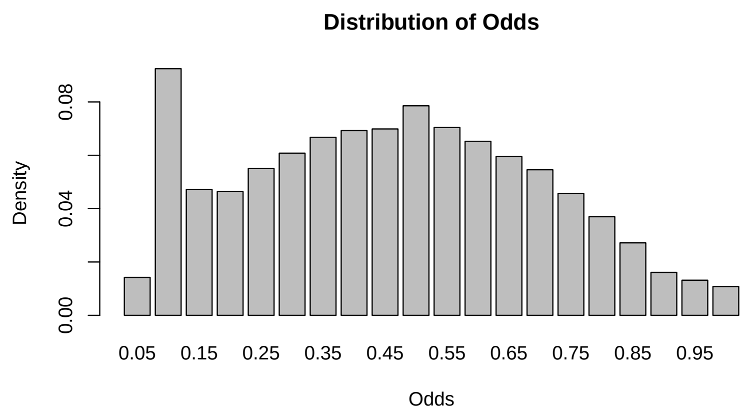


Figure 2.9: Distribution of odds in constrained sample.

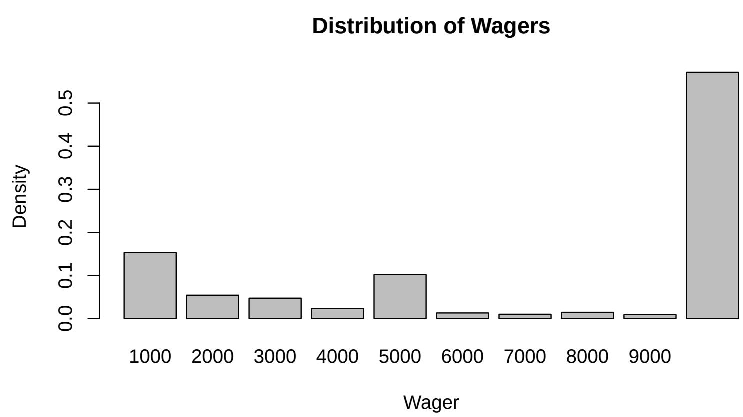


Figure 2.10: Distribution of wagers in constrained sample.

2.4 The Single Agent Decision Problem

As an econometrician, I attempt to back out features of employees' private signals from the data produced by the betting market. This requires building some intuition about what motivates employees' choices of bet intervals and wager amounts and what might cause them to revise their bets. The following section addresses the betting problem of a single agent who updates her beliefs based on her signal and observes the current distribution associated with crowd beliefs for a given forecast. I ignore strategic considerations in the interval betting game. Agents are aware of competitors but do not consider their number or relative informedness. Agents also ignore any impact they have on prices offered by the market maker.

In the context of the prediction market I examine, these assumptions seem reasonable. The participant base is relatively large and distributed across the firm. While aware that they were competing with their peers, participants had the opportunity to choose nicknames in the betting market to retain anonymity. Finally, the betting behavior of others was fairly opaque. Employees could observe the aggregate number of credits wagered on a particular forecast but could not observe the locations of individuals bets.

For tractability, I model participants as expected utility maximizers with utility over wealth. This has clear shortcomings considering the tournament incentives faced by employees in the betting market. At the end of each prizing period, the market rewards agents with the highest earnings during the period. Because participants do not receive an income and cannot borrow, losing credits makes it difficult to achieve high earnings. As a result, I expect employees to exhibit risk aversion and I model them as agents with concave utility over wealth.

The Competitive Bettor's Problem

An agent i arriving at time t solves the problem:

$$\max_{\theta_{it}, a_{it}} V_{it} = E_{it}[U_i(W_{iT})] \tag{2.1}$$

subject to the budget constraint:

$$W_{iT} = W_{it} + \theta_{it} \left(\frac{\mathbb{1}}{1 - P_t(a_{it})} - 1 \right) \tag{2.2}$$

The expectations operator E has subscripts i and t. This is meant to reflect that expectations are taken with respect to trader i's information set at time t. This captures the fact that private signals differ both across agents and through time. The term θ_{it} represents the number of credits agent i stakes on his time t bet and a_{it} represents one endpoint of the bet interval, either the upper or the lower bound.

I focus on a case where agent i's conditional beliefs first-order stochastically dominate the market prior. This suggests optimal bet lies in the upper tail of the conditional distribution and a_{it} is a lower bound. The case where agent i's signal is pessimistic is symmetric. I focus on a problem of *tail betting* rather than the interval betting problem faced by employees in the betting market to keep my model tractable. I must assume that in practice agents place an upper bound on their bet interval that leaves an arbitrarily small probability mass in the upper tail.

Moving to the budget constraint (Equation 2.2), W_{it} is the time t wealth of agent i, the sum of her initial endowment, W_{i0} and all gains and losses incurred up to time t. The term $P_t(a_{it})$ is the crowd's probability mass in the upper tail and represents the inverse of the ask price for a one-credit payoff on the specified tail. The higher is this mass, the lower will be the promised payout associated with the interval bet. Finally, $\mathbb{I}$ is an indicator variable taking on a value of 1 if the bet is a winner and 0 if the bet is a loser.

Since the following discussion concerns a single agent at a single point in time, I suppress all identity and time subscripts. I partition states of the world into two groups, one set in which the bet pays (+) and another in which the bet does not (-). Define W^+ as wealth in the winning (up) state and W^- as wealth in the losing (down) state. Define G as the probability mass in the upper tail of the trader's beliefs conditional on the agent's information set. With all of these simplifications, problem (2.1) becomes:

$$\max_{\theta,a} V = GU(W^+) + (1-G)U(W^-) \tag{2.3}$$

subject to:

$$\begin{aligned} W^+ &= W_t + \theta\left(\frac{1}{P} - 1\right) && (2.4) \\ W^- &= W_t - \theta && (2.5) \end{aligned}$$

First-order Conditions

I derive first-order-conditions for a general expected utility representation, letting $U'(W)$ represent the first derivative of agents' utility. I start with θ, holding a fixed. Differentiating with respect to θ yields:

$$\frac{\partial V}{\partial \theta} = \left(\frac{1}{P} - 1\right) GU'(W^+) - (1-G)U'(W^-) \tag{2.6}$$

Setting this equal to zero yields the first-order-condition:

$$0 = \frac{U'(W^-)}{U'(W^+)}\frac{1-G}{G} - \frac{1-P}{P} \tag{2.7}$$

or another way:

$$\frac{U'(W^-)}{U'(W^+)} = \frac{1-P}{P}\frac{G}{1-G} \tag{2.8}$$

This condition relates marginal utility across states to the difference in beliefs between the agent and the crowd.

Lemma 1. *For a fixed endpoint a, a risk-averse agent will choose a wager amount increasing in the degree to which her beliefs differ from the crowd's.*

This process is slightly more difficult for a, since P and G depend on this control. I do not report all of the steps below. I fix θ and differentiate to find:

$$0 = g(a)[U(W^+) - U(W^-)] - GU'(W^+)\theta\frac{p(a)}{P^2} \tag{2.9}$$

Reorganizing terms and defining $\Delta U = U(W^+) - U(W^-)$ gives:

$$\frac{g(a)}{p(a)} = \frac{G}{P^2}\frac{U'(W^+)\theta}{\Delta U} \tag{2.10}$$

The left-hand-side of the relation is the ratio of two density functions at the lower bound, a. If the agents selects a lower bound at the intersection of the prior density and the conditional density, this ratio is one.

The magnitude of the ratio $\frac{U'(W^+)\theta}{\Delta U}$ depends on the agent's risk preferences. Risk-aversion implies this ratio is smaller than one. On the other hand, $G > P$, so that the term $\frac{G}{P^2} > 1$. This conforms to intuition: the location of the endpoint of the bet interval depends on risk tolerance and the degree to which conditional beliefs differ from prior beliefs.

Lemma 2. *For a fixed level or risk tolerance, the endpoint of a bet interval reflects the degree to which an agent's private beliefs differ from the crowd's.*

First-order Conditions Under CARA Preferences

Suppose the agent's utility function is of the form:

$$U(W) = -e^{-\rho W} \tag{2.11}$$

where $\rho > 1$ is the coefficient of absolute risk aversion. This assumption allows me to get a little further towards a closed-form solution. The first-order condition for θ becomes:

$$\theta = \frac{P}{\rho}\log\left(\frac{1-P}{P}\frac{G}{1-G}\right) \tag{2.12}$$

I substitute this into the first-order condition for a:

$$\frac{g(a)}{p(a)} = \frac{G}{P^2}\frac{U'(W^+)\theta}{\Delta U} \tag{2.13}$$

$$= \frac{\log\left(\frac{1-P}{P}\frac{G}{1-G}\right)}{\frac{1-P}{P}\frac{G}{1-G} - 1} \tag{2.14}$$

Comparative Statics under CARA Preferences

Define $\delta = \frac{1-P}{P}\frac{G}{1-G}$, summarizing the difference in payoff probabilities implied by the agent's beliefs and the crowds beliefs. Consider an agent with CARA preferences who chooses a pair $\{\theta^0, a^0\}$ where:

$$\theta^0 = \frac{P}{\rho}\log(\delta^0) \tag{2.15}$$

$$\frac{g(a^0)}{p(a^0)} = \frac{\log(\delta^0)}{\delta^0 - 1} \tag{2.16}$$

What happens when this agent receives a new signal that generates a thickening in the upper tail of her conditional beliefs? Notice that a thicker upper tail implies $\delta > 1$, so that $\theta^0 > 0$. The ratio $\frac{\log(\delta)}{\delta-1}$ is decreasing in δ and $\lim_{\delta\to 1}\frac{\log(\delta)}{\delta-1}$ from either side, so that $\frac{g(a^0)}{p(a^0)} < 1$. Alternatively, a^0 lies in a region where the conditional density lies below the density associated with crowd beliefs.

The agent will meet a small increase in δ by moving her lower bound to a^1. I assume conditions on beliefs so that at equilibrium, the derivative of δ with respect to a is greater than zero. This means that the agent can respond by adjusting her lower bound to $a^1 > a^0$. Suppose this new choice results in $\delta^1 > \delta^0$, so that the agent revises the size of her bet to $\theta^1 > \theta^0$.

Lemma 3. *Changes in conditional beliefs, relative to the crowd, generate changes in both wager amounts and bet endpoints.*

2.5 Bets and Information

Are employees within the firm privately informed about variables that are useful to management? I begin by examining the performance of the prediction market. Managers "seed" each forecast with prior beliefs. I take these values to represent management's information set at inception. Once a forecast is open, employees place bets and an automated market maker updates crowd beliefs according to an algorithm. Each forecast *suspends* a number of days or weeks prior to the realization of the underlying variable. At realization, the forecast *closes* and bets pay out.

To evaluate the performance of the prediction market, I begin by comparing crowd beliefs at inception to crowd beliefs at suspension. If the market maker elicits and aggregates private information, then I expect beliefs at suspension to better reflect actual outcomes than beliefs at inception. For the forecasts in my sample, management employed an updating model that summarized beliefs as a normal distribution. I obtained parameters for each forecast at inception and at suspension.

I develop and compute a measure of the distance between realizations and beliefs. Suppose crowd beliefs at time t are $X \sim \mathcal{N}(\mu_t, \sigma_t^2)$ with density function $f_t(x)$ and that the realization is $\bar{x}$. Define $z = |\bar{x} - x|$ as the loss associated with x when the realization is $\bar{x}$. I compute a score s_t at time t as the expected loss at time t.

$$s_t = E_t[z] \tag{2.17}$$
$$= \int |\bar{x} - x| f_t(x) dx \tag{2.18}$$
$$= \int_{-\infty}^{\bar{x}} (\bar{x} - x) f_t(x) dx + \int_{\bar{x}}^{\infty} (x - \bar{x}) f_t(x) dx \tag{2.19}$$
$$= \bar{x}(2F_t(\bar{x}) - 1) + \int_{\bar{x}}^{\infty} x f_t(x) dx - \int_{-\infty}^{\bar{x}} x f_t(x) dx \tag{2.20}$$

This score represents the expected loss for each set of beliefs and is positive for all $\sigma_t > 0$. If $\mu_t = \bar{x}$, score varies only as a function of σ_t. For a fixed σ_t, score increases with $|\mu_t - \bar{x}|$. For each forecast, I compute s_0 and s_T. Figure 2.11 contains a histogram of $s_T - s_0$ for forecasts in my sample. Values less than zero indicate an improvement in score. The average improvement is less than zero, though the measure is very noisy in sample. This indicates that, on average, the forecasts generated by the prediction market yield a lower expected loss supporting the idea that the market elicits and aggregates private information and that managers learn from bets.

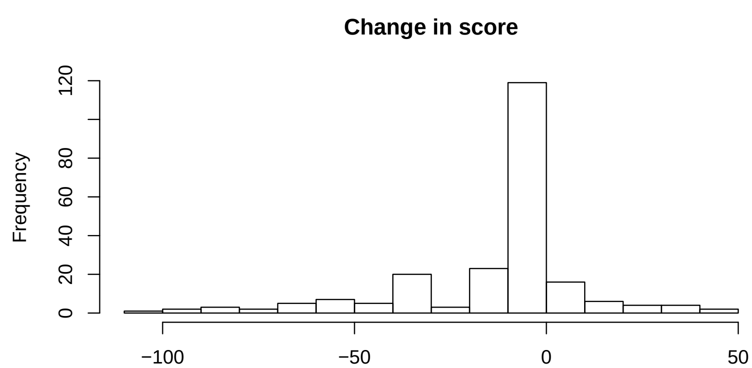


Figure 2.11: Change in expected losses, forecast initiation to forecast suspension

One problem with this evaluation is that it jointly assesses elicitation and aggregation. To get a clearer sense of whether employees' bets contain private information, I compare the payoff probabilities employees accept when they place bets with the ex-post likelihood that their bets pay out. If employees behave according to the theory I presented in Section 2.4, then they choose underpriced intervals on which to place wagers. I measure whether bets are underpriced using a logit specification:

$$\mathbb{I}_{win,it} = \alpha + \beta_1 odds_{it} + \varepsilon_{it} \tag{2.21}$$

I construct a dummy variable, $\mathbb{I}_{win,it}$ equal to one when the bet interval contains the outcome and zero otherwise. I regress this dummy on the payoff probability that employee i accepts at time t. I include payoff probabilities as a transformation:

$$odds = \log\left(\frac{P}{1-P}\right) \tag{2.22}$$

The intercept, α, estimates the average underpricing of bets in my sample. With no underpricing, I expect $\alpha = 0$, however, my estimate $\hat{\alpha}$, is significant and positive. This supports the idea that employees choose underpriced bets and implies that their actions in the prediction market communicate private information.

Bets and Job Function

I examine information dynamics by job function in Section 2.6. Here, I look at heterogeneity in information from a static perspective. Does the data generated by this prediction market reveal anything about the distribution of private information in the firm? Table III contains measures of informedness for each function. I report the number of participants and credits per participant at the end of my sample and I compute bet-level measure using only bets held until each forecast closed. These statistics communicate the final state of information about each variable prior to realization.

Function	Participants	Credits per Participant	P	P/H	Timing
Corporate	37	1337020.74	0.48	0.93	0.74
Development	235	1816122.19	0.49	0.93	0.71
Marketing	36	2750635.86	0.38	0.84	0.76
Publishing	33	4506861.87	0.36	0.67	0.79
Quality	181	2781000.08	0.45	0.89	0.76
Other	346	609152.20	0.43	0.88	0.78

Table III: Measure of informedness, by functional group.

The measure "P" represents the average payoff probability of all bets made by employees in the group, while "P/H" normalizes this by the hit rate of bets these employees make. P/H suggests the degree to which bets chosen by group members are mispriced. Employees in Corporate and Development tend to choose the least underpriced bets, while those in Publishing choose the most underpriced bets. This suggests that employees in Publishing are particularly well-informed relative to their peers. I will provide further evidence of this when I examine revised bets by job function.

Comparing employees in Marketing with those in Quality Control, Marketing employees seem to do a better job of choosing underpriced bets. However, they enjoy slightly lower net worth per employee. Turnover explains this difference. Employees in Quality Control tend to place more bets and, as I will explore in the next section, they tend to revise their bets more often than their counterparts in Marketing.

I construct "Timing" by dividing the days since inception that an employee places a bet by the total number of days that a forecast remains open. This measure suggests that on average employees in Development stop betting in forecasts before their colleagues in other groups. Employees in Publishing, though well-informed, tend to bet later than their peers.

Finally, I adapt the scoring technique introduced earlier to consider the marginal contributions of each group to the change in scores. For a subset of my data, I have changes in distributional parameters generated by each bet in the system. Where possible, I aggregate these changes by forecast and by group, and compute $S_{T,-j}$ for each group j. This is the final score at forecast suspension using all bets except those from group j. The difference $S_{T,-j} - S_T$, suggests the forecast improvement attributable to group j and I present this in Figure 2.12.

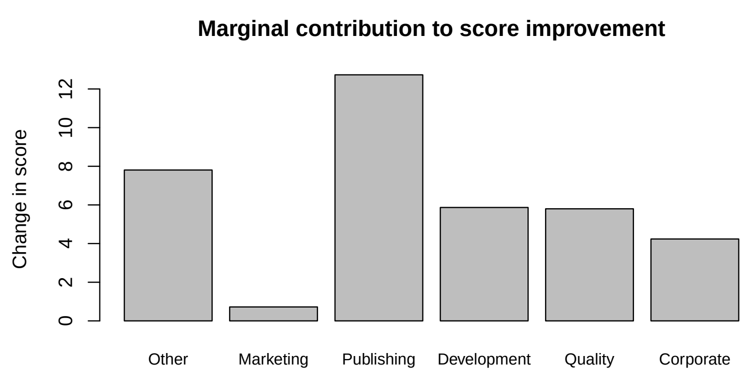


Figure 2.12: Contribution to reduction in expected losses, by group

This analysis gives a sense of who contributes private information to the betting market, but is subject to two important caveats. First, because the aggregation algorithm determines shifts in distributional parameters, this measure conflates the contributions of employees with the efficiency of the aggregation algorithm. Second, bets arrive in sequence but I compute the difference $S_{T,-j} - S_T$ without regard to transaction timing. This measure will not account for late participants who learn from those who went before.

2.6 Revisions and Information

I define *revisions* or "revision pairs" in my data as pairs of bets placed by an individual employee on a single random variable. In the analysis that follows, I will use revision activity to examine the dynamics of information arrival among market participants. I focus on revision pairs because the set of employees does not change from the first bet to the second. I pay particular attention to how revision activity differs across functional groups in the firm and the types of signals revealed over time. When I examine revisions in my data, I drop bets that were held for fewer than five-minutes. This results in 8182 pairs of bets. Revisions represent a nontrivial proportion of market activity, comprising 23% of bets submitted by traders. The sample includes activity by 56% of participants on 82% of forecasts.

My characterization of a single agent's strategy in Section 2.4 suggests that agents place bets in the market when their conditional beliefs about forecasts differ from consensus. An agent places a bet when her beliefs differ from consensus. When an agent exits a bet prior to close, she believes that the crowd's estimate of the expected value of her contract is at least as high as that justified by her private belief and that the region is no longer underpriced. If she places another bet in the same forecast, she must believe that the region she moves to is more underpriced than the region she moved from.

Full Sample

An implication of this argument is that bets agents revise *from* are overpriced relative to bets that agents revise *to*. This is testable in the data. I focus on the subsample of revisions that are held to close.

Hypothesis 1. *When employees move from one bet to another in a single forecast, the intervals they sell are less underpriced than the intervals they buy.*

I consider Hypothesis 1 a number of different ways. To begin, I compute the average probability assigned to bets ex-ante and compare this to the frequency of payoffs, ex-post. In my sample of revisions the crowd attached an average payoff probability of 41% to bets sold prior to revision. This set of bets contained the true value 47% of the time. By contrast, the crowd attached an average payoff probability of 42% to bets held after revision. These bets were 10% more likely to pay off, at 57%.

I capture the pure outperformance of bets placed second in a revision sequence by assembling a contingency table (Table IV.) Fisher's exact test strongly rejects ($p < 0.001$) the null of no difference between groups.

Finally, I obtain evidence in favor of Hypothesis 1 using a logit specification (Equation 2.23).

$$\mathbb{1}_{win,it} = \alpha + \beta_1 odds_{it} + \beta_2 \mathbb{1}_{second,it} + \beta_3[\mathbb{1}_{second,it} \times odds_{it}] + \varepsilon_{it} \tag{2.23}$$

I construct a dummy variable equal to one when the bet interval contains the outcome and zero otherwise. I regress this dummy on the payoff probability, a dummy variable indicating

	Win	Loss
First Bet	1620	1859
Second Bet	1989	1490

Table IV: Contingency table for sequences of bets. A bet is marked "Win" if the interval contained the outcome, and "Loss" otherwise.

whether this is the first or second bet in sequence, and an interaction between the second-bet dummy and the probability associated with the bet. I condition this specification on forecast, to account for systematic mispricing across forecasts.

The intercept, α, estimates the average crowd underpricing of bets in the sample of revisions. Under a null hypothesis of no difference in mispricing between the first and second bets in a sequence of revisions, the coefficient β_2 on the second-bet dummy variable should be zero. The conditional logit results strongly reject this ($p < 0.01$). The estimate $\hat{\beta}_2$ is positive and significant, suggesting systematic underpricing of second bets in revision pairs relative to first bets.

In a sequence of bets placed by the average employee on a single random variable, the crowd underprices the second bet in the sequence more than the first. Hypothesis 1 supports the idea that changes in relative information motivate revisions and contain private information about outcomes.

Revisions and Job Function

The next set of hypotheses focus on how revision activity and the accompanying information, flows vary across different functional groups within the firm. Do employees in different roles receive or acquire information about outcomes at different rates? This is a key question for any advocate of information markets within firms. If decision makers in the head office receive relatively precise signals at the same frequency as their subordinates, the benefits of eliciting information from employees may not justify the costs. The evidence I provide here will help identify sources of private information within the firm I study.

Does function within the firm have bearing on how employees' beliefs change? Employees who interact with products, customers and competing products on a regular basis likely receive or acquire information more quickly than those who do not. To the extent that revisions are consistent with information flow, I expect that agents closer to products to revise bets more often than those further away in marketing and management. I aggregate functions into two groups:

1. Close: Publishing, Quality Control and Development

2. Distant: Corporate (Management) and Marketing

Hypothesis 2. *Employees who enjoy a higher degree of contact with products revise their bets more frequently than employees whose roles in the firm are more distant from products.*

Figure 2.13 shows the distribution of revision pairs across functions within the firm, normalized by the number of employees in each group. While there is no substantial variation

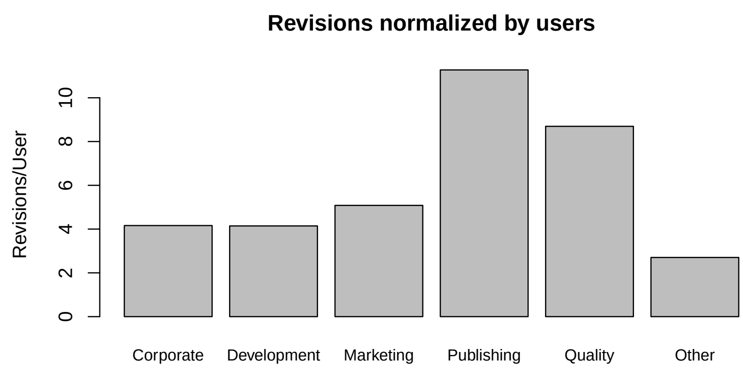


Figure 2.13: Revisions per employee, by functional group.

in the number of revisions per bet across groups, there is significant variation in the number of revisions per employee, with employees in Publishing and Quality Control placing more revisions than average. A χ^2-test of revision frequency by group confirms ($p < 0.01$) that the number of employees in a group does not explain the frequency with which members of a group revise bets. Moreover, a χ^2-test of revision frequency by Close and Distant ($p < 0.01$) suggests that the distance between employees and products might explain some variation in the propensity to revise bets.

In Section 2.6 I presented evidence that the second bet in a revision pair pays off with higher probability than the first bet in the pair. Since there is evidence that the frequency of revision varies significantly by group, a related question is whether these revisions contribute information of uniform quality. Again, I hypothesize that agents closest to products receive better information about product quality than agents further away from products.

Hypothesis 3. *Employees who enjoy a higher degree of contact with products buy bets that are more underpriced relative to the bets they sell.*

I test Hypothesis 3 with a logit regression similar to that employed in Section 2.6. In this case I append group dummies to estimate individual intercepts for each group. I focus on coefficients on interactions between group dummies and a dummy variable identifying the second bet in a revision pair. These estimates reveal the extent to which employees in each group choose second bets that are more underpriced than the first bet they entered. As before the crowd underprices bets if they pay off with higher frequency than the crowd predicts.

Tables Va and Vb contain the coefficient estimates on interaction terms from separate logit regressions. Employees in functions I define as Close and Distant appear indistinguish-

Variable	Coefficient	p
Second × Close	0.212	0.298
Second × Distant	0.134	0.598

(a) Close vs. Distant

Variable	Coefficient	p
Second × Marketing	0.153	0.601
Second × Publishing	0.662	0.008
Second × Development	0.132	0.546
Second × Quality	0.148	0.480
Second × Corporate	0.106	0.733

(b) By Functional Group

Table V: Logit regression, coefficients on group affiliation - second bet interaction.

able. However, at a finer level, I find that employees in Publishing do a significantly better job of exploiting mispricing in their revisions. In the firm I study, Publishing has sales and marketing functions but is regionally-based. Employees likely have relatively high contact with customers and the full set of the firm's products. This puts them in a good position to acquire private information and transmit it to the betting market.

2.7 Conclusions

This study examines data from a prediction market implementation. I start by characterizing the distribution of private information in my sample in a static sense. However, my data allows me to examine the dynamics of information within the firm as well. I focus on revisions - sequences of bets placed by a single employee on a single random variable. I show that these sequences appear to be generated by information flows. The set of bets that employees revise to appear more underpriced by the crowd than the set of bets that employees revise from.

Having shown that revisions contain private information, I sort these pairs of bets based on employee function within the firm. Doing so allows me to examine both revision activity and the performance of revisions at a lower level. I find evidence consistent with the idea that employees closest to products (Development, Quality Control and Publishing) revise more often than those further away. This evidence is consistent with my analysis of information in a static setting. However, I do not find that Close employees do a better job of selecting underpriced parts of the state space.

These features suggest that market mechanisms may outperform surveys in construction of forecasts within firms. Employees appear motivated to transmit private information by placing and revising bets. More importantly, my evidence suggests that, at least in the firm I study, management does not necessarily do a better job of betting than other agents in the firm. Private information, at least at this firm, does not come solely from management, rather it arises among employees at all levels.

This final point relates to corporate finance theory, where models often examine the implications of information asymmetries between managers and investors. Studying prediction market data suggests something about the source of these asymmetries. My data point out that private information can be dispersed among employees. The ability of managers to elicit and aggregate these bits of private information may deeply influence the extent of their informational advantage.

Chapter 3

Shifting Incentives in Forecasting Contests

3.1 Introduction

How do the structure of incentives affect the willingness of agents to faithfully communicate private information to a principal? In this paper, I study tournament incentives encountered by employees participating in a corporate prediction market. I explore how a change in rules governing payoffs to good forecasters alters strategies by participants and I analyze data produced by the market to detect these strategies.

The change in tournament incentives I study had two key features. First, management awarded prizes to fewer participants in each month of the prediction market game. Second, the value of prizes decreased across the board. I find evidence that causing participants to compete over a pool of fewer prizes increases the riskiness of their bets. However, I fail to find any measurable effects of lower-value prizes on participation.

The prediction market game I describe in Section 3.3 has the flavor of a forecasting contest. Participants were employees at a large software company asked to place bets on the outcome of various metrics, listed in Section 3.3. Management ranked employees on their performance in each month and awarded prizes based on their rank relative to others. Roughly, employees earned rewards for communicating precise and rare information to the market. Employees enjoy limited liability and face no penalties for poor performance. When management raises the threshold for earning prizes, employees have every incentive to increase the riskiness of their bets.

My results resemble the "exaggeration" of private information predicted by previous work by Ottaviani and Sorensen (2006). When agents earn rewards for making a report closest to the realization of a random variable and lose nothing for being distant, they have extra incentives to differentiate themselves more than they might in the absence of a tournament. In their model of a forecasting contest, the authors consider the symmetric equilibrium of a simultaneous-move game. In choosing her forecast, an agent faces a trade-off between informational efficiency and competition with other agents.

Outside of the context of tournaments, many authors have documented a bias in betting markets towards long-shots and away from favorites. Ali (1977), Snyder (1978) and Asch, et al (1982) use bets on horse-races to show that favorites tend to be underpriced relative to long-shots. Woodland and Woodland (1984), however, fail to observe this effect in baseball betting markets. Hodges, et al (2003) find evidence consistent with this bias in the market for index options.

Others have attempted to explain this bias towards low-probabilty events. Kahneman and Tversky (1979) suggest that bettors' may overestimate probabilities when payouts are high. Alternatively, they might derive utility directly from the chance of a big win, however small. More recent work by Bordalo et al (2011), models agents whose attraction to salient payoffs generates risk-seeking behavior.

I also interpret the change in incentives I exploit in this study in the context of managerial risk-taking. Work by Jensen and Meckling (1976), Gavish and Kalay (1983) and others consider how managerial incentives change with leverage. This work suggests that managers with levered claim on firm assets might exhibit risk-seeking in their investment decisions that is not necessarily in line with the interests of debt holders. Tournament incentives in the prediction market resemble a claim on prize payoffs similar to a call option. Shrinking the size of the prize pool means that earning a prize requires more winnings in a given month. In the prediction market game, I find a result that evokes asset substitution - employees ramp up the risk of their bets as managers might ramp up the risk of their investments.

Turning to incentives and participation, I find no significant effect of the change in prizes on activity in the betting market, either in the number of participants of the number of bets participants place or modify. This is surprising. Employees incur small non-monetary costs in time and cognition when they place bets. As rewards decrease after the change in incentives, I expect employees to spend less time and effort placing bets. I attribute this result either to the confounding effects of an adjustment in endowments made at the time of the change in incentives or to employees getting direct utility from game-play.

3.2 Model

In this section, I present a stylized model of a forecasting tournament to show the effects that underlie a testable hypotheses in the data relating betting behavior to a change in incentives. A principal sets up a forecasting game for a continuum of *competitors*. Competitors make forecasts of a random variable X on the real line. Nature chooses both the realization of X and the distribution of competitors' forecasts with density $f_M(m)$. I model the problem of an agent who places a forecast m_a. The agent stands to win a payoff π if her forecast is closer to the realization x than a proportion ρ of competitors. If the agent's forecast is further away from x than ρ competitors, she receives no payoff.

The agent knows $f_M(m)$ and receives a noisy signal of X, denoted s. The density $f_{X|s}(x|s)$ describes the agent's beliefs conditional on s. Both $f_{X|s}(x|s)$ and $f_M(m)$ are continuous. I assume the agent is risk-neutral and chooses m_a to maximize the probability of winning π. Begin by computing the value of forecast $V(m_a)$ conditional on x. Suppose the agent chooses $m_a > x$, she wins if the cumulative density of competitors' forecasts between $2x - m_a$ and m_a is less than ρ. Let $F_M(m)$ denote the cumulative density of competitors' forecasts:

$$V(m_a|x)|x < m_a = \begin{cases} \pi & \text{if } F_M(m_a) - F_M(2x - m_a) < \rho \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

At $x = m_a$, $F_M(m_a) - F_M(2x - m_a) = 0 < \rho$ and the agent wins. Further, by choosing a small x, I can push $F_M(2x - m_a)$ arbitrarily close to zero. So long as $F_M(m_a) > \rho$ and $F_M(\cdot)$ is continuous, there exists a threshold value $\underline{x}_a$ where $F_M(m_a) - F_M(2\underline{x}_a - m_a) = \rho$. This implies that $V(m_a|x) = \pi$ for $x \in [\underline{x}_a, m_a]$ and zero otherwise. The case where $x > m_a$ is symmetric with:

$$V(m_a|x)|x > m_a = \begin{cases} \pi & \text{if } F_M(2x - m_a) - F_M(m_a) < \rho \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

By a similar argument, so long as $F_M(m_a) < 1 - \rho$, $\exists$ a value $\bar{x}_a$ such that $F_M(2\bar{x}_a - m_a) - F_M(m_a) = \rho$. This implies that $V(m_a|x) = \pi$ for $x \in [m_a \bar{x}_a]$ and zero otherwise.

Using these arguments, I compute the following conditional expectations:

$$E[V(m_a)|x < m_a] = \pi[F_{X|s}(m_a|s) - F_{X|s}(\underline{x}_a|s)] \tag{3.3}$$
$$E[V(m_a)|x > m_a] = \pi[F_{X|s}(\bar{x}_a|s) - F_{X|s}(m_a|s)] \tag{3.4}$$

and express the unconditional expected value as:

$$E[V(m_a)] = \pi\{F_{X|s}(m_a|s)[F_{X|s}(m_a|s) - F_{X|s}(\underline{x}_a|s)] + [1 - F_{X|s}(m_a|s)][F_{X|s}(\bar{x}_a|s) - F_{X|s}(m_a|s)]\} \tag{3.5}$$

While the expected value function is fairly intuitive, rather than trying to compute an analytical solution, I provide numerical results. To motivate the hypotheses that follow in Section 3.4, I focus on on the impact of changing the proportion of competitors that receive a payoff. For the calibration, I begin by specializing this general structure with normal distributions. The agent's beliefs conditional on s are normal $X|s \sim \mathcal{N}(\mu_a, \sigma_a^2)$ and Nature chooses competing forecasts that are also normal: $M \sim \mathcal{N}(\mu_m, \sigma_m^2)$.

Figure 3.1 displays the payoff probability as a function of the agent's forecast. The x-axis represents the set of possible forecasts m_a, such that $F_M(m_a) \in (\rho, 1 - \rho)$. In this simulation, the agent's conditional beliefs have a higher mean and are less dispersed than the distribution of competitors' forecasts. The effects of tournament incentives are evident in the figure. The expected-payoff maximizing forecast (4.45) is greater than the agent's conditional mean. The agent biases his forecast away from the mass of competitors.

This "exaggeration" result resembles the result of Ottaviani and Sorensen's (2006) model of a winner-take-all forecasting contest and occurs for the same reason. When locating

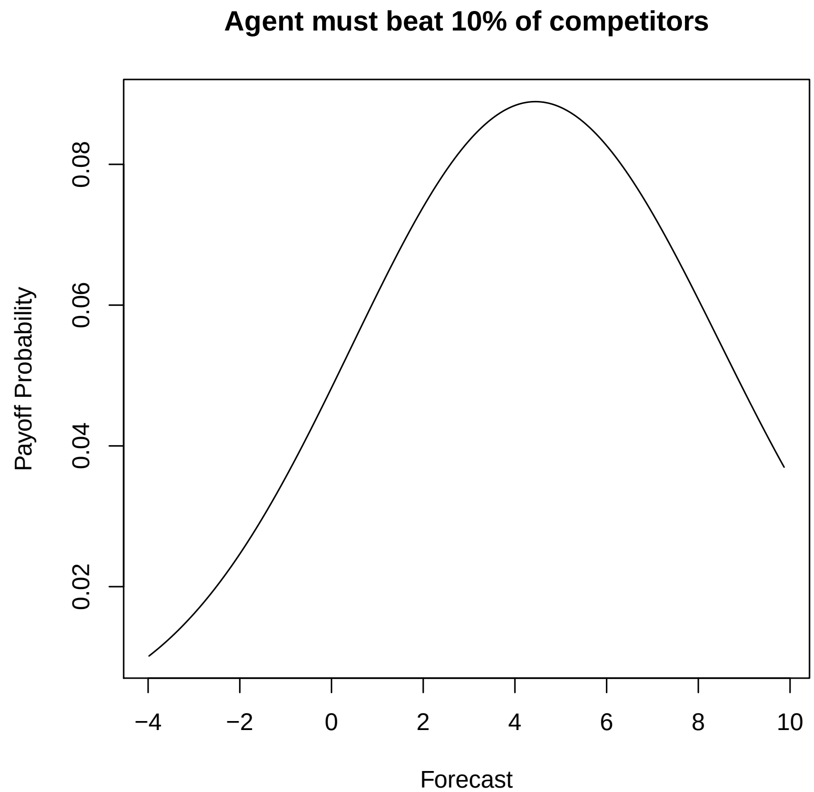


Figure 3.1: Payoff probabilities associated with possible forecasts. $M \sim \mathcal{N}(3, 30)$, $X|s \sim \mathcal{N}(4, 10)$, $\rho = 0.10$

her forecast, the agent faces a trade off between choosing values close to her conditional expectation of X and choosing values that set her apart from the crowd. For the agent depicted in Figure 3.1, making a forecast of 4 would bring her closer to the realization of X but she would win over a smaller set of states, since a larger mass of competitors lie in the vicinity of 4 relative to her optimal forecast of 4.45.

A key difference between this model and the winner-take-all contest is that I can examine what happens when the standards for winning a payoff change. Suppose an adjustment to the incentive structure wherein the agent must place her forecast within the best 5% of competitors forecasts in order to earn a payoff. How will her best response change? Figure 3.2 shows the results of a calibration exercise identical to that shown in Figure 3.1, only with a smaller proportion ρ. Again, the x-axis represents the set of forecasts m_a such that $F_M(m_a) \in (\rho, 1-\rho)$. The agent further exaggerates her forecast, this time choosing 4.47. Allocating fewer payoffs to the population of forecasters makes the agent more sensitive to

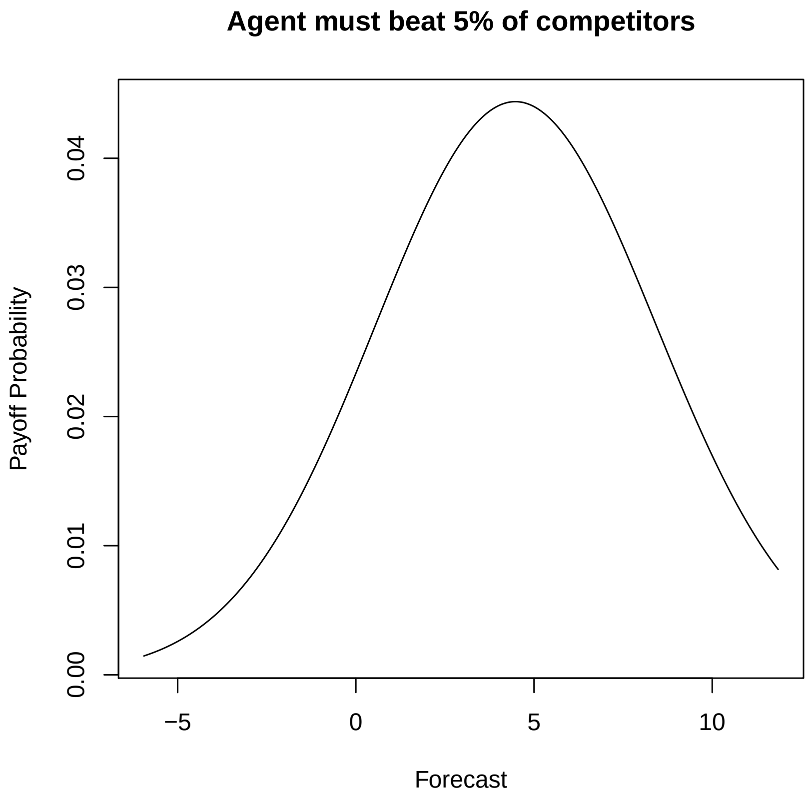


Figure 3.2: Payoff probabilities associated with possible forecasts. $M \sim \mathcal{N}(3, 30)$, $X|s \sim \mathcal{N}(4, 10)$, $\rho = 0.05$

the mass of competitors and causes her to intensify her exaggeration.

I term the change in incentives depicted above a "narrowing" of the incentive base. Lemma 4 summarizes the impact that such a narrowing has on the agent's behavior.

Lemma 4. *A narrowing of the incentive base causes the agent to exaggerate her best response.*

3.3 Data

This paper uses the same basic data as Phatak (2012) which comes from a single, ongoing implementation of a prediction market within a software firm. The sample contains bets placed by employees between 20 March, 2009 and 9 March, 2011.

Market Design

Employees access the betting market using a computer interface. The market is composed of a set of *forecasts*, chosen at the discretion of management. Employees log onto the application and enter with an endowment of *credits*, or currency within the system. Credits have three important features. First, the site administrator chooses the initial endowment of credits allocated to employees who participate. Second, employees may not transfer credits between one another. Finally, employees may not redeem credits for cash; the site administrator redeems credits for prizes according to a *prizing schedule*.

Figure 2.1 contains a menu of variables on which employees in a fictional betting market might place bets. On entering the market, an employee may bet in any of four forecasts. Each of these forecasts corresponds to a random variable whose realization the firm wishes to predict. An employee may believe she has information about the number of new product units ordered by customers for July delivery. The computer system allows her to click a link and place a bet on this particular variable.

An employee who clicks to bet on the number of units ordered for July delivery sees a screen depicting the aggregate beliefs of the *crowd* of other participants in the form of a distribution function. These beliefs are formed from (1) a prior chosen by the market administrator; and (2) signals inferred from all preceding bets. Figure 2.2 shows the crowd density for units ordered for July delivery.

In this prediction market game, participants place *bets* on continuous variables by specifying a bet *interval* and a number of credits to wager. In Figure 2.2 an employee has selected an interval between 4.2 million units and 5.5 million units. If the firm experiences unit sales within the interval specified by the agent, she will receive a *promised payoff* in credits.

In this example, the promised payoff associated with the employee's bet is approximately 20 times the bet amount (Figure 2.3), corresponding to the crowd's belief that the likelihood of realized unit sales within the bet interval is about 5%. Had the employee instead chosen an interval carrying 10% probability in the crowd's estimate, then the promised payoff would decrease to 10 times the wager.

Once she submits a bet to the market the employee receives a *contract* specifying a bet interval, a wager amount and the promised payoff. The system distributes payoffs once forecasts close and a measurement of the underlying variables becomes available. In this example, if the realization of unit sales is outside of the user's bet interval, she loses her stake. If her bet pays off, then she receives her promised payment of 18728 credits.

Management determines the length of time that forecasts remain open. So long as a forecast is open and not yet suspended, employees may withdraw bets or revise them. Users may hold one bet in a forecast at a time.

Between trades, prices reset as a result of activity according to an updating algorithm. Only transactions move prices. When an employee submits a bet through the interface, an algorithm estimates parameters of the employee's signal from from the interval and wager size. The algorithm infers the location of an employees beliefs based on the location of the bet interval and measures signal precision using the size of the wager and the width of the bet interval.

This conforms to intuition. High signal precision implies less dispersion in conditional beliefs and a more peaked posterior distribution. This is consistent with a contraction in bet intervals. A more peaked distribution results in a larger degree of perceived underpricing and hence a higher quantity demanded. I will return to this in Section 3.2 when I examine how information leads employees to place and revise bets. Once the software has imputed the signal that produced an employee's bet, her signal is added to those of other bettors in the system.

The market provides incentives by awarding prizes. Employees receive a ranking based on the number of credits they earn during each month. These rankings map to Amazon gift cards of different denominations.

Sample

I base my study on a sample of bet data taken from a prediction market consisting of transactions over a two-year period. The plurality (130) of questions asked by management concerned product quality scores. Management also asked employees to help forecast unit sales, dollar sales, ship dates, growth rates and market share. I provide a detailed breakdown of forecasts by topic in Table I.

The shortest forecast lasted one week from inception to close while the longest lasted for one year. The average forecast lasted for 114 days from inception to close. Forecasts suspended prior to close. The period between forecast suspension and realization of the underlying random variable was typically between one week and one month.

868 employees registered to participate in the prediction market. Of these registrants, 598 chose to place at least one bet in the system. Conditional on placing a bet in the system, employees placed 46 bets, including revisions, on average. Participation was highly variable, with a standard deviation of 77. Figure 2.6 is a histogram of the number of bets, including revisions, placed by employees in my sample.

To the transactions data from the betting market, I merge information about incentives. Recall that participants in the market placed bets denominated in credits which they could not transfer or redeem for cash. Management ranked employees based on their earnings from month to month and mapped rankings to prizes according to a prizing schedule. Table II contains the prizing schedules in effect during my period of observation. For nineteen months from June, 2009 to January, 2001, I obtained detailed prize information. This includes identifying information for all users who received prizes along with their rank in that month. For many months, I also have information about the change in credits on which management based rankings.

How difficult was it to win prizes in this environment? Of the 593 employees who placed a bet in the market, 389 won some sort of prize. Table I partitions these users by ranking and givens an indication of turnover among the "elites" of this prediction market. In my sample of 19 prizing periods, there were 190 prizes awarded in Tier 1 with an aggregate value

Table I: Unique appearing in each prize tier.

	# of Users	# of Prizes	Aggregate Value ($)
Tier 1	75	190	17000
Tier 2	142	300	7100
Tier 3	356	1070	10700

Table II: Gini coefficients for subsamples of prizing data.

(a) Gini coefficients for all employees who placed bets.

	Full Sample	High Slope	Low Slope
Credits	0.798	0.687	0.959
Prizes	0.768	0.732	0.911

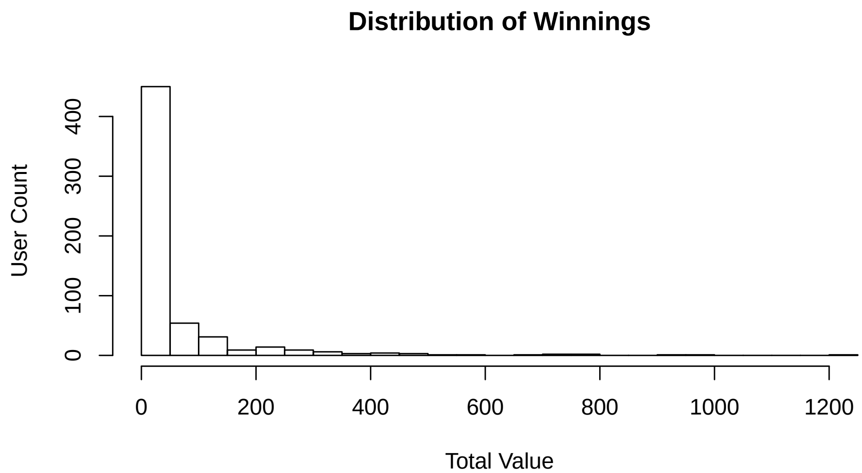
(b) Gini coefficients for all prize winners.

	Full Sample	High Slope	Low Slope
Credits	0.807	0.764	0.836
Prizes	0.661	0.632	0.552

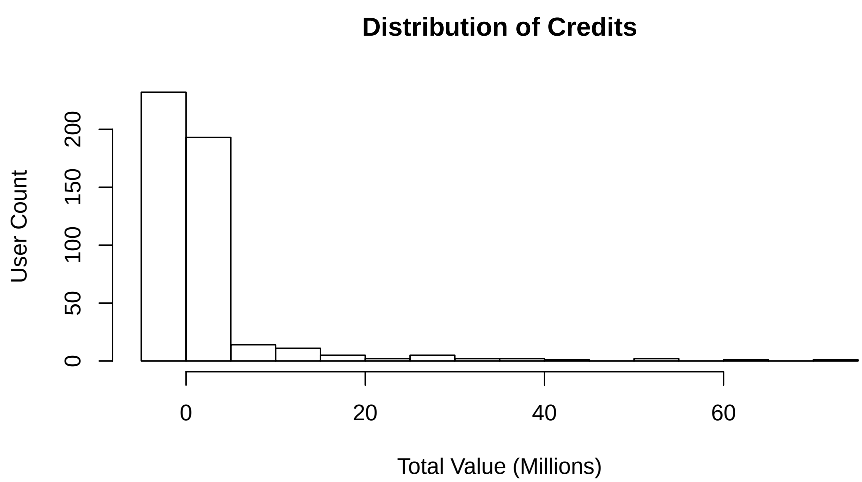
of $17000. However, only 75 employees shared this pot. Tournament incentives produced a highly skewed distribution of winnings, as presented in Figure 3.3a.

Compare this to Figure 3.3b, where I plot the distribution of credits earned by agents in the betting market. Gini coefficients provide a second way to examine the relationship between credits and prizes. I compute this statistic (Table II) for the full sample, the "high-slope" period ending in May 2010 and the low-slope period from May 2010 to the end of my sample.

Table IIa contains Gini coefficients estimated on the full set of employees who placed bets in each period. Based on these estimates, inequality in prizes resembles inequality in credits. One possible explanation is that employee's responses to tournament incentives in the prediction market are moderated by their beliefs about their relative ability to earn credits. Those who believe they have a low chance of winning anything put low effort into placing bets and earn fewer credits than their more motivated colleagues. This generates a skew in the distribution of credits closely resembling the skew in the distribution of winnings.



(a) Distribution of total cash winnings.



(b) Distribution of credits

Figure 3.3: Inequality in prizes and credits

3.4 Results

The intuition underlying the model I presented in Section 3.2 has a few implications testable in the prediction market data I collected. In May 2010, prize values decreased and management began distributing half the number of prizes (Table II). A reduction in prize amounts likely reduced incentives to participate. I expect fewer employees to place bets after the change in incentives. Moreover, I expect employees who choose to participate to bet less frequently than before.

I test whether or not employees' betting behavior changes as a result of the change in incentives. Awarding 50 prizes rather than 100, could cause more intense competition among employees, ultimately resulting in much riskier bets. I also attempt to measure the information content of bets before and after the change in incentives. Did increased competition encourage more risk-taking without any more information acquisition? I hope to understand whether changing the payoffs to bets changes the informational efficiency of the betting market.

Participation

Did the adjustment to incentives have any effect on participation in the betting market game? Though the prediction market software was simple to use and internet-accessible, employees still faced costs in time and cognition when placing bets. To the extent that the change in prizes reduced the marginal benefits of placing bets, I expect to observe fewer users placing bets and fewer bets placed, conditional on participation.

Hypothesis 4. *Fewer employees participate in forecasts when the magnitude of incentives decrease.*

A natural way to measure employee participation is by the number of unique users who participate in each forecast. Normalizing by available forecasts helps account for the fact each additional forecast represents a new opportunity for gain. I measure participation as opening or cashing out a bet. I compare the number of employees who participate in each forecast, per day, before and after the change in incentives.

Casual inspection suggests that participation diminishes through time, perhaps as the novelty of the betting game wears off, or as uninformed employees learn how poorly they perform relative to their peers. As a result, I choose to test this hypothesis using a linear model attempts to identify the effect of the incentive change separately from the time trend in participation:

$$p_{ut} = \alpha + \beta_1 t + \beta_2 \mathbb{1}_{low-slope,t} + \beta_3 \mathbb{1}_{low-slope,t} t + \varepsilon_t \tag{3.6}$$

where p_u is the number of unique employees, per question, per day and t counts the number of days from the beginning of the sample. Under the null hypothesis of no change in participation, $\beta_2 = 0$. I fail to reject the null hypothesis, suggesting no significant difference in the number of employees participating in forecasts each day. I provide coefficient estimates in Table III.

	Estimate	Std. Error	t value	Pr(>|t|)
$\hat{\alpha}$	0.5853	0.0329	17.82	0.0000
$\hat{\beta}_1$	-0.0011	0.0001	-8.37	0.0000
$\hat{\beta}_2$	0.1846	0.1194	1.55	0.1225
$\hat{\beta}_3$	0.0002	0.0002	0.92	0.3571

Table III: Results of estimating sensitivity of users per forecast per day, to days from inception and an indicator variable for the change in incentive structure.

Another implication of lowering benefits to betting is that, conditional on participation, employees are less active in the prediction market.

Hypothesis 5. *Employees place and modify bets in forecasts less often when the magnitude of incentives decrease.*

Here, I face the same problem of a general downward trend in user activity, per question per day. I estimate a version of Equation 3.6 where I replace p_{ut} with a measure of the number of bets placed or changed, per question, in each day. As before, $\beta_2 = 0$ implies

	Estimate	Std. Error	t value	Pr(>|t|)
$\hat{\alpha}$	2.1369	0.1609	13.28	0.0000
$\hat{\beta}_1$	-0.0032	0.0007	-4.76	0.0000
$\hat{\beta}_2$	2.5072	0.5873	4.27	0.0000
$\hat{\beta}_3$	-0.0029	0.0012	-2.39	0.0172

Table IV: Results of estimating sensitivity of bets per forecast per day, to days from inception and an indicator variable for the change in incentive structure.

no change in the average participation level before and after the change in incentives. The estimate, $\hat{\beta}_2$ (Table IV), is positive and significant, suggesting that not only did this measure of participation not decrease, it actually increased with a decrease in prize amounts. This is surprising and possibly due to a “reset” of the market at the same time as the change in the prizing schedule. In the course of the reset, participants with net worth less than 500000 credits received a grant and those with net worth above 500000 sustained a withdrawal of credits from their accounts. At the time, most employees were below the threshold and received a grant of credits. An attempt by these employees to deploy their new wealth might explain the increase in activity I detect.

Betting Strategy

The change in prize amounts was relatively small. The top 10 employees lost \$25 each, the next 10 lost only \$5 each and employees ranking 21-30 each lost \$15. Of larger magnitude was the fact that 50 fewer employees received prizes in each period. I expect the increase in competition to place in the top 50 earners to outweigh the effect of small changes in rewards for the top 50 employees. In particular, I expect employees to take more long-shot bets as the number of prizes shrinks.

Hypothesis 6. *Probabilities associated with bets during the high-slope period are lower than those during the low-slope period.*

A key identifying assumption is that employees were no more informed at the end of my sample than during the beginning. Without this condition, I might observe lower odds after the change in incentives because employees possessed more precise private information leading them to choose more underpriced bets. I return to this question with Hypothesis 7.

I test Hypothesis 6 with a regression of bet probabilities on an intercept and an indicator variable to condition on whether an employee places a bet before or after the change in incentives. Because probabilities are bounded, I employ a logit transformation on the dependent variable:

$$l(P) = \log\left(\frac{P}{1-P}\right) \tag{3.7}$$

and estimate:

$$l(P) = \alpha + \beta \mathbb{I}_{low-slope} + \varepsilon \tag{3.8}$$

where $\mathbb{I}_{low-slope} = 1$ if an employee placed the bet in a low-slope period and zero otherwise. I estimate a model with fixed effects to account for unobserved heterogeneity in employees and priors for the random variables available for betting. I also limit attention to bets placed by users who The coefficient β measures the change in average probabilities resulting from the change in incentives. Under the null, $\beta = 0$. The estimate, $\hat{\beta}$, is negative and significant using forecast-level fixed effects (Table Va). Including employee-level fixed effects weakens the result (Table Vb).

Note however, that I report two-tailed probabilities for a one-tailed null hypothesis. Moreover, my results strengthen significantly if I limit my analysis to the 22-month period surrounding the change. This window excludes employees who stopped participating very early on and might not behave as those with more experience under both prizing schedules. On balance, the evidence from this regression indicates that bet probabilities decreased with the change in incentives. Under the assumption that employees were no more informed after

	Estimate	Std. Error	t value	Pr(>|t|)
$\hat{\alpha}$	-0.8063	0.0740	-10.90	0.0000
$\hat{\beta}$	-0.1345	0.0336	-4.00	0.0000

(a) Forecast-level fixed effects

	Estimate	Std. Error	t value	Pr(>|t|)
$\hat{\alpha}$	-0.6018	0.2767	-2.18	0.0296
$\hat{\beta}$	-0.0510	0.0301	-1.69	0.0903

(b) Forecast- and employee-level fixed effects

Table V: Fixed-effects regression of payoff probabilities on an indicator variable identifying the period post incentive change.

the the change in incentives, employees responded to the reduction in the number of prizes by placing riskier bets.

The model I describe in Section 3.2 suggests that the change in forecasting behavior stemming from a change in the incentive structure should yield more extreme forecasts that are no more informative of the outcome variable. In my data, this means that the long-shot bets I observe after the incentive change should be no more underpriced than corresponding bets prior to the change.

Hypothesis 7. *High-risk bets placed by employees responding to the change in incentives are no more likely to yield a positive payoff.*

I test Hypothesis 7 using a logit specification similar to Phatak (2012). I estimate:

$$\mathbb{1}_{win,it} = \alpha + \beta_1 l(P_{it}) + \beta_2 \mathbb{1}_{low-slope,t} + \beta_3 [\mathbb{1}_{low-slope,t} l(P_{it})] + \varepsilon_{it} \tag{3.9}$$

I construct a dummy variable, $\mathbb{1}_{win,it}$ equal to one when the bet interval contains the outcome and zero otherwise. I regress this dummy on $l(P_{it})$, the logit of the payoff probability (Equation 3.8), a dummy variable indicating whether an employee placed the bet before or after the change in incentives, and an interaction between this dummy and the probability associated with the bet. I provide coefficient estimates in Table VI. I fail to reject the null hypothesis that $\beta_2 = 0$. There is no evidence that the lower-probability bets selected by employees after the change in incentives are any more underpriced than those selected before the change. In other words, the change in odds I observe in testing Hypothesis 6 is not the result of more precise private information.

	Estimate	Std. Error	z value	Pr(>|z|)
$\hat{\alpha}$	0.1494	0.0179	8.34	0.0000
$\hat{\beta}_1$	0.7381	0.0175	42.25	0.0000
$\hat{\beta}_2$	0.0266	0.0280	0.95	0.3430
$\hat{\beta}_3$	-0.2248	0.0236	-9.51	0.0000

Table VI: Estimating the sensitivity of payoff probability to an indicator variable identifying whether bets occurred after the incentive change.

3.5 Conclusions

I study an exogenous shock to tournament incentives in a prediction market game to examine effects on risk-taking and participation. My results are mixed. I find significant and interesting effects on the propensity of participants to make long-shot bets, but no real impact on the frequency with which they participate in the betting game.

Confronted with an incentive structure that required employees to be in the top 50 performers, rather than the top 100, to win prizes, employees responded by increasing risk-taking. They exaggerated their bets, consistent with prior theoretical work on forecasting contests by other authors, in an attempt to differentiate themselves from their peers. I confirm that the additional risk-taking I observe is not associated with more precise private information about the variables on which employees place bets.

This work has implications for the design of forecasting contests. A principal who elicits information from agents motivated by tournament incentives may, in theory, still efficiently aggregate signals that agents communicate. However, efficient aggregation requires adjusting for the structure of prizes. As mentioned in the introduction, this study also touches the literature on managerial incentives in corporate finance. I provide empirical evidence that the participants in my prediction market game respond to the option-like payoffs associated with tournament prizing by taking on riskier bets.

However, a reduction in the size of prizes did not have the expected effect on participation in the betting market. Future work will consider possible explanations for this. A change in net worth accompanied the change in incentives, which might have encouraged credit-rich participants to enter the market and search for new wagers. Alternatively, employee activity might be sensitive to both the scope of incentives and their size. Reducing the number of prizes awarded in each period might have caused participants to be more vigilant about acquiring information or placing bets in the system. This effect could have dominated any discouragement arising from lower payoffs. Understanding and separately identifying the effects of reducing the size and scope of incentives remains a topic for future work.

Appendix A

Supplement for Menu-Based Complexity

A.1 Summary Statistics

Total	2606
Alternatives	1229
Order	777

Table I: Unique user interactions by treatment. I log an interaction whenever a subject logs into one of the trial servers.

Unique Workers	1660
One Trial	1180
Two Trials	331
Three Trials	68
Four Trials	45
Average Participation	1.47

Table II: Participation by Workers recorded by MTurk. Some workers participated in multiple trials. I use data on acceptances and submissions to describe how often workers in my sample participated across trials.

	Accept to Login	Login to Bet	Bet to Submit	Total
Alternatives	17.06	81.90	19.72	118.68
Order	16.52	83.43	19.45	119.40

Table III: Average sub-task durations for Workers recorded on MTurk. These are the average number of seconds workers spent on each of three activities in the experiment.

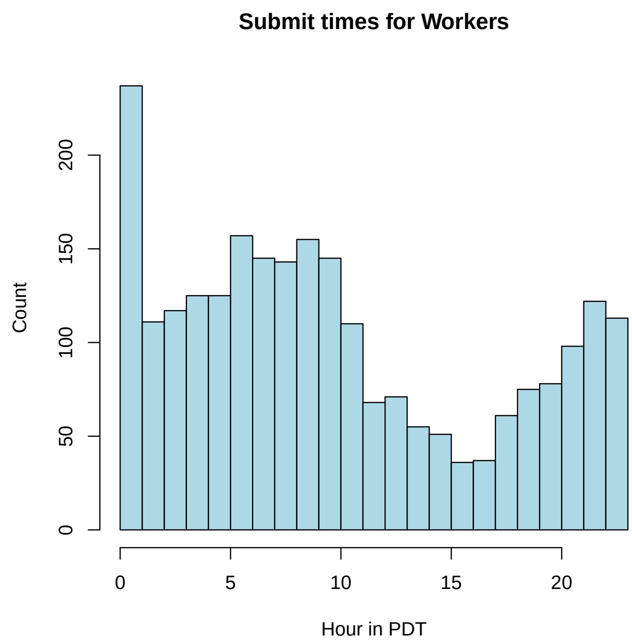


Figure A.1: Submission hours, in Pacific Daylight Time

A.2 Amazon Mechanical Turk

Workers and Requesters populate the MTurk environment. Requesters post "Human Intelligence Tasks" (HITs). Each HIT is a collection of assignments for workers to complete. Along with a task description, Requesters post a price they are willing to pay for each assignment in a HIT. Requesters also specify an expiration and a maximum duration per assignment.

Workers are anonymous and identified only with a unique Worker ID. Workers choose from a large selection of HITs varying in task complexity and price. If a worker accepts a given HIT he must complete and submit the task within the duration specified by the Requester. Workers are free to return assignments mid-task for no reward.

The Requester receives the completed task and may approve or reject submitted work. If the Requester approves a task, Amazon debits the Requester's account and credits the

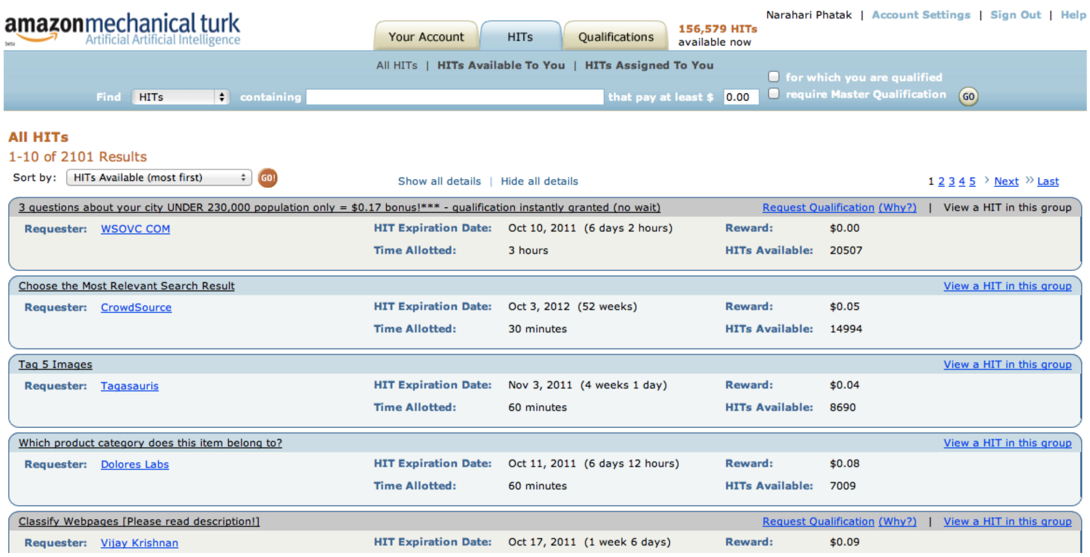


Figure A.2: Menu of tasks available on MTurk

Worker's account. If the Requester rejects a submission, no transfer takes place. Rejections are recorded to a Worker's account. If the Requester approves work, he may also grant the Worker a bonus of any amount.

Example of Interaction

An interaction on MTurk begins with a subject selecting a task I post from the set of different tasks available to her on MTurk (Figure A.2). Upon selecting my task, I present the subject with a consent form that she is able to read completely before proceeding (Figure A.3). A subject who chooses not to participate at this stage, returns to the menu of tasks without incurring any penalty.

If a subject clicks the "Accept HIT" button, I direct her to either a landing page with a link to a survey question or to the survey question itself, as depicted in Figure A.4. In this particular case, having read the task description and information on the right panel, the subject would solve the problem by choosing the interval width she wants, by clicking the appropriate radio button. She has information on how to make this choice. In this case, an expected payoff maximizer would evaluate the expected payoff of the five-cent lottery as \$0.05, the expected payoff of the ten-cent lottery as \$.0.055, and so on. The ten-cent lottery is optimal in this set.

Once the subject selects a lottery using the radio buttons, I enable a "Place Bet" button that submits information about the selection to the server. The subject now has the option

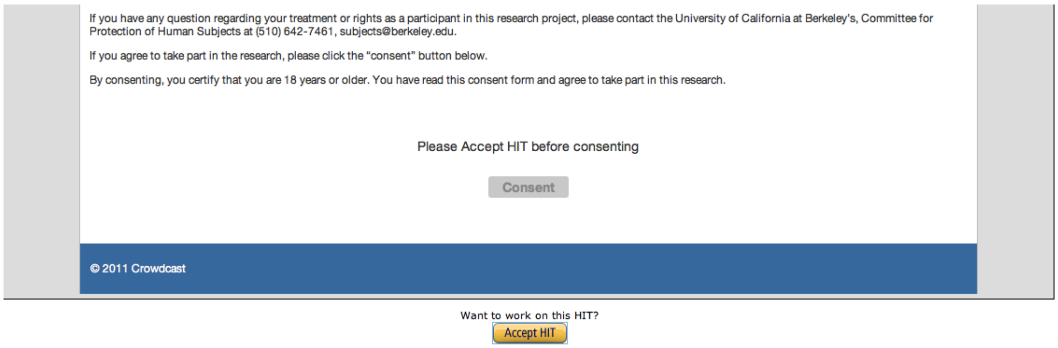


Figure A.3: Consent form and "Accept HIT" button

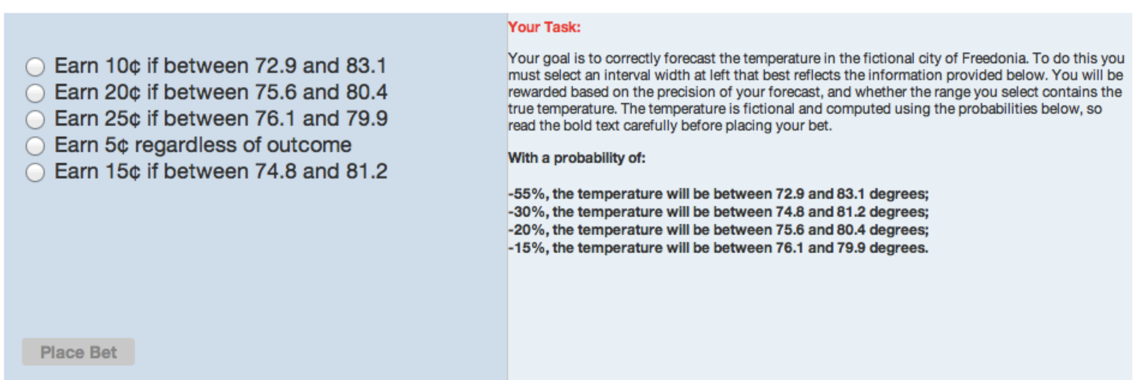


Figure A.4: Game screen after bet location entry

of revising her bet and resubmitting if she chooses (Figure A.5). Alternatively, she can click a "Submit HIT" button that transmits a confirmation of task completion to Amazon and ends her interaction with the experiment. At the conclusion of each trial, I randomly awarded payoffs to participants, consistent with the probabilities in their prompts and the payoffs they selected.

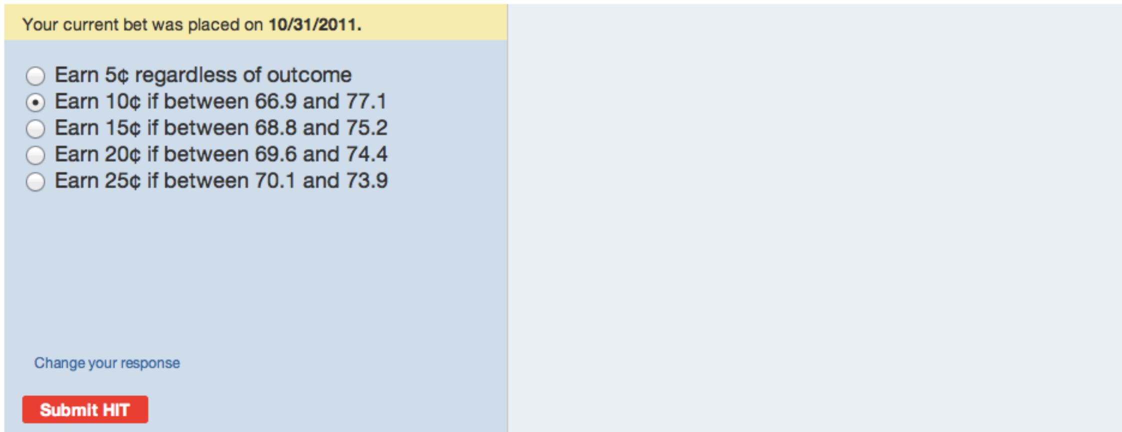


Figure A.5: Game screen before HIT submission

A.3 Inference from Interval Bets

This section details how I construct the menus of interval bets that I offer subjects in my treatments. In particular, it shows why subjects' best response reveals both their signal realization the the precision of their private information. I also explain how I calibrate the data I present in Tables XIII and XX.

Constructing Menus of Bets

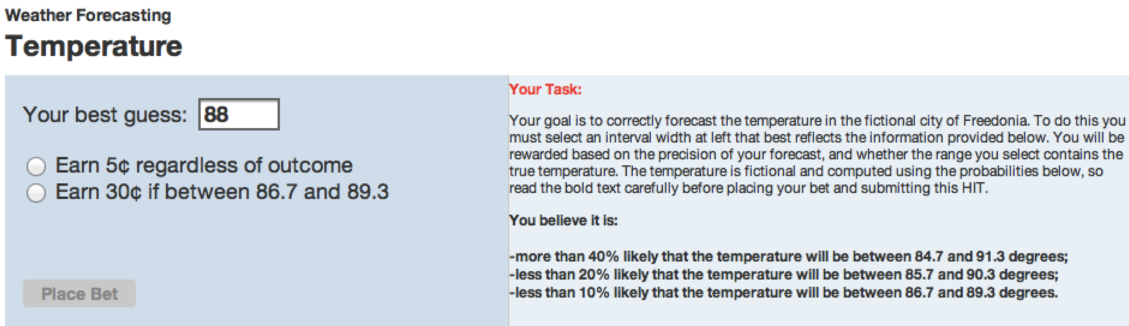
To construct menus, I have three controls at my disposal:

1. The distribution of interval widths offered for bets;
2. The set of payoffs, one for each interval width;
3. The number of widths offered to subjects.

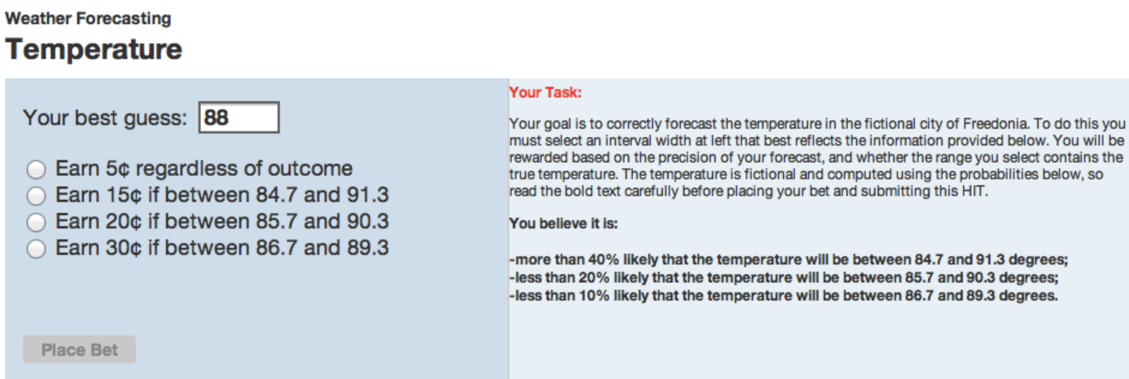
The number of widths represents the key difference between treatments in experiments on alternative-based complexity. I choose the corresponding set of payoffs to conform roughly to the MTurk environment. Base pay for subjects was either five or ten cents so I chose lottery payoffs of the same order of magnitude.

To illustrate, suppose the random variable $X \sim \mathcal{N}(72, 10)$ and this distribution is common knowledge. An agent of type i receives a private signal of the form:

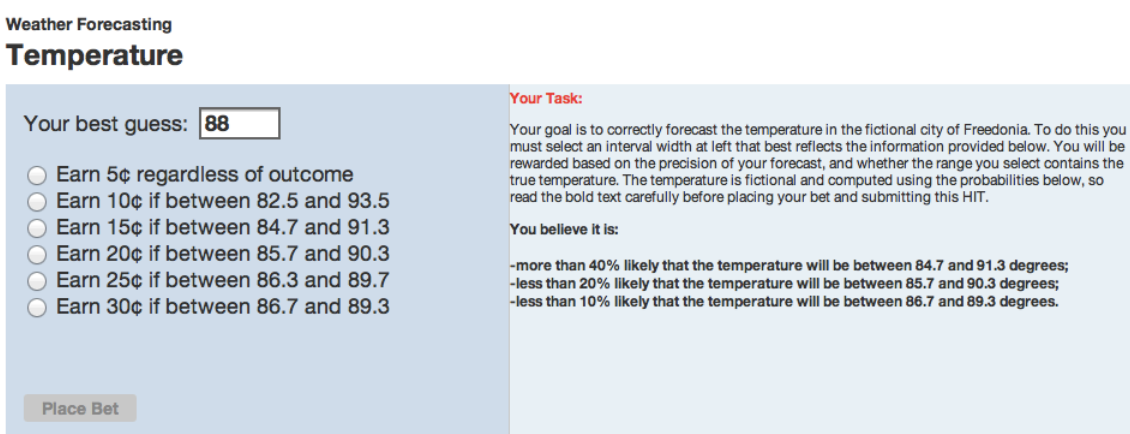
$$y_i = x + \varepsilon_i$$



(a) SHORT



(b) MEDIUM

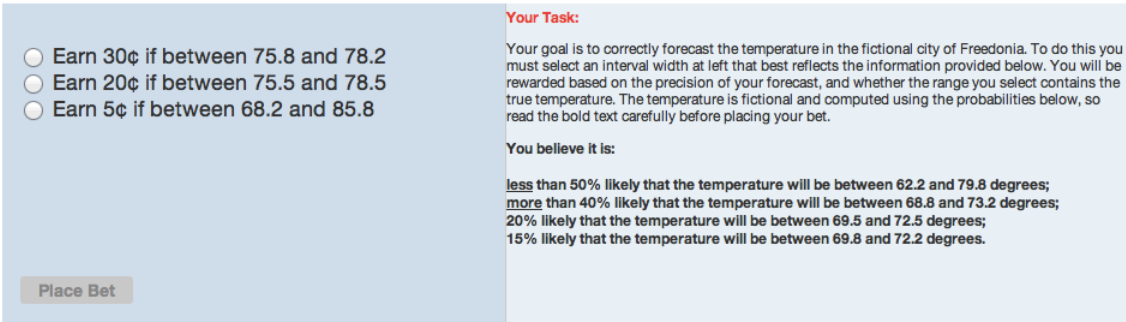


(c) LONG

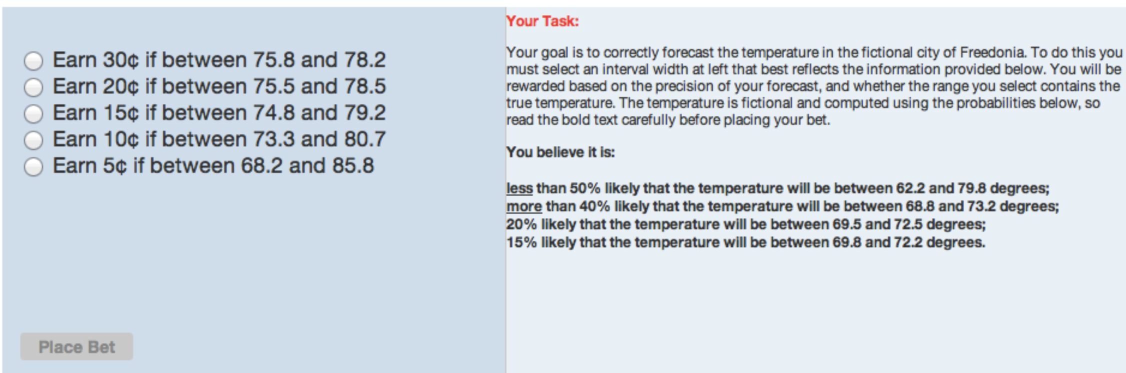
Figure A.6: Alternative-Based Complexity and Choice, version A

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_i)$. In this example, $i \in \{H, L\}$, where $\sigma_H = 10$ and $\sigma_L = 20$. If one agent receives $y_H = 68$ and another receives $y_L = 62$ then their posterior beliefs are:

$$X|\{y_H = 68\} \sim \mathcal{N}(70, \sqrt{50})$$
$$X|\{y_L = 62\} \sim \mathcal{N}(70, \sqrt{80})$$



(a) SHORT



(b) MEDIUM

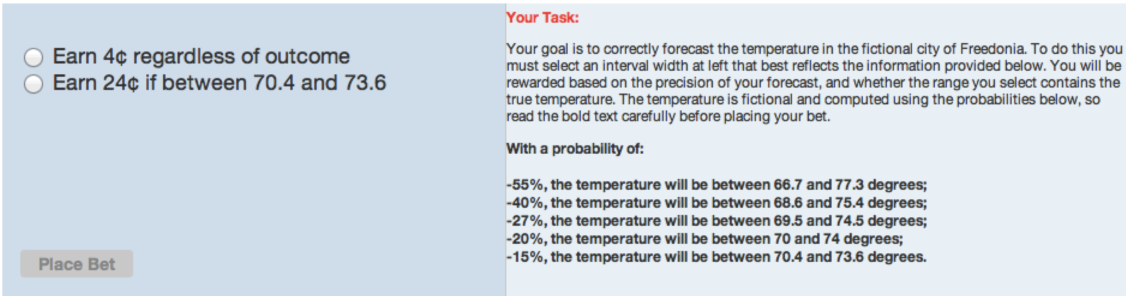
Figure A.7: Alternative-Based Complexity, version B

These two agents have conditional distributions centered at precisely the same point and are indistinguishable from one another on the basis of a best guess (Figure A.11). Let $w_i(m)$ be an interval of width $w_i, i \in \{H, L\}$, centered at m. An agent of type i assigns probability $p_i(w_i(m))$ to the event that X is realized on the interval $w_i(m)$. Let T_i be a contingent payoff associated with the width w_i. Incentive compatibility (IC) in this context means choosing payoffs/width pairs such that:

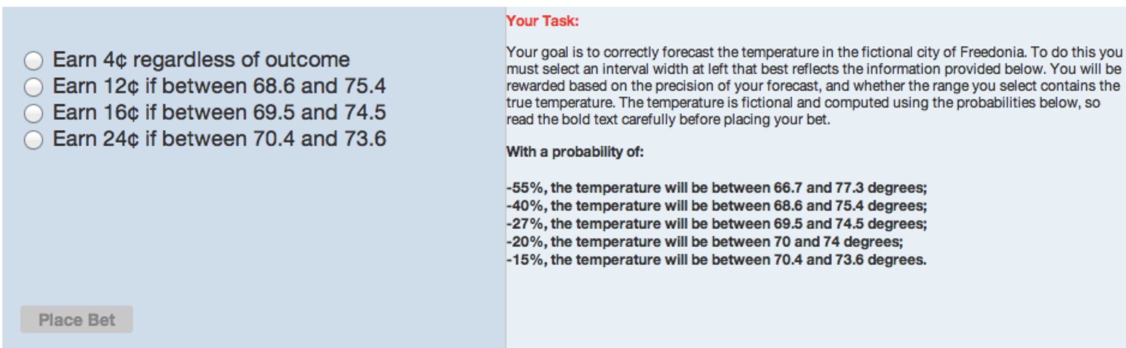
$$\begin{aligned} p_H(w_H(m))T_H &\geq p_H(w_L(m))T_L \\ p_L(w_L(m))T_L &\geq p_L(w_H(m))T_H \end{aligned}$$

To separate the two types, I choose widths, corresponding to payoffs, that enforce IC. Suppose payoffs are $T_L = 5$ and $T_H = 10$. A natural way to choose widths is to find a width that causes the low type's participation constraint to bind. If the low type's option is to leave with a risk-free payoff of 3, then:

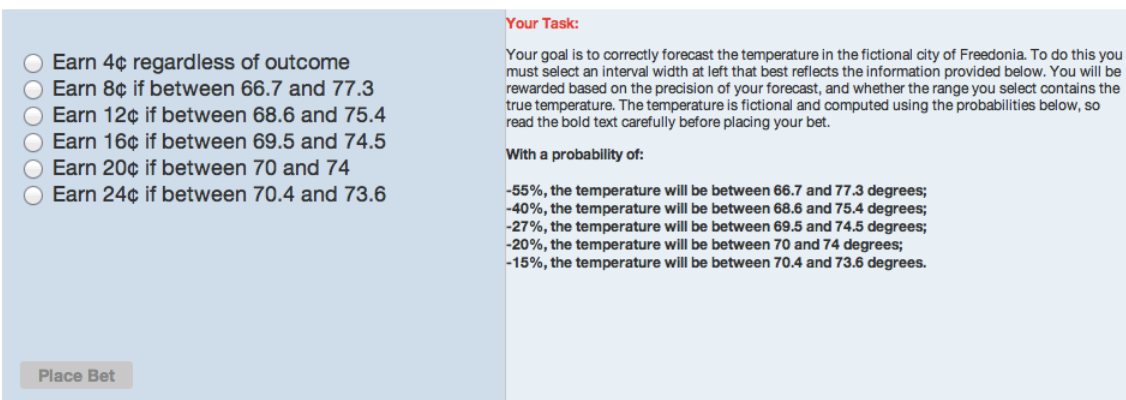
$$w_L(m) = p_L^{-1}\left(\frac{3}{5}\right)$$



(a) SHORT



(b) MEDIUM

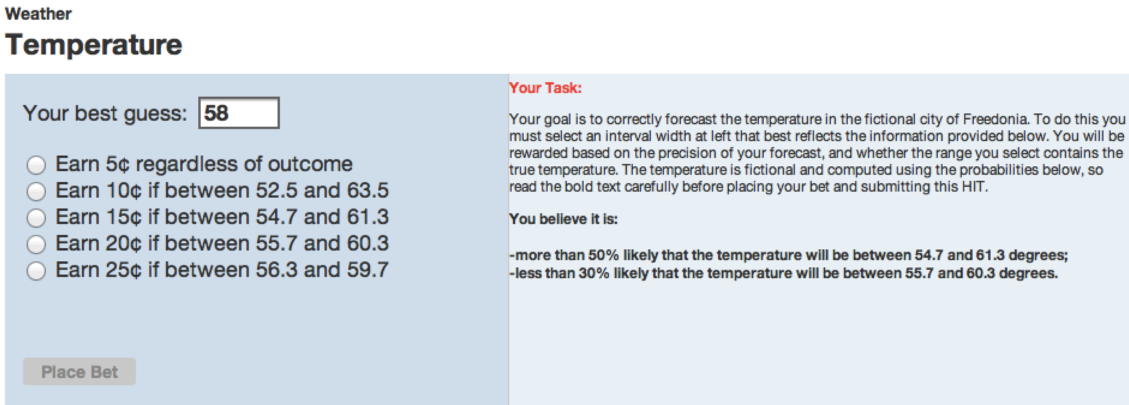


(c) LONG

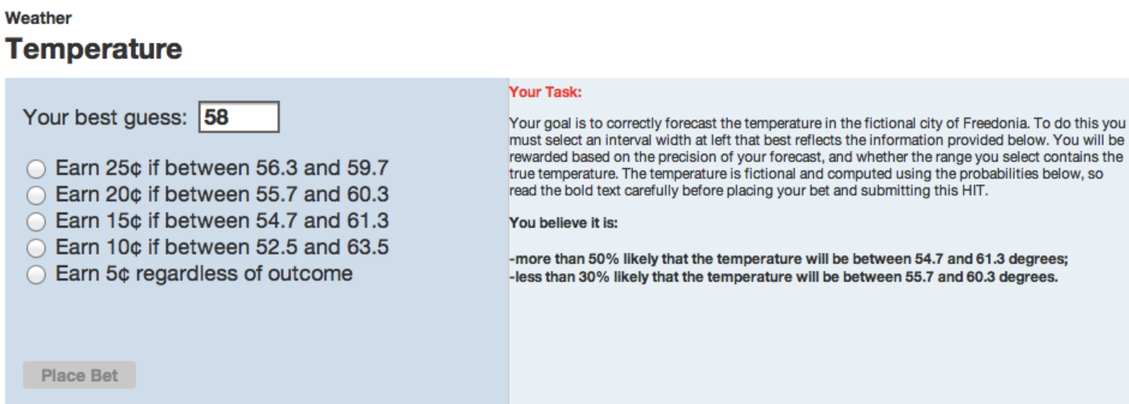
Figure A.8: Alternative-Based Complexity, version C

To force the high-type's IC constraint to bind requires:

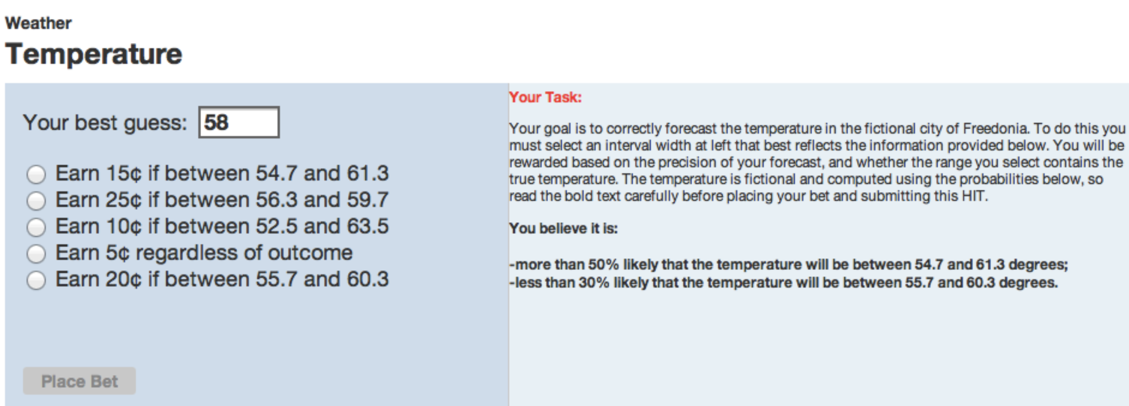
$$w_H(m) = p_H^{-1}\left(\frac{p_H(w_L(m))}{2}\right)$$



(a) ASCEND

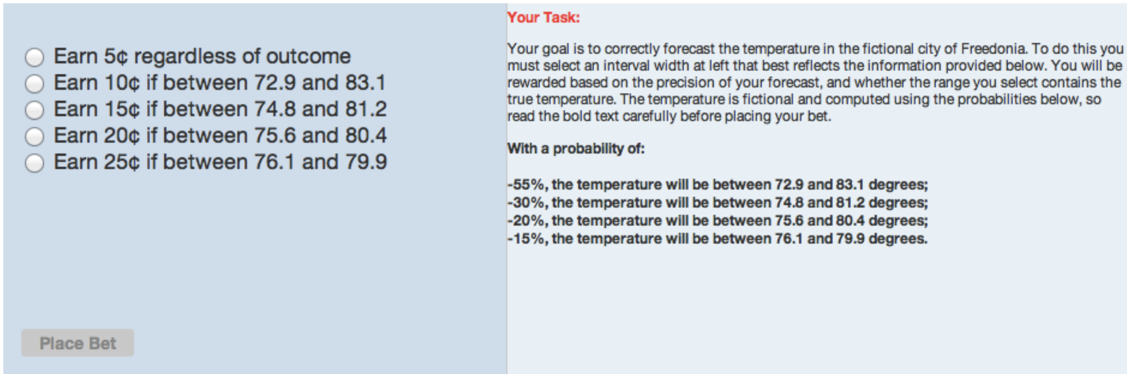


(b) DESCEND

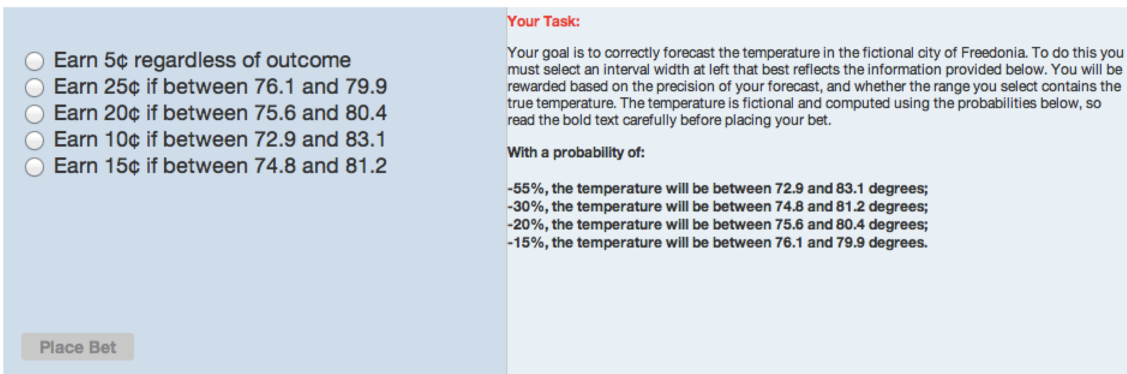


(c) GARBLE

Figure A.9: Menu Order and Choice, version F



(a) ASCEND



(b) GARBLE

Figure A.10: Menu Order and Choice, version G

which immediately implies a narrower interval for the high type. I solve for these interval widths numerically using the assumptions about conditional distributions laid out above:

$$w_L = 15.06$$
$$w_H = 6.54$$

Inference and Calibration

In Tables XIII and XX, I present the results of thought experiments that relate the choices made by subjects in the MTurk forecasting game to the conditional variance of forecasts made by a bookmaker who observes these choices. If I construct interval widths as described in section A.3, then each width-payoff pair maps to a threshold signal precision - the minimum level of precision required to make a lottery dominate all less-risky alternatives in expected payoff terms.

I construct forecasts using the data based gathered from subjects in each treatment. Due to random assignment, the number of participants assigned to each treatment is different,

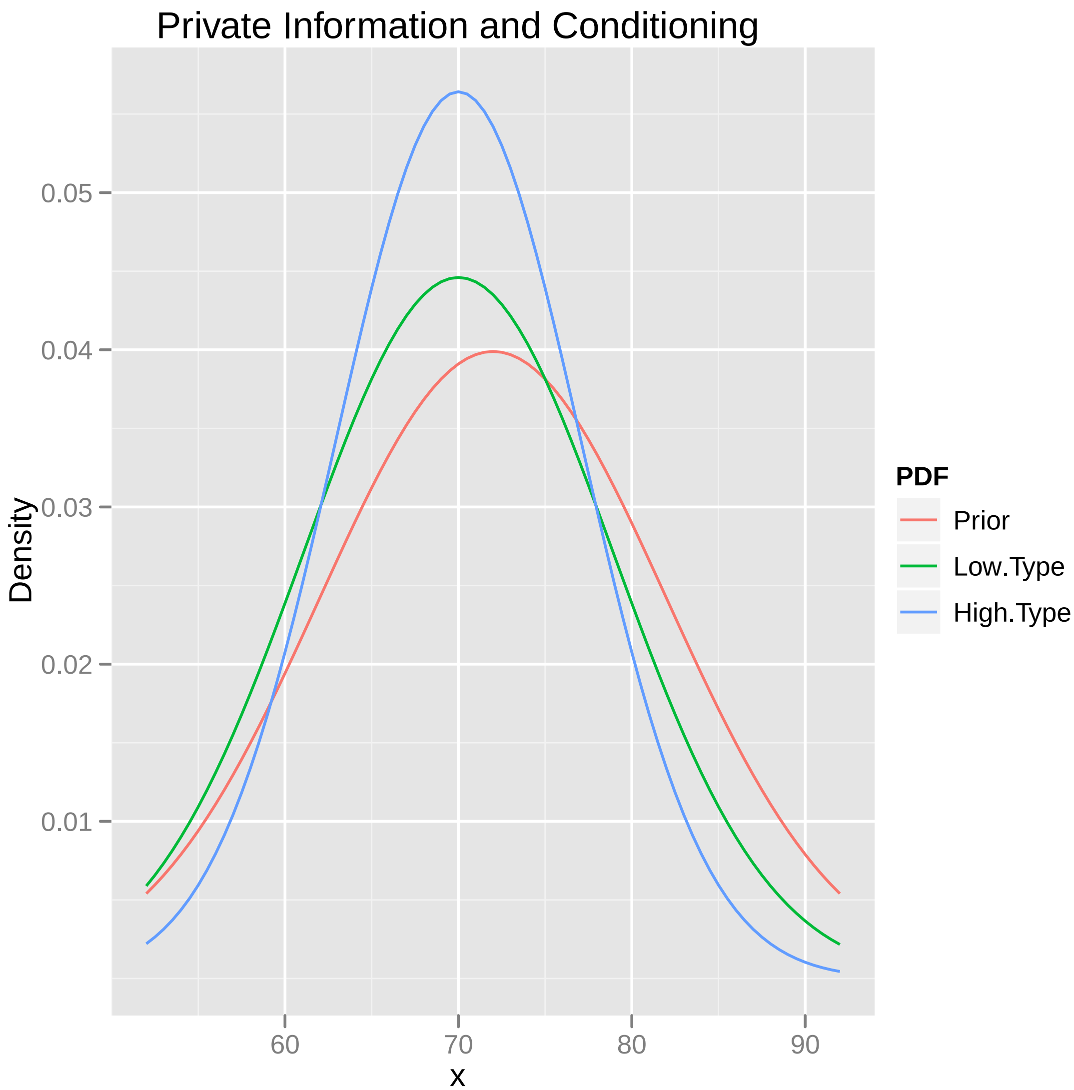


Figure A.11: Conditional densities of high- and low-type agents

	Lottery Payoffs					
Payoff	4	8	12	16	20	24
Precision	0.000	0.006	0.009	0.012	0.015	0.018
SMALL	100					0
MEDIUM	0		100	0		0
LARGE	0	0	100	0	0	0

Table IV: Selections under a null hypothesis of efficient choice by all subjects

	Lottery Payoffs					
Treatment	5	10	15	20	25	30
Precision	0.000	0.006	0.009	0.012	0.015	0.018
SMALL	28					72
MEDIUM	10		54	16		20
LARGE	6	28	19	28	11	8

Table V: Observed selections normalized to 100 subjects per treatment

so I normalize the number of participants to 100. For both the benchmark and observed forecasts, I assume:

1. Subjects come in a finite set of types based on their signal precision and the set of types is common knowledge.
2. If the bookmaker cannot distinguish between types based on a bet, he assumes a uniform distribution over types for whom the bet is incentive-compatible.

Example: Alternative-Based Complexity and Choice

Consider the treatments described in Table VII. In the benchmark case I assume optimal behavior by all agents, with 100 agents assigned to each treatment (Table IV). I set the prior precision, $\nu = 0.01$ and apply a normal updating model. For bets $i \in 1...N$, conditional variance is:

$$\frac{1}{\nu + \sum_{i=1}^{N} \tau_i}$$

Table V takes the observed count data from version C and normalizes them so that 100 subjects encounter each menu. I repeat the updating process described above using these normalized count data. Using observations from LONG, conditional variance is 1.01. The degree to which the subjects underestimate variance in LONG relative to their benchmark is $1 - \frac{1.01}{1.10} = 8.20\%$

A.4 Alternative Measures of Disorganization

One way I identify complexity is in the effect of menu ordering on sensitivity to expected payoffs, as evidenced by subjects' choices. I estimate a conditional logit model that includes $shuffle_i$ an indicator variable that takes a value of 1 if subject i observed a menu that is disordered and zero otherwise. Can I gain further insight into the effects of disorder by using a measure of disorder that captures the extent to which a menu is disorganized?

Distance to Benchmark

One way to measure disorganization is to look at how an ordering deviates from a benchmark. I define the benchmark as ascending order, represented by the vector $b = (1, 2, 3, 4, 5)$ and the menu ordering of an arbitrary menu as $\omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5)$, then let:

$$d_{euc}(\omega) = \|b - \omega\| \tag{A.1}$$

where $\|\cdot\|$ represents the Euclidean norm. The value of d_{euc} increases as the menu order deviates from the benchmark. A menu that ascends in order of payoffs would set $d_{euc}(b) = 0$. However, this measure may not capture disorder in an appropriate way. Consider $\bar{\omega} = (5, 4, 3, 2, 1)$, a menu with payoffs in descending order. This menu maximizes the total deviation from b. Why should I consider a descending menu any more disorganized than an ascending menu?

Distance Between Adjacent Alternatives

A second measure of disorganization considers a menu more disorganized the more distant are adjacent alternatives. For a menu with N alternatives, I define:

$$d_{adj}(\omega) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(\omega_{i+1} - \omega_i)^2} \tag{A.2}$$

This measure is immune to the criticism I level at d_{euc}. For a menu in descending order, $d_{adj}(\bar{\omega}) = d_{adj}(b) = 0$. On this measure, a descending menu is no more complex than an ascending menu since the increments between alternative payoffs are equal.

This menu $(3, 2, 5, 1, 4)$ is disordered because there are large gaps between each adjacent pair of alternatives. I argue that these gaps make it difficult for subjects to match probabilities to payoffs and compute expected values. Even when subjects compute expected payoffs, the gaps make ranking alternatives based on expected values confusing.

VARIABLES	choice
$\hat{\beta}$	0.386***
	(0.146)
$\hat{\gamma}$	-0.069*
	(0.039)
$\hat{\delta}_2$	0.204
	(0.134)
$\hat{\delta}_3$	0.235*
	(0.141)
$\hat{\delta}_4$	-0.028
	(0.167)
$\hat{\delta}_5$	-0.353*
	(0.192)
Observations	2,045
df	6

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table VI: Conditional logit results for disordered menus (N = 409); γ is the loading on d_{euc}.

Estimation

To assess the effect of menu disorganization with more nuance, I replace $shuffle_i$ in (1.2) with the measures $d_{euc,i}$ and $d_{adj,i}$. If these measures reflect the fact that certain disorganized menus are more difficult to interpret than others then I should observe negative loadings on interaction terms between these measures and expected payoffs with the magnitude of the interaction effect increasing in the degree of disorder.

Moreover, each measure of disorganization I propose carries a slightly different interpretation of order. A difference in effects across specifications may provide more precise information about why disordered menus reduce subjects' sensitivity to expected payoffs.

I consider the measures separately, starting with distance to benchmark. I estimate:

$$U_{ij} = \beta ev_j + \gamma d_{euc,i} ev_j + \delta_j position_j(\tilde{j}) + \varepsilon_{ij} \tag{A.3}$$

I present the results of this model in Table (VI). The loading on the interaction $d_{euc,i}ev_j$ is negative and significant at the 10% level, providing weak evidence that menu disorder, as measured using deviations from a vector representing ascending payoffs, reduces sensitivity to expected payoffs.

To get a sense of the economic significance of these results, I compute the marginal effect of expected payoffs on selection probability for the *average* alternative. The average

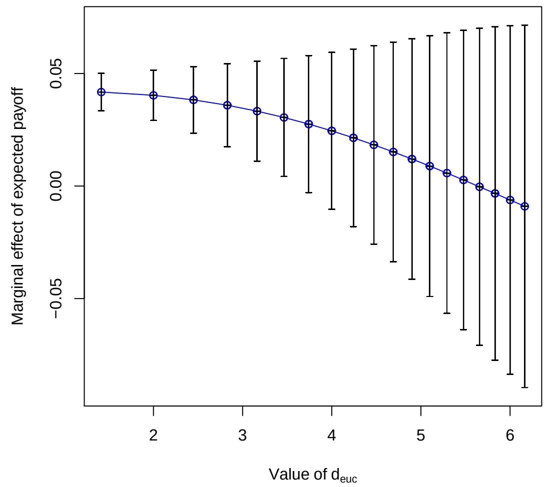


Figure A.12: Marginal effects for disordered menus. On the x-axis is d_{euc}, a measure of distance from the ascending menu benchmark. Bars cover a 95% confidence interval.

alternative carries an expected payoff equal to the average expected payoff of the alternatives on the menu (4.55) and occupies the middle position on the menu. When $d_{euc} = 0$, or the menu is ordered, the marginal effect of an increase in expected payoffs is 4.08%. When I set $d_{euc} = 4.28$, its average value for shuffled menus, the marginal effect of expected payoffs on selection probabilities is nearly halved to 2.09% and loses statistical significance. In Figure (A.12), I show the marginal effect of expected payoffs for positive values of d_{euc}. For values of d_{euc} greater than four, the marginal effect of expected payoffs on selection probabilities for the average alternative is indistinguishable from zero.

I present the estimated interaction effect for different value of d_{euc} in Figure (A.13). This shows that at higher levels of disorganization, d_{euc} exerts negative pressure on subjects' sensitivity to expected payoffs. However, even at the highest levels of d_{euc}, the interaction effect is not significant. I conjecture that this is due to fact that high values of d_{euc} are associated with descending menus. Despite being different from the benchmark ascending menu, these menus are still organized in a way that facilitates easy comparison of alternatives.

Now I turn to the second proxy for disorder, d_{adj} and estimate:

$$U_{ij} = \beta ev_j + \gamma d_{adj,i} ev_j + \delta_j position_j(\tilde{j}) + \varepsilon_{ij} \tag{A.4}$$

I present the full results of this estimation in Table (VII). As above, the loading on expected payoffs, β, is significant. This model also significantly loads on the interaction term $d_{adj,i}ev_j$ ($p < 0.01$). The marginal effect of expected payoffs on selection probability for the average

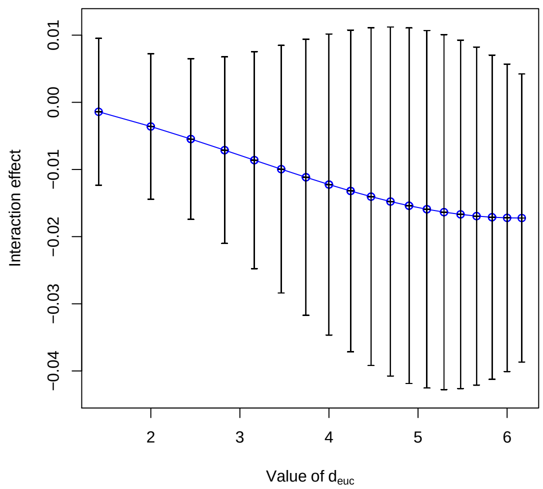


Figure A.13: Interaction effect for disordered menus. On the x-axis is d_{euc}, a measure of distance from the ascending menu benchmark. Bars cover a 95% confidence interval.

alternative on an ordered menu is 3.54%. This effect drops to 0.81% for a shuffled menu with a average value of d_{adj}.

Figures (A.14) and (A.15) present the marginal effects of expected payoff and interaction effects associated with d_{adj}, respectively. Comparing the interaction effects across models in Figures (A.13) and (A.15) suggests that d_{adj} more accurately proxies for the effects of disorder across different shuffled menus. For high levels of d_{adj}, the interaction effect is significant and negative. At high levels of disorganization, a small increase in complexity, as measured by d_{adj} significantly reduces subjects' sensitivity to expected payoffs.

VARIABLES	choice
$\hat{\beta}$	0.488***
	(0.139)
$\hat{\gamma}$	-0.185***
	(0.067)
$\hat{\delta}_2$	0.161
	(0.136)
$\hat{\delta}_3$	0.240*
	(0.138)
$\hat{\delta}_4$	-0.112
	(0.159)
$\hat{\delta}_5$	-0.335*
	(0.180)
Observations	2,045
df	6

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table VII: Conditional logit results for disordered menus (N = 409). γ is the loading on d_{adj}.

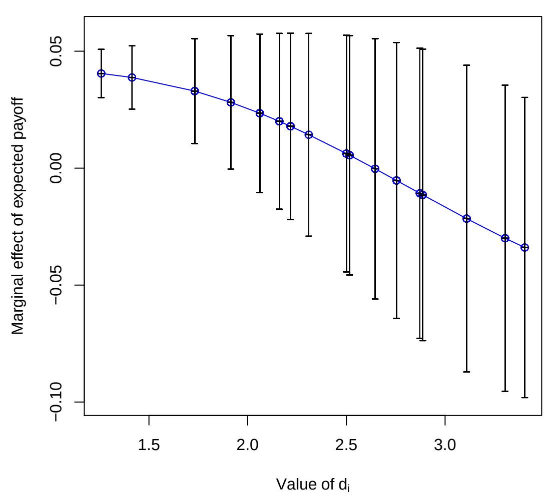


Figure A.14: Marginal effects for disordered menus. On the x-axis is d_{adj}, a measure of payoff variability for adjacent alternatives. Bars cover a 95% confidence interval.

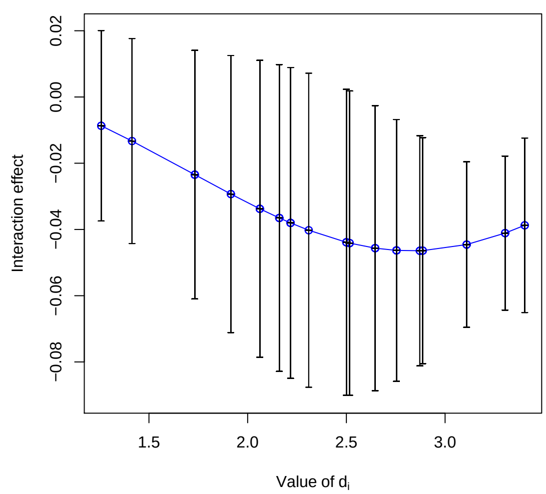


Figure A.15: Interaction effect for disordered menus. On the x-axis is d_{adj}, a measure of payoff variability for adjacent alternatives. Bars cover a 95% confidence interval.

Bibliography

[1] Mukhtar M. Ali. "Probability and Utility Estimates for Racetrack Bettors". English. In: *Journal of Political Economy* 85.4 (1977), pp. 803–815. ISSN: 00223808. URL: http://www.jstor.org/stable/1828704.

[2] Franklin Allen. "Stock markets and resource allocation". In: ed. by Colin Mayer and Xavier Vives. Center of Economic Policy Research, 1995. Chap. 4, pp. 81–108.

[3] Peter Asch, Burton G. Malkiel, and Richard E. Quandt. "Racetrack betting and informed behavior". In: *Journal of Financial Economics* 10.2 (1982), pp. 187 –194. ISSN: 0304-405X. DOI: 10.1016/0304-405X(82)90012-5. URL: http://www.sciencedirect.com/science/article/pii/0304405X82900125.

[4] Ulf Axelson. "Security Design with Investor Private Information". English. In: *The Journal of Finance* 62.6 (1993), pp. 25872632. ISSN: 00221082. URL: http://www.jstor.org/stable/2329041.

[5] Shlomo Benartzi and Richard H. Thaler. "Naive Diversification Strategies in Defined Contribution Saving Plans". English. In: *The American Economic Review* 91.1 (2001), pp. 79–98. ISSN: 00028282. URL: http://www.jstor.org/stable/2677899.

[6] J.E. Berg et al. "Results from a dozen years of election futures markets research". In: *Handbook of Experimental Economic Results* 1 (2001), pp. 486–515.

[7] Joyce E. Berg and Thomas A. Rietz. "Prediction Markets as Decision Support Systems". In: *Information Systems Frontiers* 5 (1 2003). 10.1023/A:1022002107255, pp. 79–93. ISSN: 1387-3326. URL: http://dx.doi.org/10.1023/A:1022002107255.

[8] Philip Bond and Itay Goldstein. "Government Intervention and Information Aggregation by Prices". English. In: *SSRN eLibrary* (2010).

[9] Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. *Salience Theory of Choice Under Risk*. Working Paper 16387. National Bureau of Economic Research, 2010. URL: http://www.nber.org/papers/w16387.

[10] Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. "Amazon's Mechanical Turk". In: *Perspectives on Psychological Science* 6.1 (2011), pp. 3–5. DOI: 10.1177/1745691610393980. eprint: http://pps.sagepub.com/content/6/1/3.full.pdf+html. URL: http://pps.sagepub.com/content/6/1/3.abstract.

[11] Bruce I. Carlin. "Strategic price complexity in retail financial markets". In: *Journal of Financial Economics* 91.3 (2009), pp. 278 –287. ISSN: 0304-405X. DOI: 10.1016/j.jfineco.2008.05.002. URL: http://www.sciencedirect.com/science/article/pii/S0304405X08002092.

[12] Bruce I. Carlin, Simon Gervais, and Gustavo Manso. "Libertarian Paternalism, Information Sharing, and Financial Decision-Making". English. In: *SSRN eLibrary* (2010).

[13] Bruce I. Carlin and Shimon Kogan. "Trading Complex Assets". English. In: *SSRN eLibrary* (2010).

[14] Bruce Ian Carlin and Gustavo Manso. "Obfuscation, Learning, and the Evolution of Investor Sophistication". In: *Review of Financial Studies* 24.3 (2011), pp. 754–785. DOI: 10.1093/rfs/hhq070. eprint: http://rfs.oxfordjournals.org/content/24/3/754.full.pdf+html. URL: http://rfs.oxfordjournals.org/content/24/3/754.abstract.

[15] Alexander Chernev. "When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice". English. In: *Journal of Consumer Research* 30.2 (2003), pp. 170–183. ISSN: 00935301. URL: http://www.jstor.org/stable/10.1086/376808.

[16] Bo Cowgill, Justin Wolfers, and Eric Zitzewitz. *Using Prediction Markets to Track Information Flows: Evidence from Google*. Working Paper. 2008.

[17] S. Figlewski. "Subjective information and market efficiency in a betting market". In: *The Journal of Political Economy* (1979), pp. 75–88.

[18] Bezalel Gavish and Avner Kalay. "On the Asset Substitution Problem". English. In: *The Journal of Financial and Quantitative Analysis* 18.1 (1983), pp. 21–30. ISSN: 00221090. URL: http://www.jstor.org/stable/2330802.

[19] F. A. Hayek. "The Use of Knowledge in Society". English. In: *The American Economic Review* 35.4 (1945), pp. 519–530. ISSN: 00028282. URL: http://www.jstor.org/stable/1809376.

[20] David Hirshleifer. "Investor Psychology and Asset Pricing". English. In: *The Journal of Finance* 56.4 (2001), pp. 1533–1597. ISSN: 00221082. URL: http://www.jstor.org/stable/2697808.

[21] Jack Hirshleifer. "The Private and Social Value of Information and the Reward to Inventive Activity". English. In: *The American Economic Review* 61.4 (1971), pp. 561–574. ISSN: 00028282. URL: http://www.jstor.org/stable/1811850.

[22] Stewart D. Hodges, Robert G. Tompkins, and William T. Ziemba. "The Favorite/Long-Shot Bias in S&P 500 and Ftse 100 Index Futures Options: The Return to Bets and the Cost of Insurance". English. In: *SSRN eLibrary* (). DOI: 10.2139/ssrn.424421.

[23] Sheena Iyengar, Gur Huberman, and Wei Jiang. "How Much Choice is Too Much? Contributions to 401(k) Retirement Plans". In: ed. by Olivia S. Mitchell and Stephen P. Utkus. Oxford, 2004. Chap. 5.

[24] Sheena S. Iyengar and Emir Kamenica. "Choice proliferation, simplicity seeking, and asset allocation". In: *Journal of Public Economics* 94.7-8 (2010), pp. 530 –539. ISSN: 0047-2727. DOI: DOI:10.1016/j.jpubeco.2010.03.006. URL: http://www.sciencedirect.com/science/article/B6V76-4YTV7V9-1/2/8bb993511622cb211c991864b9993e14.

[25] Michael C. Jensen and William H. Meckling. "Theory of the firm: Managerial behavior, agency costs and ownership structure". In: *Journal of Financial Economics* 3.4 (1976), pp. 305 –360. ISSN: 0304-405X. DOI: 10.1016/0304-405X(76)90026-X. URL: http://www.sciencedirect.com/science/article/pii/0304405X7690026X.

[26] Daniel Kahneman and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk". English. In: *Econometrica* 47.2 (1979), pp. 263–292. ISSN: 00129682. URL: http://www.jstor.org/stable/1914185.

[27] Naresh K. Malhotra. "Information Load and Consumer Decision Making". English. In: *The Journal of Consumer Research* 8.4 (1982), pp. 419–430. ISSN: 00935301. URL: http://www.jstor.org/stable/2489029.

[28] Ulrike Malmendier and Young Han Lee. "The Bidder's Curse". In: *American Economic Review* 101.2 (2011), pp. 749–87. DOI: 10.1257/aer.101.2.749. URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.2.749.

[29] Charles F. Manski. *Interpreting the Predictions of Prediction Markets*. Working Paper 10359. National Bureau of Economic Research, 2004. URL: http://www.nber.org/papers/w10359.

[30] Winter Mason and Siddharth Suri. "Conducting behavioral research on Amazon's Mechanical Turk". In: *Behavior Research Methods* (). 10.3758/s13428-011-0124-6, pp. 1–23. ISSN: 1554-351X. URL: http://dx.doi.org/10.3758/s13428-011-0124-6.

[31] Nathan Novemsky et al. "Preference Fluency in Choice". In: *Journal of Marketing Research* (). DOI: 10.1509/jmkr.44.3.347.

[32] Terrance Odean. "Volume, Volatility, Price, and Profit When All Traders Are above Average". English. In: *The Journal of Finance* 53.6 (1998), pp. 1887–1934. ISSN: 00221082. URL: http://www.jstor.org/stable/117456.

[33] G. Ortner. "Forecasting Markets—An Industrial Application". In: *Universität Wien* (1998).

[34] Marco Ottaviani and Peter Norman Sørensen. "The strategy of professional forecasting". In: *Journal of Financial Economics* 81.2 (2006), pp. 441 –466. ISSN: 0304-405X. DOI: 10.1016/j.jfineco.2005.08.002. URL: http://www.sciencedirect.com/science/article/pii/S0304405X06000286.

[35] Charles R. Plott and Kay-Yut Chen. *Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem*. Working Papers 1131. California Institute of Technology, Division of the Humanities and Social Sciences, 2002. URL: `http://EconPapers.repec.org/RePEc:clt:sswopa:1131`.

[36] C.R. Plott. "Markets as information gathering tools". In: *Southern Economic Journal* 67.1 (2000), pp. 2–15.

[37] R. Roll. "Orange juice and weather". In: *The American Economic Review* 74.5 (1984), pp. 861–880.

[38] Joel Ross et al. "Who are the crowdworkers?: shifting demographics in mechanical turk". In: *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*. CHI EA '10. Atlanta, Georgia, USA: ACM, 2010, pp. 2863–2872. ISBN: 978-1-60558-930-5. DOI: `http://doi.acm.org/10.1145/1753846.1753873`. URL: `http://doi.acm.org/10.1145/1753846.1753873`.

[39] Benjamin Scheibehenne, Rainer Greifeneder, and Peter M. Todd. "Can There Ever Be Too Many Options? A MetaAnalytic Review of Choice Overload". English. In: *Journal of Consumer Research* 37.3 (2010), pp. 409–425. ISSN: 00935301. URL: `http://www.jstor.org/stable/10.1086/651235`.

[40] Wayne W. Snyder. "Horse Racing: Testing the Efficient Markets Model". English. In: *The Journal of Finance* 33.4 (1978), pp. 1109–1118. ISSN: 00221082. URL: `http://www.jstor.org/stable/2326943`.

[41] Justin Wolfers and Eric Zitzewitz. *Interpreting Prediction Market Prices as Probabilities*. Working Paper 12200. National Bureau of Economic Research, 2006. URL: `http://www.nber.org/papers/w12200`.

[42] Justin Wolfers and Eric Zitzewitz. *Prediction Markets*. Working Paper 10504. National Bureau of Economic Research, 2004. URL: `http://www.nber.org/papers/w10504`.